LNCS 6768

Constantine Stephanidis (Ed.)

Universal Access in Human-Computer Interaction

Applications and Services

6th International Conference, UAHCI 2011
Held as Part of HCI International 2011
Orlando, FL, USA, July 2011, Proceedings, Part IV

Part IV

HCI2011
INTERNATIONAL

Springer

Lecture Notes in Computer Science 6768

Constantine Stephanidis (Ed.)

Universal Access in Human-Computer Interaction

Applications and Services

6th International Conference, UAHCI 2011
Held as Part of HCI International 2011
Orlando, FL, USA, July 9-14, 2011
Proceedings, Part IV

Springer

Volume Editor

Constantine Stephanidis
Foundation for Research and Technology - Hellas (FORTH)
Institute of Computer Science
N. Plastira 100, Vassilika Vouton, 70013, Heraklion, Crete, Greece
and
University of Crete
Department of Computer Science
Crete, Greece
E-mail: cs@ics.forth.gr

ISSN 0302-9743 e-ISSN 1611-3349
ISBN 978-3-642-21656-5 e-ISBN 978-3-642-21657-2
DOI 10.1007/978-3-642-21657-2
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2011928824

CR Subject Classification (1998): H.5, K.6, H.3-4, C.2, D.2, J.1, J.3

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

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

Printed on acid-free paper

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

Foreword

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

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

This volume, edited by Constantine Stephanidis, contains papers in the thematic area of universal access in human-computer interaction (UAHCI), addressing the following major topics:

- Speech, communication and dialogue
- Interacting with documents and images
- Universal access in complex working environments
- Universal access to education and learning
- Well-being, health and rehabilitation applications

The remaining volumes of the HCI International 2011 Proceedings are:

- Volume 1, LNCS 6761, Human–Computer Interaction—Design and Development Approaches (Part I), edited by Julie A. Jacko
- Volume 2, LNCS 6762, Human–Computer Interaction—Interaction Techniques and Environments (Part II), edited by Julie A. Jacko
- Volume 3, LNCS 6763, Human–Computer Interaction—Towards Mobile and Intelligent Interaction Environments (Part III), edited by Julie A. Jacko
- Volume 4, LNCS 6764, Human–Computer Interaction—Users and Applications (Part IV), edited by Julie A. Jacko
- Volume 5, LNCS 6765, Universal Access in Human–Computer Interaction—Design for All and eInclusion (Part I), edited by Constantine Stephanidis

- Volume 6, LNCS 6766, Universal Access in Human–Computer Interaction—Users Diversity (Part II), edited by Constantine Stephanidis
- Volume 7, LNCS 6767, Universal Access in Human–Computer Interaction—Context Diversity (Part III), edited by Constantine Stephanidis
- Volume 9, LNCS 6769, Design, User Experience, and Usability—Theory, Methods, Tools and Practice (Part I), edited by Aaron Marcus
- Volume 10, LNCS 6770, Design, User Experience, and Usability—Understanding the User Experience (Part II), edited by Aaron Marcus
- Volume 11, LNCS 6771, Human Interface and the Management of Information—Design and Interaction (Part I), edited by Michael J. Smith and Gavriel Salvendy
- Volume 12, LNCS 6772, Human Interface and the Management of Information—Interacting with Information (Part II), edited by Gavriel Salvendy and Michael J. Smith
- Volume 13, LNCS 6773, Virtual and Mixed Reality—New Trends (Part I), edited by Randall Shumaker
- Volume 14, LNCS 6774, Virtual and Mixed Reality—Systems and Applications (Part II), edited by Randall Shumaker
- Volume 15, LNCS 6775, Internationalization, Design and Global Development, edited by P.L. Patrick Rau
- Volume 16, LNCS 6776, Human-Centered Design, edited by Masaaki Kurosu
- Volume 17, LNCS 6777, Digital Human Modeling, edited by Vincent G. Duffy
- Volume 18, LNCS 6778, Online Communities and Social Computing, edited by A. Ant Ozok and Panayiotis Zaphiris
- Volume 19, LNCS 6779, Ergonomics and Health Aspects of Work with Computers, edited by Michelle M. Robertson
- Volume 20, LNAI 6780, Foundations of Augmented Cognition: Directing the Future of Adaptive Systems, edited by Dylan D. Schmorrow and Cali M. Fidopiastis
- Volume 21, LNAI 6781, Engineering Psychology and Cognitive Ergonomics, edited by Don Harris
- Volume 22, CCIS 173, HCI International 2011 Posters Proceedings (Part I), edited by Constantine Stephanidis
- Volume 23, CCIS 174, HCI International 2011 Posters Proceedings (Part II), edited by Constantine Stephanidis

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

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

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

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

July 2011 Constantine Stephanidis

Organization

Ergonomics and Health Aspects of Work with Computers

Program Chair: Michelle M. Robertson

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

Human Interface and the Management of Information

Program Chair: Michael J. Smith

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

Human–Computer Interaction

Program Chair: Julie A. Jacko

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

Engineering Psychology and Cognitive Ergonomics

Program Chair: Don Harris

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

Universal Access in Human–Computer Interaction

Program Chair: Constantine Stephanidis

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

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

Virtual and Mixed Reality

Program Chair: Randall Shumaker

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

Internationalization, Design and Global Development

Program Chair: P.L. Patrick Rau

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

Online Communities and Social Computing

Program Chairs: A. Ant Ozok, Panayiotis Zaphiris

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

Augmented Cognition

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

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

Digital Human Modeling

Program Chair: Vincent G. Duffy

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

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

Human-Centered Design

Program Chair: Masaaki Kurosu

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

Design, User Experience, and Usability

Program Chair: Aaron Marcus

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

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

HCI International 2013

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

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

Table of Contents – Part IV

Part I: Speech, Communication and Dialogue

Part II: Interacting with Documents and Images

Part III: Universal Access in Complex Working Environments

Part IV: Well-Being, Health and Rehabilitation Applications

Part V: Universal Access to Education and Learning

Part I

Speech, Communication and Dialogue

Greek Verbs and User Friendliness in the Speech Recognition and the Speech Production Module of Dialog Systems for the Broad Public

Christina Alexandris[1] and Ioanna Malagardi[2]

[1] National and Kapodistrian University of Athens, Greece
[2] Department of Informatics & Telecommunications
Educational & Language Technology Laboratory
National and Kapodistrian University of Athens
calexandris@gs.uoa.gr, imalagar@di.uoa.gr

Abstract. It has been observed that for the Greek language, in Service- Oriented Dialog Systems targeted towards the broad public, verbs display particular features to be considered both in the System's Speech Recognition (ASR) and Speech Synthesis (or Speech Production) Module. Additionally, the function of verbs, both in respect to their role in the semantic content of the spoken utterance and in respect to their prosodic features in the spoken utterance, is observed to be directly linked to the step and the related Speech Act in the dialog structure. From a prosodic aspect, it is observed that, in spoken input, "Multitasking" verbs do not receive prosodic emphasis, whereas in spoken output, prosodic emphasis is given on words signalizing the User-System Relationship.

Keywords: prosodic emphasis, verb ontology, User-friendliness, "Multitasking" verbs, User-System Relationship expressions.

1 Introduction

Service- Oriented Dialog Systems targeted towards the broad public involve a higher percentage of non-sublanguage specific vocabulary and a lower percentage of terminology and professional jargon. Unlike highly specialized Task-related Dialog Systems, in Service- Oriented Dialog Systems, the Human-Computer interaction taking place is directed towards two equally significant goals, namely the successful performance of the activated or requested task and User satisfaction and User-friendliness. These goals are related to requirements on the Satisfaction Level in respect to a System's evaluation criteria, namely perceived task success, comparability of human partner and trustworthiness [10]. In some cases of successful design and implementation, even a sense of attachment of the User in respect to the System can be possible [9]. It should be noted that the more goals to be achieved, the more parameters in the System Design and System Requirements, and subsequently Dialog Design are to be considered [14]. Prosodic modelling for the Speech Acts in Service-Oriented Dialog Systems may, therefore, be characterized as a complex task.

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 3–11, 2011.

In User input recognized by Service- Oriented Dialog Systems, verbs play an essential role in the vocabulary of User input. In Service- Oriented Dialog Systems, verbs play an equally essential role when they are produced as spoken output by the System's Speech Processing Module. It has been observed that for the Greek language, in Service- Oriented Dialog Systems targeted towards the broad public, verbs display particular features to be considered both in the System's Speech Recognition (ASR) and Speech Synthesis (or Speech Production) Module.

The function of verbs, both in respect to their role in the semantic content of the spoken utterance and in respect to their prosodic features in the spoken utterance, is observed to be directly linked to the step and the related Speech Act in the dialog structure. In spoken Greek, non-sublanguage specific verbs in user input may be characterized by polysemy and absence of prosodic emphasis (I). In contrary, non-sublanguage specific verbs in the System's spoken output contain a semantic feature expressing the System-User relationship and receive prosodic emphasis for the achievement of friendliness towards the User (II). Furthermore, in User input, the correct interpretation of non-sublanguage specific verbs in user input requires the identification of the correct Speech Act (A). Similarly, in the System's spoken output, the context of a so-called "Non-Task-related" Speech Act in dialog structure [1] is the condition for non-sublanguage specific verbs to express the System-User relationship (B).

1.1 Verb Classes and Ontologies: Automatic Grouping Practices

Here, we will address two types of verb classes: (1) "Multitasking" verbs, usually occurring high in the verb ontology, some of them constituting "primitive" verbs and (2) verbs constituting "User-System-Relationship" expressions, usually occurring lower in the verb ontology, most of them constituting verbs with one or more objects.

The "Multitasking" verbs used in input phrases must be interpreted in their basic "primitive" interpretation and the verbs constituting "User-System-Relationship" expressions, usually occurring lower in the verb ontology, are interpreted in the form they occur. The analysis of the definitions of the above-mentioned classes of verbs is used for the application of the understanding and description of actions.

Verb classes can be identified and grouped automatically. Previous studies [5], [8] have demonstrated that computational processing of verbs involves a process of automatic grouping of verbs, namely that the definitions of the verb's features are given as input to a System producing an output that renders a grouping of the verbs and synthesized definitions of these verbs using "primitives". The set of verb entries related to the above categories used as input to a System can be identified on the basis of chains containing their definitions. The verb at the end of a chain was used as the criterion of verb identification and verb grouping [5], [8].

Here, we note that prior to using these chains, it was necessary to eliminate the cyclic parts of the definition chains which were also automatically detected by the System. The definitions of the verbs in each definition are, in turn, retrieved from the lexicon and, in this way, chains of definitions are formed. These chains end up in circularities that correspond to reaching basic verbs, "basic", representing primitive actions [5]. The elimination of the circularity that occurs in certain chains requires the choice of the suitable verb as a terminal of the chain. The choice for each case of elimination of circularity requires the adoption of some ontology [5], [8].

1.2 Verb Classes and Ontologies – The Prosody and Semantics Relation

In respect to Speech Processing, "Multitasking" verbs are observed to be related to the special prosodic feature of the absence of prosodic emphasis in regard to the recognition of the Speaker's or User's spoken input. In contrary, verbs constituting "User-System-Relationship" expressions are observed to be related to the prosodic feature of prosodic emphasis in respect to the production of the Systems's spoken output. In respect to User-friendliness, within the Speech-Act framework of Service-Oriented Dialog Systems, the absence of prosodic emphasis on "Multitasking" verbs contributes to the achievement of clarity and processability of User-input, while the possibility of using these verbs in User input allows the User to speak in a more natural way to the System, without too many restrictions in utterance content. On the other hand, prosodic emphasis on "User-System-Relationship" verbs contributes to the achievement of naturally sounding, comprehensible and friendly System-output.

The relationship of "Multitasking" and "User-System-Relationship" verbs in respect to prosodic emphasis may be depicted in the following table (Table 1).

Table 1. Relationship of "Multitasking" and "User-System-Relationship" verbs in respect to prosodic emphasis

[-prosodic emphasis]:	[+prosodic emphasis]:
System Input	System Output
Verb type: "Multitasking" verbs	Verb type: "User-System-Relationship" expressions
Speech Act: Task-Related	Speech Act: Non Task-Related
User-friendliness Goal: clarity and processability for System with few restrictions in User-Input	User-friendliness Goal: naturally sounding, comprehensible and friendly System-Output

Furthermore, it may noted that, at least within the Speech-Act framework of Service-Oriented Dialog Systems, verbs with a relatively high content of specifying features in their semantics, such as verbs signalizing the User-System-Relationship, have a higher likelihood to receive prosodic emphasis, whereas verbs with a more vague or general semantic content and often occurring higher in semantic ontologies ("primitives") have a higher likelihood not to receive prosodic emphasis. The question of whether similar observations can be made in respect to other verb classes and/or contexts other than the Speech-Act framework of Service-Oriented Dialog Systems, remains an issue under investigation.

2 Prosody and "Multitasking Verbs" in Spoken Input

From the aspect of the System's Speech Recognition (ASR) Module, one basic problem that requires to be addressed is the existence of "Multitasking" verbs which may be described as verbs related to multiple semantic meanings and used in a variety of expressions, existing in Modern Greek, and possibly in other languages as well. For example, in the sublanguage related to the communication context of the CitizenShield system for the Service sector (consumer complaints) [11], the semantically related verbs “buy”, “get” and “purchase” may be used in similar expressions as the (primitive) verbs “is” and “have”, as well as the verbs “give” and “receive”, to convey the same semantic meaning from the speaker. This possibility is illustrated in the following examples (Table 2).

Table 2. “Multitasking” verbs in the communication context of the CitizenShield system for the Service sector (consumer complaints) (User-input)

[-prosodic emphasis]	Examples (User-input):
have, get	“I got this cereal” / “I have this cereal here” “They give misguiding information on the package”
be, is	“They are from a convenience store”
	“This is about a cereal”
see	“I saw some misguiding information on the package”

The rather rigid and controlled nature of the CitizenShield System’s dialog structure allows the application of constraints in regard to the processing of the correct interpretation of “Multitasking” verbs. The CitizenShield System’s dialog structure is based on Task-related Speech Acts involving Directed Dialogs [15], [16], most of which involve Yes-No Questions or questions directed towards Keyword answers. From a prosodic aspect, it is observed that, unlike sublanguage-specific verbs with more restricted semantics, “Multitasking” verbs do not receive prosodic emphasis when uttered by Greek native speakers as spoken input to be processed by the System’s Speech Recognition (ASR) Module and are, therefore, of less significance in the semantics of the spoken utterance. This feature enables their efficient processing, since they are recognizable by the absence of prosodic emphasis and, consequently, can be identifiable by an attached marker signifying absence of prosodic emphasis. For the achievement of clarity and processability as spoken input to be processed by the System’s Speech Recognition (ASR) Module, “Multitasking” verbs do not receive prosodic emphasis and, therefore, an additional prosodic marker signifying absence of prosodic emphasis is proposed.

Both in monolingual and in multilingual applications, the traditional approach constructed around a verb-predicate signalizing the basic semantic content of the utterance, the so-called “frame”, also typically used in traditional Interlinguas [6], may not always be adequate for the polysemy of “Multitasking” verbs in User input. Thus lexical-based alternatives linked to Speech Act types in respect to the steps dialog context are proposed. Specifically, the correct interpretation of the semantics of “Multitasking” verbs is directly related to the type of Speech Act expressed in the spoken utterance.

In an attempt to meet the needs of the diverse community of foreign residents in Greece, Interlinguas (ILTS) are used as semantic templates for a possible multilingual extension of the CitizenShield dialog system for consumer complaints. The proposed Interlinguas are designed to function within a very restricted sublanguage related to a specific task, but at the same time, for a very diverse range of languages and language families, allow minimum interference of language-specific factors.

Table 3. Examples of Simple Interlinguas with "Multitasking" verbs

S-FRAME	Keyword and Lexical Content of "S-ILT"
[S-FRAME = GOT [-prosodic emphasis]	WHO (PERSON) WHAT (ITEM) QUANTITY (QUANT), PRICE (NUM-EURO) WHERE (PLACE) WHEN (DAY)]
[S-FRAME = HAVE [-prosodic emphasis]	WHO (PERSON) WHAT (ITEM) QUANTITY (QUANT), PRICE (NUM-EURO) WHERE (PLACE) WHEN (DAY)]
[S-FRAME = IS [-prosodic emphasis]	WHO (PERSON) WHAT (ITEM) QUANTITY (QUANT), PRICE (NUM-EURO) WHERE (PLACE) WHEN (DAY)]

The proposed S-Interlinguas (Table 3) may be characterized by a very simple structure and to be more of Interlinguas with an accepting or rejecting input function rather than the traditional Interlinguas with the function of summarizing the semantic content of a spoken utterance [6], [12]. Thus, in the present application, the role of the "frame" in the S-Interlingua structure is weakened and the core of the semantic content is shifted to the lower level of the lexical entries. The lexical level allows the possibility to add a large number of variants of terms in different languages.

The simple structure of the S-Interlingua (S-ILT) may be differentiated in three levels: (1) the Template Level, the Keyword-Content Level (2) and (3) the Lexical-Level. The Template Level constitutes the top level and is loosely associated with the verb of the speakers utterance, however, its function is more that of "wrapping up" the key-word contents of the speakers response than the original function of signalizing the sentence content. The "frame" level (S-FRAME) will not signalize the meaning of the sentence: this task will be performed by the lexical entries. However, in most cases, the "frame" level is retained for purely functional purposes, namely for connecting the lexical entries to each other and facilitating the rejection of possible input whose semantic content is irrelevant and not contained within the sublanguage of the application. This rejected input, however, may contain individual words initially recognized by the system at the Lexical Level

3 Prosody and "System-User Relationship" Verbs in Spoken Output

In respect to spoken output by the System's Speech Processing Module, it has been observed that for languages like Greek, where friendliness is related to directness and spontaneity, constituting features of Positive Politeness [13], from a prosodic aspect, User-friendliness can be achieved with prosodic emphasis on verbs functioning as elements expressing the User-System Relationship.

Words signalizing the User-System Relationship can be subsumed under the general category of expressions involving the System's or User's positive intention or cooperation and may be related to respective Non-Task-related Speech Acts. These word categories may be described as expressions related to the System's positive attitude toward the User and be categorized as (1) system-service verbs, (2) system-intention verbs, (3) system-service nouns, (4) system-intention nouns, (5) user-intention verbs and (5) user-action verbs.Typical examples of system-service verbs are the verbs "give" and "show" and nominalized user-intention verbs such as "cooperation". Examples of system-intention verbs (including nominalized system-intention verbs) are the verbs "help" and "assist". The verbs "wish"/"want" and "finished" are typical examples of user-intention (or user intended action) and user action verbs respectively.

Specifically, User-friendliness can be achieved with prosodic emphasis on verbs (or nominalized verbs) functioning as elements expressing the User-System Relationship in the context of Non-Task-related Speech Acts such as "Apologize" and "Delay" [1].

3.1 Defining Non-task Related Speech Acts in Dialog Structure

Data from European Union projects in Speech Technology for social services and Human-Computer Interaction in English, German and Greek [17],[18],[19],[20] allows the formulation of a general categorization scheme of Non-Task-related Speech Acts. Specifically, Non-Task-related Speech Acts can be divided into three main categories: Speech Acts constituting an independent step in dialog structure (Category 1, for example, "Close Dialog":"Thank you for using the Quick-Serve Interface"), Speech Acts attached to other Task-Related Speech Acts [4] constituting with them a singular step in dialog structure (Category 2, for example, "I am sorry" ("Apologize") following or preceeding the Task-Related Speech Act ("Inform"), "Your input cannot be processed" or "I cannot understand your request" ("Justify") following or preceeding the Task-Related Speech Act ("Request")) and Speech Acts constituting an optional step in dialog structure of Service-Oriented dialogs (Category 3), for example, "Reminder": "You still have two minutes to complete this transaction" [1].

3.2 Prosodic Modeling and the User-System Relationship in Greek

In the following examples (Table 4) from the CitizenShield dialog system [11], the above listed types of words signalizing the User-System Relationship receive prosodic emphasis ("Usr-Sys-Rel prosodic emphasis", indicated as [Usr-Sys-Rel prosod] following the emphasized word). These words are the expressions "sorry" (system-intention noun – in Greek), "hear", "request" (system-service verbs), "thank" (system-intention verb), "cooperation" (nominalised user-intention verb) and "completed" (user-action verb).

At this point, it is important to stress that not all Non-Task-related Speech Acts necessarily contain "Usr-Sys-Rel" expressions. It should, additionally, be noted that expressions signalizing negations, temporal and spatial information, quantity and quality, as well as sublanguage-specific task-related expressions receive prosodic emphasis by default ("default prosodic emphasis", indicated in italics) [2]. In Table 4, these are the expressions “not”, “some”, “very much”, “more”, “additional” and “obviously” [3].

Additionally, it should be stressed that in Non-Task-related Speech Acts, “Usr-Sys-Rel prosodic emphasis” has a priority over “default prosodic emphasis” in respect to amplitude. Specifically, in Non-Task-related Speech Acts, the amplitude of the prosodic emphasis on Usr-Sys-Rel expressions is intended to be slightly higher than the amplitude of the prosodic emphasis on expressions receiving default prosodic emphasis.

Table 4. Examples with User-System-Relationship Expressions (CitizenShield System)

[+prosodic emphasis]: for “Usr-Sys-Rel” expressions and “default” prosodic emphasis

1. **Συγγνώμη**[Usr-Sys-Rel prosod] ***δεν*** σας **άκουσα**[Usr-Sys-Rel prosod]
(Non-Task-related Speech Act: “Justify”)

2. Θα σας **ζητήσω**[Usr-Sys-Rel prosod] ***μερικές*** πληροφορίες ***ακόμα***
(Non-Task-related Speech Act: “Introduce-new-task”)

3. Σας **ευχαριστούμε**[Usr-Sys-Rel prosod] ***πολύ*** για την **συνεργασία**
[Usr-Sys-Rel prosod]σας
(Non-Task-related Speech Act: “Thank”)

4. ***Προφανώς*** **ολοκληρώσατε**[Usr-Sys-Rel prosod] με τις ***επιπλέον*** πληροφορίες
(Non-Task-related Speech Act: “Reminder”)

Translations close to the syntax of original spoken utterances

1. I am **sorry**[Usr-Sys-Rel prosod]**,** I could ***not*** **hear**[Usr-Sys-Rel prosod] you**.**
(Non-Task-related Speech Act:“Justify”)

2. I will **request** [Usr-Sys-Rel prosod] from you ***some more*** information.
(Non-Task-related Speech Act: “Introduce-new-task”)

3. We **thank**[Usr-Sys-Rel prosod] **you** ***very much*** for your **cooperation**
[Usr-Sys-Rel prosod]
(Non-Task-related Speech Act: “Thank”)

4. You have ***obviously*** **completed**[Usr-Sys-Rel prosod] providing the
additional input
(Non-Task-related Speech Act: “Reminder”)

We note that, in Greek, as a verb-framed and pro-drop language (like Spanish or Italian), the prosodic emphasis is directly matched to the finite verb, containing the features of the verb’s subject - in this case the System or the User. This difference in respect to languages such as English may also influence the process of identifying Usr-Sys-Rel expressions.

For example, the Greek verb “zi΄tiso” (“request”) in the context of “make questions” contains the features of the verb’s subject. In the context of Service-oriented HCI applications, the Greek verb “zi΄tiso” may be identified as an Usr-Sys-Rel expression. In another example, the Greek verb “olokli΄rosate” (“finished-completed”) is equivalent to the verb “finished” in English. We note that the semantics of the Greek verbs “zi΄tiso” and “olokli΄rosate” allows them to be classified as a Usr-Sys-Rel expression, whereas the respective verbs “request” and “finished” in English are classified as verbs signalizing an ACTION-TYPE [7], a task-related expression receiving default emphasis [1],[3],[7].

4 Conclusions and Further Research

In Service- Oriented Dialog Systems, the semantic content and the prosodic features of verbs are observed to be directly linked to the step and the related Speech Act in the dialog structure.

In spoken Greek Input, non-sublanguage specific verbs in user input may be characterized by polysemy and absence of prosodic emphasis (1). In contrary, non-sublanguage specific verbs in the System’s spoken output contain a semantic feature expressing the System-User relationship and receive prosodic emphasis for the achievement of User-friendliness (2).

Specifically, it is observed that, unlike sublanguage-specific verbs with more restricted semantics, “Multitasking” verbs do not receive prosodic emphasis when uttered by Greek native speakers as spoken input to be processed by the System’s Speech Recognition (ASR) Module and, therefore, play a less important role in the semantics of the spoken utterance. Thus, traditional practices based on a verb-predicate signalizing the basic semantic content of the utterance, the so-called “frame”, also typically used in traditional Interlinguas [6], may not always be efficient for the processing of “Multitasking” verbs in User input. Therefore, lexical-based alternatives linked to Speech Act types in dialog context are proposed, especially in the case of multilingual applications in diverse languages and language families.

On the other hand, in spoken Greek output, User-friendliness can be achieved with prosodic emphasis on verbs functioning as elements expressing the User-System Relationship in the context of Non-Task-related Speech Acts, as defined, based on data within the framework of European Union projects. Prosodic emphasis can be directly matched to the finite verb, constituting a User-System Relationship expression and containing the features of the verb’s subject - in this case the System or the User. This is possible in a verb-framed and pro-drop language such as Greek, however, may be problematic in non verb-framed and non pro-drop languages.

Further research is required to determine whether similar phenomena are observed in respect to verbs or to other word categories in Service- Oriented Dialog Systems of other languages, possibly also in Dialog Systems of different domains.

References

1. Alexandris, C.: Speech Acts and Prosodic Modeling in Service-Oriented Dialog Systems. In: Computer Science Research and Technology. Nova Science Publishers, Hauppauge (2010)
2. Alexandris, C.: Show and Tell: Using Semantically Processable Prosodic Markers for Spatial Expressions in an HCI System for Consumer Complaints. In: Jacko, J.A. (ed.) HCI 2007. LNCS, vol. 4552, pp. 13–22. Springer, Heidelberg (2007)

3. Alexandris, C.: Word Category and Prosodic Emphasis in Dialog Modules of Speech Technology Applications. In: Botinis, A. (ed.) Proceedings of the 2nd ISCA Workshop on Experimental Linguistics, ExLing 2008, Athens, Greece, pp. 5–8 (August 2008)
4. Heeman, R., Byron, D., Allen, J.F.: Identifying Discourse Markers in Spoken Dialog. In: Proceedings of the AAAI Spring Symposium on Applying Machine Learning to Discourse Processing, Stanford (March 1998)
5. Kontos, J., Malagardi, I., Pegou, M.: Processing of Verb Definitions from Dictionaries. In: Proceedings of the 3rd International Conference in Greek Linguistics, Athens, pp. 954–961 (1997) (in Greek)
6. Levin, L., Gates, D., Lavie, A., Pianesi, F., Wallace, D., Watanabe, T., Woszczyna, M.: Evaluation of a Practical Interlingua for Task-Oriented Dialog. In: Proceedings of ANLP/NAACL 2000 Workshop on Applied Interlinguas, Seattle, WA (April 2000)
7. Malagardi, I., Alexandris, C.: Verb Processing in Spoken Commands for Household Security and Appliances. In: Stephanidis, C. (ed.) UAHCI 2009, Part II. LNCS, vol. 5615, pp. 92–99. Springer, Heidelberg (2009)
8. Malagardi, I. , Kontos, J.: Motion Verbs and Vision. In: Proceedings of the 8th Hellenic European Research on Computer Mathematics & its Applications Conference (HERCMA 2007), Athens (2007), http://www.aueb.gr/pympe/hercma/proceedings2007
9. Matsumoto, N., Ueda, H., Yamazaki, T., Murai, H.: Life with a Robot Companion: Video Analysis of 16-Days of Interaction with a Home Robot in a "Ubiquitous Home" Environment. In: Jacko, J.A. (ed.) HCI International 2009. LNCS, vol. 5611, pp. 341–350. Springer, Heidelberg (2009)
10. Moeller, S.: Quality of Telephone-Based Spoken Dialog Systems. Springer, New York (2005)
11. Nottas, M., Alexandris, C., Tsopanoglou, A., Bakamidis, S.: A Hybrid Approach to Dialog Input in the CitzenShield Dialog System for Consumer Complaints. In: Proceedings of HCI 2007, Beijing, Peoples Republic of China (2007)
12. Schultz, T., Alexander, D., Black, A., Peterson, K., Suebvisai, S., Waibel, A.: A Thai speech translation system for medical dialogs. In: Proceedings of the conference on Human Language Technologies (HLT-NAACL), Boston, MA, USA (2004)
13. Sifianou, M.: Discourse Analysis. An Introduction. Leader Books, Athens (2001)
14. Wiegers, K.E.: Software Requirements. Microsoft Press, Redmond (2005)
15. Williams, J.D., Witt, S.M.: A Comparison of Dialog Strategies for Call Routing. International Journal of Speech Technology 7(1), 9–24 (2004)
16. Williams, J.D., Poupart, P., Young, S.: Partially Observable Markov Decision Processes with Continuous Observations for Dialog Management. In: Proceedings of the 6th SigDial Workshop on Discourse and Dialog, Lisbon (September 2005)
17. The Agent-DYSL Project, http://www.agent-dysl.eu/
18. The HEARCOM Project, http://hearcom.eu/main.html
19. The ERMIS Project, http://www.image.ntua.gr/ermis/
20. The SOPRANO Project, http://www.soprano-ip.org/

Intercultural Dynamics of First Acquaintance: Comparative Study of Swedish, Chinese and Swedish-Chinese First Time Encounters

Jens Allwood, Nataliya Berbyuk Lindström, and Jia Lu

SSKKII/SCCIIL and Dept. of Applied IT,
Chalmers & University of Gothenburg, Sweden
jens@ling.gu.se
{berlinds,lulux}@chalmers.se

Abstract. Today, intercultural first acquaintance meetings are becoming more and more frequent. The aim of this study is to describe, analyze and compare Swedish, Chinese and Swedish-Chinese first acquaintance interactions. Our focus lies on a classification of the topics in mono- and intercultural first-time encounters. The analysis is based on 12 arranged face-to-face first acquaintance interactions between Chinese-Chinese, Swedish-Swedish and Swedish-Chinese students (4 of each dyad). The interactions are video-recorded and transcribed. In addition, semi-structured interviews with the participants have been conducted to get a better understanding of their communication. The method of activity-based communication analysis is used to analyze the data. The result of the study is a classification and a cross-cultural comparison of topics and the order of their occurrence in first time encounters. In addition, the study sheds light on the similarities and differences between Chinese and Swedish communication patterns.

Keywords: first-time encounters, intercultural communication, Chinese, Swedes, cross-cultural comparison.

1 Introduction. First-Time Encounters and Culture

In everyday life people often meet and talk to each other for the first time, often without giving it much thought. Many such first-time encounters are spontaneous and brief, e.g. talking to a stranger at the bus stop, interacting with another customer in a shop, etc. The initiation of interaction is usually related to a certain shared context and activity for the people involved, and the topics are related to the context (e.g. talking about traffic delays while waiting at the bus stop, discussing sale prices with other customers, etc). Some first-time encounters are spontaneous (like meeting at the bus stop) while others are arranged, like a booked job interview. In the arranged meetings, there is often a more or less pre-defined topic for interaction, e.g. previous work experiences are in focus in a job interview.

Today many first-time encounters are intercultural. Meeting a person with a differ-

The original version of this chapter was revised: an error in the title of the paper was corrected. The correction to this chapter is available at https://doi.org/10.1007/978-3-642-21657-2_73

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 12–21, 2011.

ent cultural background can be an exciting experience, but also a challenge, causing stress, anxiety and uncertainty about appropriate communicative behavior. Language problems and cultural differences can often though not always have a negative impact on the interactions [1]. This study aims to throw light on people's communicative behavior in first-time encounters in general, and in intercultural encounters in particular, with a focus on the choice of topics in interactions.

2 Aim and Specific Questions

The aim of the study is to describe, analyze and to compare Swedish, Chinese and Swedish-Chinese first acquaintance interactions. Our focus lies on a classification and comparison of topics in mono- and intercultural first time encounters. The specific questions of the study are: (1) What are the topics discussed in first time encounters? (2) Are there any differences and/or similarities between monocultural (Swedish-Swedish and Chinese-Chinese) and intercultural (Chinese-Swedish) interactions in terms of topics discussed and the order in which the topics are introduced in conversations? (3) How can the observed differences and similarities be explained?

3 Material and Method

The analysis is based on 12 arranged video-recorded face-to-face first acquaintance interactions between Chinese and Swedish university students (4 of each dyad, Chinese-Chinese, Swedish-Swedish and Swedish-Chinese) in combination with semi-structured interviews with the participants. Four Swedes and four Chinese (two male and two female of each, eight participants in total) participated in the study. The age range of the participant is 23-30 years. All of them at the recording time were students at the University of Gothenburg. Each interaction involved two students who had not met each other before the recording. The languages used in the interactions are the participants' native languages (Swedish and Mandarin Chinese) while the Swedish-Chinese interactions were conducted in English.

The participants were asked to get acquainted, then the project assistant left the room to come back in about 10 minutes. The research team set no restrictions on the topics to be discussed by the participants. The camera was placed outside the participants' field of vision to minimize the possible effect of video recording on participants' behavior.

Semi-structured interviews were conducted with the participants after the recording. The interviews included questions about the topics discussed, reasons for choosing these topics and differences in communication with regard to people from the same and different cultural backgrounds. All interviews were audio-recorded.

The interactions were transcribed directly from the recordings by the native speakers of Swedish and Chinese. GTS and MSO (Göteborg Transcription Standard + Modified Standard Orthography [2], developed at the Department of Linguistics and Center SSKKII/SCCIIL, University of Gothenburg were used. The standard allows inclusion of such features of spoken language as overlaps, pauses, comments on communicative body movements and other events. The transcriptions were checked. The recordings and transcribed interactions were then analyzed using Activity-based communication analysis [3]. The result of the analysis is a detailed analysis of first acquaintance as a social activity, with a special focus on the topics discussed. A number of topics have been identified, and the order in which the topics are taken up in interactions has been analyzed. The analysis is supported by examples from the transcriptions and comments from the interviews.

4 Results

4.1 First Acquaintance as a Social Activity. Some General Comments on the Interactions

Since the recorded encounters are arranged, this has an impact on the behavior of the participants. For example, one of the **purposes** of the interactions is to participate in a research project, another is to get know each other and perhaps to some extent to "kill time" and have fun. The **participants** are native speakers of Swedish and Chinese, and, though all have a good command of English as a second language, difficulties in using English might have had some influence on their communicative behavior. All participants live in Gothenburg or in the districts around Gothenburg. The recordings took place in a room at the university (**physical environment**), which might have created a more formal atmosphere than if the recordings had been made outside the university. Consider a comment from one of the Swedish respondents, which highlights the impact of physical context on the topics discussed:

> *"If you know that you are in school and you ask what the other person studies… however when a new person comes to my job it is a different thing."*
>
> *Swedish female participant*

The respondents commented that in general it was challenging to participate in interactions of this kind. Some report being nervous. Another general comment about communication was that the participants experienced more tension and difficulties in interactions with interlocutors from a different cultural background than their own:

> *"Feel happy and enjoy in the communication* [with the Chinese]… *Feel relax. I felt it was a little bit hard to find topics and shared experiences* [with the Swedes]. *The topics were more narrow and public"*
>
> *Chinese female participant*

Table 1. Topics in the first-time encounters

<table>
<tr><th>Type</th><th colspan="8">Topic</th></tr>
<tr><td rowspan="3">I.
Personal identity</td><td rowspan="2">1. Greeting</td><td rowspan="2">2. Name Introduction</td><td rowspan="2">3. Age</td><td colspan="2">4. Family</td><td rowspan="2">5. Religion</td><td rowspan="2">6. Nationality</td><td rowspan="2">7. Money and econo my</td></tr>
<tr><td>a) partners</td><td>b) parents/siblings</td></tr>
<tr><td>Saying hello, hi, ni hao (Chinese: hello)</td><td>Introducing/both parties presenting names</td><td>Age-related information</td><td colspan="2">Talking about marriage, partners, parents, siblings (e.g. their origin and education, etc)</td><td>Talking about religion and church</td><td>Talking/asking about nationality</td><td>Economic issues (private economy, debt)</td></tr>
<tr><td rowspan="3">II.
Occupation and residence</td><td colspan="5">8. Occupation/studies</td><td colspan="3">9. Place of living</td></tr>
<tr><td colspan="2">a) current</td><td>b) previous</td><td colspan="2">c) future</td><td colspan="2">a) current</td><td>b) previous</td></tr>
<tr><td colspan="5">Talking about current work/studies, earlier work/studies and future plans concerning work/studies</td><td colspan="3">Talking about current, earlier and future plans concerning residence, such as city, district, apartment, etc.</td></tr>
<tr><td rowspan="3">III. Purpose of meeting/acquaintance and personal experiences</td><td colspan="8">10. Reason for the meeting (usually participation in research project)</td></tr>
<tr><td colspan="4">11. (Searching for) common acquaintances
Talking about people both might know</td><td colspan="4">12. (Searching for) common interests /experiences (outside of occupation/studies)
Talking about similar interests/experiences, e.g. hobbies, travel experiences</td></tr>
<tr><td colspan="8">13. Individual acquaintances and interests/experiences other than occupation/studies
Talking about people only one person knows
Talking about the interests/experiences only one person had</td></tr>
<tr><td>IV.
Sweden and China</td><td colspan="3">14. Talking about Sweden (as a country)</td><td colspan="3">15. Talking about China (as a country)</td><td colspan="2">16. Comparing China and Sweden (as countries)</td></tr>
<tr><td>V.
Emotional support</td><td colspan="8">17. Emotional support</td></tr>
</table>

4.2 Topics in First Acquaintance

In total 17 topics have been identified. The topics are grouped in five types (I-V) and numbered (1-17). Some topics (4, 8 and 9) have sub-topics (a, b, c). The topics are presented in Table 1 together with a brief explanation.

The topics of Type I include greetings and the topics related to personal information. Type II includes occupation/studies and residence. Type III contains the purpose for meeting, common or non-common acquaintances, interests and other experiences (occupation/studies excluded). Talking about Sweden and China and comparing them are the topics in Type IV. Giving emotional support (e.g. consolation and encouragement) is included in Type V. In Table 2, the distribution of topics in relation to the type of encounter is presented.

Table 2. Distribution of topics in relation to the type of encounter

<table>
<tr><th colspan="4">Topics present in</th></tr>
<tr><th>the Chinese, Swedish and Swedish-Chinese interactions</th><th>the Swedish interactions</th><th>the Chinese interactions</th><th>the Swedish-Chinese interactions</th></tr>
<tr><td rowspan="5">1. Greeting
2. Name introduction
8. Occupation/studies ((a)current, b) previous and c) future)
9. Place of living (b. previous)
10. Reason for meeting
12. (Searching for) common interests/experiences
13. Individual acquaintances, interests/experiences</td><td>5. Religion
7. Money and economy
9. Place of living (a. current)</td><td>15. Talking about China
17. Emotional support</td><td rowspan="3"></td></tr>
<tr><th colspan="2">both the Chinese and the Swedish ineractions</th></tr>
<tr><td colspan="2">3. Age, 4. Family, 11. (Searching for) common acquaintances</td></tr>
<tr><td rowspan="2"></td><th colspan="2">both the Chinese and the Swedish-Chinese interactions</th></tr>
<tr><td colspan="2">6. Nationality, 14.Talking about Sweden 16. Comparing China and Sweden</td></tr>
</table>

The data is scarce which enables us to draw only tentative conclusions. As can be seen from the table, it is customary for the participants to greet each other and introduce themselves. Being quite neutral, such topics as occupation/studies, reasons for meeting, previous places of living, searching for common interests/experiences as well as talking about the people one knows and the interests/experiences appear in both mono- and intercultural conversations. Religion, money and economy as well as current place of living are discussed only by the Swedes, while China and emotional support are in focus in the Chinese interactions. Topics that require more shared background and sensitivity are discussed in the monocultural dyads, perhaps because the participants feel at ease with more topics. In line with this, topics like age, family and searching for common acquaintances are discussed in monocultural interactions only, but not in the intercultural interactions. Probably, it is easier to find a common acquaintance with a person from the same country than with a foreigner. Nationality and talking about Sweden and comparing China and Sweden are topics present in the Chinese-Chinese and intercultural encounters. The reason for this might be that all three topics are relevant when the participant is an outside visitor. Below, we will provide some more explanation to the distribution of topics and also comment on the frequency and order of occurrence of topics in the analyzed interactions.

In Table 3, the distribution of topics in order of their occurrence in the interactions for each dyad is presented.

Table 3. Distribution of topics in the encounters

	Type, conversation numbers and order of topics in interactions (order 1- 15)											
	Chinese-Chinese interactions				*Swedish-Swedish interactions*				*Swedish-Chinese interactions*			
	CC1	***CC2***	***CC3***	***CC4***	***SS5***	***SS6***	***SS7***	***SS8***	***SC9***	***SC10***	***SC11***	***SC12***
1. Greeting	1	1	1	1	1	1	1	1	1	1	1	1
2. Name introduction	2, 4	2, 5	2	2	2	2	2	2	2	2	2	2, 4
3. Age	-	-	9	-	13	-	5	-	-	-	-	-
4. Family	-	-	5, 8	-	10, 14	11	14	-	-	-	-	-
5. Religion	-	-	-	-	-	-	7	-	-	-	-	-
6. Nationality	5	-	3	-	-	-	-	-	4	-	-	3
7.Money and Economy	-	-	-	-	-	-	13	-	-	-	-	-
8.Occupation/studies												
a) current	3, 6	3, 6	6	3, 5	3	3	8	4,6,8,10	7	4, 6	4, 6	8
b) previous	9	7, 9	-	4	8	7	10	9	8	-	-	-
c) future	-	-	10, 12	-	-	4, 13	11	7	11	5	5, 7	
9. Place of living												
a) current	-	-	-	-	5	12	14	-	-	-	-	-
b) previous	8	8, 10	4	-	6, 9, 11, 15	5, 9, 10	3	-	3, 5	3	3, 8	5, 10
10. Reason for meeting	-	12	-	-	4	6, 8	-	3, 5	-	-	-	7
11.(Searching for) common acquaintances	13	4	-	-	12	-	-	-	-	-	-	-
12. (Searching for) common interests /experiences	7, 14	11, 14	-	6	7	-	4, 12	-	10	-	-	-
13. Individual acquaintances, interests/experiences	-	13	7	-	-	-	9, 6	11	6	7	-	6
14. Talking about Sweden	12	-	-	-	-	-	-	-	9	9	10	-
15. Talking about China	10	-	-	-	-	-	-	-	-	-	-	-
16. Comparing China Sweden	11	-	-	-	-	-	-	-	8	8	9	-
17. Emotional support	-	-	11	-	-	-	-	-	-	-	-	-

Some but not all topics are present in all interactions; some topics are taken up in several times (e.g. previous studies), and some occur only once (greeting and name introduction). For example, the topic **6. Nationality** is in the 5th place in CC1 (Chinese-Chinese conversation 1). If the topic is not present in the interaction, it is marked by (-). The largest number of topics is 16 (SS6) and the lowest is 6 (CC4).

Below, some observations are provided concerning frequency and the order of occurrence of the topics in the interactions. Examples are given illustrating the topics, taken from the interviews and the transcriptions.

4.3 Topics Present in the Chinese, Swedish and the Swedish-Chinese Interactions

Conversation opening – greeting and name introduction

As we can see in Tables 2 and 3, such topics as **greeting** and **name introduction** appear at the beginning of all interactions. The Swedes greet each other with an informal *hej/tjena* (hi/hello), the Chinese use *ni hao* (how are you) and *hi/hello, nice to meet you* are used in intercultural interactions. In the Swedish interactions the participants mention their first names only, e.g. *Robert, Simon, Hanna*, which is customary in Sweden and also a result of the informality of the situation. The Chinese introduce themselves by using full names first, but they call each other differently after they get known of each other, depending on how many words they have in the name. A general principle is that the Chinese use no more than two words to refer to the other person rather than first name or full name. For example, Gegen Tana has four words in her name, so the Chinese speakers will call her Tana after they get to know her. The participants also tend to comment about their names, e.g. *Erik is a Swedish name* and, try to find a resemblance between their names, e.g. *Gegentana and Hanna are similar in a way* and relate them to the names of the people they know (see Excerpt 1):

Excerpt 1. SS6

Speaker	**Transcription**	**Eng. translation**
$L:	< hej /[1] [lisa] >[2]	*< hi / [lisa] >*
@ < *gesture: shake hands* >		
$K:	[som] [3] min syster	*[as] my sister*

Thus, though scarcely knowing each other, the participants attempt to find something similar, starting with names (as in the example above) and continuing with common acquaintances and interests.

Career plans, place of residence and reason for meeting

It is not surprising that **current occupation/studies** is a common topic in all interactions, as the participants are students. Talking about studies e.g. *går du också här på universitetet* (do you go to this university as well), is common and it appears in the initial and/or middle part of all conversations. Unlike greeting and name introduction, the participants turn to current occupation/studies repeatedly during interactions, e.g. in SS8.

Past experiences, more specifically related to **previous place of living** and **previous occupation/studies** occur in 10 and 8 interactions respectively. It is possible to see that the topic **previous place of living** is discussed somewhat earlier in intercultural interactions (as early as on the 3rd place in three out of four interactions) than in the Chinese (4th place in one and 8th in two) and in the Swedish (3rd place in one and 5th – 15th place in two) interactions. A possible explanation for this is that in the intercultural encounters it is common that people come from different countries. Thus, inquiries about previous places of living are reasonable and might appear earlier in intercultural interactions compared to the monocultural ones.

The topic **previous occupation/studies** appear in the middle and closer to the end in the majority of Swedish and the Chinese interactions (4th place in CC4 and 7th- 10th place in other interactions). The topic is discussed in only one intercultural interaction (SC9). A possible explanation is that the focus in intercultural interactions is primarily on the current state of things and on the experiences of a new culture, rather than on personal history. It is possible that past experiences that are more "rooted" in one's background are more difficult to talk about compared to the present experiences which are more shared for the Chinese and the Swedes who are living in Sweden both in general, and when meeting a person for the first time in particular.

Future occupation/studies is another topic in the interactions, which can be expected since the participants' backgrounds (students) motivate talking about a future career with fellow students.

---

[1] /, //, /// - pauses

[2] < >, @ < > - comments about non-verbal behavior, comment on standard orthography, other actions

[3] [ ] - overlap

Participants discussing **reason for meeting (participation in the research project)** is observed mainly in the Swedish interactions. This can perhaps be explained by the greater familiarity of the Swedish students with the university environment.

Individual acquaintances and (searching) for common interests/experiences
Talking about people you know, your personal experiences and interests are other common topics in the interactions. In the excerpt below, the Swedish participants talk about a shared experience:

Excerpt 2. SS5

Speaker	Transcription	English translation
$M:	där har jag åkt skridskor många [gånger]	*I skated there many [times]*
$J:	[ja det har] ju jag // också gjort några gånger // vilket sammanträffande	*[yeah I did] it // as well a couple of times // what a coincidence*

The participants stressed the problems of finding something in common with people of different cultural backgrounds. A Swedish respondent claims that it might also be necessary to be more careful in the intercultural interactions:

> *"As a Swede you can discuss the things that we have experienced as we grew up… I cannot discuss them with a foreigner. I would be also more careful with a foreigner… one doesn't know what can be sensitive"*
>
> *Swedish female participant*

Though the data is scarce, it can be observed that the participants in general tend to attempt to talk more about shared interests/experiences than common acquaintances (6 compared to 3 interactions). Probably talking about people one knows is more sensitive than talking about interests and experiences? In addition, as mentioned above, (searching for) common acquaintances is present only in monocultural encounters as it might be difficult to think about common acquaintances in intercultural first encounters.

4.4 Topics Common for the Chinese and Swedish Encounters

Age, **family** and **(searching for) common acquaintances** are the topics discussed in monocultural interactions only. An interesting observation can be made concerning the Chinese and the Swedes discussing **family**. The Chinese students talked about their partners, while the Swedes focused on parents and siblings. Below, an excerpt from the Chinese interaction is provided:

Excerpt 3. CC3

Speaker	Transcription	English translation
$B:	ran hou ta men dou shi jia li mian yao qiu qu de er xi fu er a shen me de	*then do they have specific requirements to have son-in-laws or so*
$C:	dui ta men shi you zhe ge yao qiu dan shi wo men mei you wo men jia mei you	*yes they have specific requirements but we don't my family do not have*
$B:	ni men wu suo wei	*it does not matter to you*
$C:	dui < yao kan na ge kan >	*no < it is dependent on >*
@ < *gaze down to the left, hands open trying to find words* >		
$B:	chuan tong bu chuan tong ba	*whether the family is traditional or not*

Here, we can observe the discussion of a family issue that is more common for the Chinese than for the Swedes, namely the role of the family in choosing a marital partner.

4.5 Topics Present in the Swedish and in the Chinese Interactions Only

The current place of living is only discussed in the Swedish interactions. The Swedes (locals) tend to talk more in detail about Gothenburg. The reason might be that the Chinese (non-locals) lack a specific knowledge about the city, and that in the intercultural encounters the focus seems to lie on Sweden as a country. Below, the Swedish students discuss a "non-fancy" district (Bergsjön) and a "fancy" part (Majorna) in Gothenburg:

Excerpt 4. SS7

Speaker	Transcription	English translation
$P:	e va bo du i stan	*er where do you live in the city*
$M:	e ja bor i bergsjön	*er I live in < bergsjön >*
@ < *områdesnamn/district name* >		
$P:	okej // var	*okay // where*
$M:	ja: / < teleskopgatan > du	*hm: / < teleskopgatan > and you*
@ < *gatunamn/ street name* >		
$P:	i < majorna > i den [< fin+ >]	*in < majorna > in the [< fan+ >]*
@ < *områdesnamn/district name* >, @ < *cutoff: fina/fancy* >		
$M:	[fina stadsdelen]	*[fancy district]*

The participants ($P, $M) talk freely about the districts and share a common knowledge of the city. On the contrary, the excerpt below illustrates how the Swedish participant ($H) experiences difficulties in initiating interaction with the Chinese student ($B) about Vasastan, a locally famous part of Gothenburg:

Excerpt 5. SC12

Speaker	Transcription
$H:	you know vasastan
$B:	< no >
@ <*head: shake, disagreement* >	
$H:	< [you don't] >
@ < *facial expression: eyebrow: eyebrow raise; surprise*>	
$B:	[I don't think so]

Unlike in Excerpt 4, the Swedish speaker soon realizes that the Chinese interlocutor has no idea where this part of the city is, which surprises him. Thus, later in interaction, he attempts to provide some explanation.

Religion and **economy** appear in one Swedish interaction. The interactants comment briefly on going to church and study loans. It is interesting as money is a sensitive topics in many cultures, including Sweden. The fact that the participants discuss it can be explained by the context of interaction – talking about student issues at the university. One of the Swedish female respondents comments about it:

> *"Some things one should not start talking about … at the same time in some circumstance such things can be okay to talk about … if it is the reason why one start interacting because one is ill and is sitting at the waiting room at the hospital"*
>
> *Swedish female participant*

The Chinese **talk about China**, their home country. In addition, in one interaction, one of the Chinese girls provides emotional support, consolation, to the interlocutor, which might indicate some closeness in monocultural encounters.

4.6 Topics Common for the Chinese and the Intercultural Encounters

Nationality, **talking about Sweden** and **comparing China and Sweden** are the topics discussed in the Chinese-Chinese and in the intercultural interactions. An interesting example is presented below, when the Chinese and Swedish participants talk about the Swedish weather:

Excerpt 6. SC10

Speaker	Transcription
$D:	I like the weather here
$F:	< ok > interesting
@ < surprise, facial expression: general face: laughter; uncertainty / surprise >	
$D:	er // except winter

The Swedish respondent is surprised that the Chinese girl likes the Swedish rainy and cold weather. The Chinese girl comments on it in the interview:

"I wanted to tell why I like Swedish weather. She talked about some bad things about Sweden…"
Chinese female participant

Avoiding criticizing the host country might be one of typical features in intercultural encounters as well as a tendency for talking about more neutral topics than those in monocultural encounters.

5 Discussion

The paper provides an overview, analysis and comparison of the topics present in the monocultural and multicultural first-time encounters between Swedish and Chinese university students. Though the data is scarce, the analysis shows the following tendencies in the first-time encounters.

The topics discussed in the encounters for this study are related to the personal background of the participants; i.e. their career, residence, (searching for) common interests, acquaintances and reason for interaction. Moreover, issues related to the countries of origin of the participants (Sweden and China) are present. The analysis shows that there are both similarities and differences in the topics discussed between the mono- and intercultural encounters.

To start with the **similarities in topics used in the interactions**, they are primarily due to the activity influences. Greetings and introductions are used as a way to establish social recognition. As the purpose of this interaction is to get to know each other, name introduction is a starting point in this process. Talking about the career/studies is expected, as the participants are students. Moreover, a tendency to search for and/or talk about something in common can be observed, which is again a consequence of the fact that the participants attempt to get to know each other better. Focusing on finding similarities, rather than differences is a natural way to enhance interaction, to motivate the interlocutor to talk more about him/herself. Talking about the reason for interaction can be also expected. Concerning **differences**, one of the factors influencing the interaction is whether the participants are of the same or different background. The fact that the Swedes talk about such sensitive topics as money and religion, but not with the Chinese, and that the Chinese girl gives emotional support to another

Chinese girt, but not to any Swedish participant, might be due to less tension in monocultural encounters. Searching for common acquaintances is common for the monocultural interactions, as one can presume that there is a better chance to find the people both know; in intercultural encounters this tendency is not present at all. Another influencing factor is whether the participants are local inhabitants (the Swedes) or sojourners (the Chinese). The Swedes talk more in detail about the districts and streets in Gothenburg while the Chinese informants do not discuss it at all. This can be explained by the Swedes having better knowledge about the city compared to the Chinese who are new in Sweden. The participants in this study are Swedes and Chinese; thus, comparing and discussing China and Sweden are expected topics in the interactions between the Swedish and the Chinese students, but also in the Chinese-Chinese interactions since they are new in Sweden.

References

1. Berbyuk Lindström, N.: Intercultural communication in health care - Non-Swedish physicians in Sweden, in Department of Linguistics, University of Gothenburg, Gothenburg (2008)
2. Nivre, J., et al.: Göteborg transcription standard. V. 6.4. Department of Linguistics. Göteborg University, Gothenburg (2004)
3. Allwood, J.: An Activity Based Approach to Pragmatics. In: Bunt, H., Black, B. (eds.) pp. 47–80. John Benjamins, Amsterdam (2000)

Greek WordNet and Its Extension with Terms of the Computer Science Domain

Panagiotis Blitsas[1] and Maria Grigoriadou[2]

[1] I.P.G.S. in Basic and Applied Cognitive Science, National and Kapodistrian University of Athens, Ilissia, Panepistimiopolis, 15771, Athens, Greece
[2] Dept. of Informatics and Telecommunications, National and Kapodistrian University of Athens, Ilissia, Panepistimiopolis, 15784, Athens, Greece
{pblitsas,gregor}@di.uoa.gr

Abstract. In the present work, the extension of the Greek WordNet database in the Computer Science domain is presented. This extension is based on the Baudet & Denhière text comprehension model, so that, the Greek WordNet can describe in a more efficient way: (1) the static states of a system's units, as well as the relations among terms expressing technical entities within the dictionary, and (2) more relations referring to events that happen during the operations of these system's units, as well as to goals that these operations and/or units have to achieve. The extension of the Greek WordNet was implemented with the contribution of VisDic Editor for the enrichment of a semantic database.

Keywords: macrostructure, microstructure, Semandix, text comprehension model, VisDic Editor, WordNet.

1 Introduction

The present work aims to extend the Greek WordNet with terms related to the Computer Science (CS) area, especially in the Computer Memory domain, as well as the semantic relations among the terms. It is observed that in the existing Greek WordNet, several terms used in the CS domain are also widely used in the general vocabulary in the Greek language. These terms have a totally different interpretation from that, which they are supposed to have in the CS area. Thus, their use in such a domain is not possible. Additionally, the English WordNet does not support some semantic relations that are needed for the fuller description of a technical system. These relations define non-standard features, such as attributes, operations, goals or intentions etc. It is necessary though to examine the depth of the features that we want to give to each term, determining both the quantity and quality of the dictionary extension. The quantity gives the advantage of designing new semantic statistical analyses, as well as the opportunity of defining new relations among the already listed terms, while the quality is determined by the different ways, in which the computational dictionary could be used.

To provide all coverage of the CS terminology, to predict possible language-specific interpretations of CS terms and to cover the needs of a broad range of applications

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 22 – 31, 2011.

intended to a wide range of users, an analytical framework of semantic relations is used. The specifications presented here are not only restricted within the Greek language, but facilitate possible extensions to other languages.

2 WordNet and VisDic Editor

The first WordNet (Figure 1) (http://Wordnet.princeton.edu/) was implemented in Princeton and it was in English followed by other WordNets like BalkaNet and EuroWordNet. WordNet, commonly used tool in CS and Natural Language Processing (NLP), is a large English vocabulary database based on the concepts of words and semantic relations that exist among them. Adjectives, verbs, nouns and adverbs are grouped into sets of synonyms (synsets); each synset expresses something unique and is ontologically [1] linked to other synsets by lexical semantic relations. Under the support of the semantic relations, the meanings of words in a language can be linked each other, forming a network of concepts in WordNet. The tool has been developed in such way that facilitates the navigation of the semantically related words and entities.

WordNet contains and expresses the following types of semantic relations:

- Synonyms: X is a synonym of Y, namely that X could replace Y or the opposite, without any semantic effect, e.g. “catch” is a synonym of “fetch”.
- Hyponyms/hypernyms: X is a hyponym of Y / Y is a hypernym of X, namely that Y is a type of X, e.g. “dog” is a hyponym of “animal” (a dog is a kind of animal).
- Meronyms/holonyms: X is a meronym of Y / Y is an holonym of X, namely that the computer is part of X, e.g. the “bumper” is a meronym of “car” (the bumper is a part of the car).
- Familiarity: Y is a familiar term to X, if X and Y share a hyperonym, e.g. “wolf” is a familiar term of “dog” and “dog” is a familiar term of “wolf”, because they share the same hypernym “animal”.

The EuroWordNet project has produced WordNets for various European languages (http://www.illc.uva.nl/EuroWordNet/). This program has produced BalkaNet WordNets for six languages: Bulgarian, Czech, Greek, Romanian, Turkish and Serbian. These WordNets are aligned to the Princeton WordNet, following the principles established by the EuroWordNet consortium. For the purposes of this project VisDic was developed, which is an XML-based authoring tool and was used in the present work.

The Greek WordNet was developed at the University of Patras by a group of linguists with the contribution of the National and Kapodistrian University of Athens. Some indicative statistics for the Greek WordNet are: 18.677 sets of synonyms, 24,811 different words/terms, 1.33 words/synonym Inter-Lingual and 18.649 definitions. The following tools were used for the Greek WordNet implementation: (a) tools for Linguistic Information Extraction from online dictionaries and text corpora, (b) VisDic Editor, which is a tool of semantic development of words and (c) the WMS (WordNet Management System).

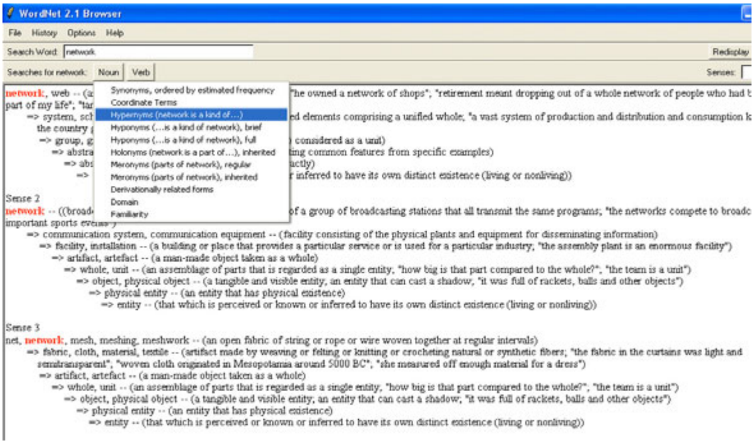


Fig. 1. WordNet example for searching the term "network"

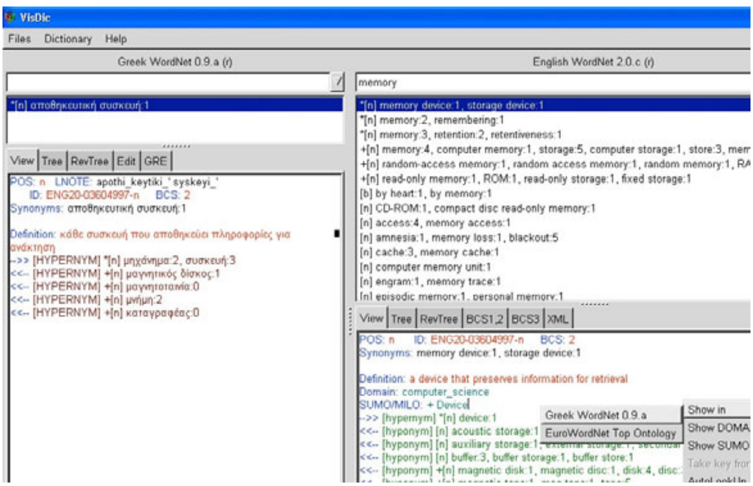


Fig. 2. VisDic Editor example for searching the term "memory"

VisDic Editor (http://nlp.fi.muni.cz/projects/visdic/) is a graphical application [2] for finding terms and their relations with other terms, and for modifying the dictionary database, which is stored in XML (eXtensible Markup Language) format; it is the main program for browsing the BalkaNet WordNets. Additionally, this editor enables the user to extract the relations among the entities/terms in XML file format that can be used by other applications. It is also possible to add an already constructed dictionary. This capability makes VisDic Editor flexible, in order the tool to be used in combination with other existing dictionaries. The main VisDic working window is shown in Figure 2. There is an open frame for each dictionary, which has a field placement and a list of records found during the search. The name of the dictionary is displayed on the top of each frame. In the lower section there is a status display area containing information.

3 Greek WordNet Extension

3.1 WordNet and Greek-Speaking User Requirements

It is observed from empirical data in the field of CS Education (Greek public schools: 1st Technical High School of Ioannina, 2nd Technical High School of Agrinio, 1st & 2nd General High Schools of Piraeus and 1st Technical High School of Piraeus, as well as the National University of Athens: Dept. of Informatics and Telecommunications) that Greek users tend to seek a cause-result relation in terms, commonly encountered in the morphological structure of Greek words. Specifically, it is observed that Greek students/users are more motivated to learn and comprehend new terms when they can detect and possibly question logical relations among them, rather than to directly incorporate them in their vocabulary. A research has pointed out that when Greek students learn new CS terms, they face problems (contradictions/conflicts) in understanding the causal relations between these terms [3]. In this case, and considering the popularity of CS studies in Greece and the reinforcement of CS classes in Greek public education, a more efficient way for describing the causal relations among CS terms is necessary. Consequently, a theory supporting structures, which describe this kind of relations, would be suitable for the extension of the Greek WordNet.

3.2 Theoretical Framework

A basic requirement for the extension of the Greek WordNet is the use of an appropriate theoretical framework allowing the consistent description of CS terminology, in order to predict possible language-specific interpretations of CS terms and to cover the needs of a broad range of applications for Greek and Balkanet users.

The method of the terms entry plays a particularly important role because an inconsistent and incomplete analysis could lead to cyclic references for both existing and new added terms. In this case, the semantic database of the Greek WordNet would contain false logical relations among the terms, and subsequently leads to misinterpretations of the semantic ontology. Thus, it is necessary to extend the Greek WordNet with more synsets and relations among them after careful analysis and under the prospect of merging it with other similar extensions within the CS domain [4]. The extension of the Greek WordNet, in respect to the relations among the terms, is based to the Baudet & Denhière text comprehension model [5].

The above-mentioned model supports that a person reading a text builds gradually its Microstructure and Macrostructure. The Microstructure involves the descriptions of the states of the system's units, the sequence of events performed on these units and the causes that change their static states, as well as the static states of the entire system. Specifically, the Microstructure is constituted of two other structures: the Relational and the Transformational Structure. The Relational Structure contains simple or complex static states. The hierarchy is determined by taxonomical (hyponyms/hypernyms) and meronymic (meronyms/holonyms) relations. The Transformational Structure represents the sequence of the events that change the system's state and the causal or/and temporal relations that are shown among these events. The causal relations are expressed by the causality among events, i.e. when an event causes another. These relations are temporal too, since the cause always precedes the result. The temporal relations reflect the temporal sequence of events. It should be noted that the temporal sequence of events is not necessarily causal. This means that an event may precede another one without causing it. This structure is transformative through Macroevents, which describe the general event leading to an event goal. A Macroevent is a general term covering all the events of the specific operation. The Macrostructure, except for the Microstructure, includes the Teleological Structure of the goals and the subgoals of the various system operations. Specifically, the Teleological structure is constructed upon the Microstructure and firmly interrelated to it, and represents the structures and behaviors of natural or artificial being ("organisms") as a whole set of functional modules. The hierarchical form of the Teleological structure is defined by the meronymic relations, which are organized as a tree of goals or intentions (in case of human being).

3.3 Greek WordNet Extension Procedure

The implementation of the extension of the Greek WordNet was performed in four stages, namely "Comprehension", "Comparison", "Modeling & Development" and "Extension". At the "Comprehension" stage, text abstracts from the book "Computer Science: An Overview" [6] were used, read and comprehended. The terms, the definitions and the relations, which were about to be used, were identified. During searching the terms - definitions within the texts, the need to find new terms that did not existed within them was revealed. At the "Comparison" stage, the "Dictionary of Common Greek" issued by the Modern Greek Studies Institute of the Aristotle University of Thessaloniki was used, as an additional source, in order to get the missing terms of the texts and cover in this way the "semantic gaps".

Table 1. New relation types

Relation Types	Examples	Model Structures
has_attribute	main memory *has_attribute* speed	relational
has_operation	computer *has_operation* data storage	transformational
causes	rotation pending *causes* block transfer	transformational
caused_by	block transfer is *caused_by* rotation pending	transformational
precedes	searching *precedes* rotation pending	transformational
comes_after	rotation pending *comes_after* searching	transformational
mero_event	block transfer is a *mero_event* of data access	transformational
holo_event	data access is a *holo_event* of block transfer	transformational
intends_to	hard disk *intends_to* permanent storage	teleological

At the "Modeling and Development" stage, the new and necessary relation types based on the text comprehension model were manually added. Adding relations into the Greek WordNet was realized through the VisDic Editor. In detail, the XML file that contains the final extended relations and a new directory gre_db were created in the path of VisDic software. The files of the directory contained in the Greek WordNet were copied to the new directory. Additionally, the new extended .inf file, which contains the new relation sets of synonyms, has been copied. The new records/rows were added by the name of each new relation type under the default relations. Examples of the new relation types added while extending the Greek WordNet are shown in Table 1. Finally, at the "Extension" stage, the Greek WordNet has been manually extended with new terms concerning the Computer Science domain. The entry has been made through the VisDic Editor interface. The words/terms not existing in the Greek WordNet were registered. The entire director of the Greek WordNet dictionary has been attached, and the .inf file has been modified, in order for the new relations and terms to be defined in VisDic Editor. From the 128 synsets of the present work, 75 are new, while the rest 53 have been modified. The definitions of all synsets are new. Finally, 745 relations among the synsets were imported and every synset has an average of 5.82 relations with the other synsets.

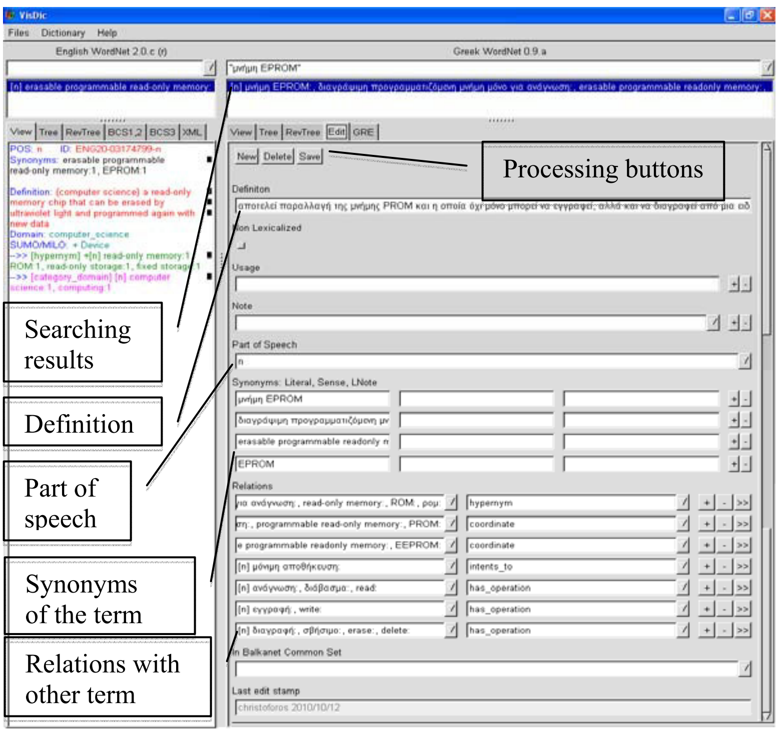


Fig. 3. Adding new terms on VisDic Editor

Figure 3 shows a screenshot of the Visdic editing environment, which explains its fields. In the field of word searching in the Greek WordNet, the new term is being added by selecting the option "New", and associated with other words/terms. In the Definition field, a definition of the same word is added. In the field "Part Of Speech", the appropriate part of speech of the new added word is being recorded (e.g. verb, adjective, noun). In the field "Synonyms", the word/term is being added and its synonyms. Finally, in the field “Relations”, the associations of the added terms with other terms are being declared.

Figure 4 shows the result of the search entry “σκληρός δίσκος (hard disc)” of the Computer Memory domain, under VisDic Editor “View tab”. Example 1 describes the teleological structure relation “σκληρός δίσκος (hard disc) – intends_to – μόνιμη αποθήκευση (permanent storing)”. Example 2 describes the transformational structure relation “σκληρός δίσκος (hard disc) – has_operation – εγγραφή (write)”. Example 3 describes the relational structure relation “σκληρός δίσκος (hard disc) – has_attribute - χωρητικότητα (capacity)”.

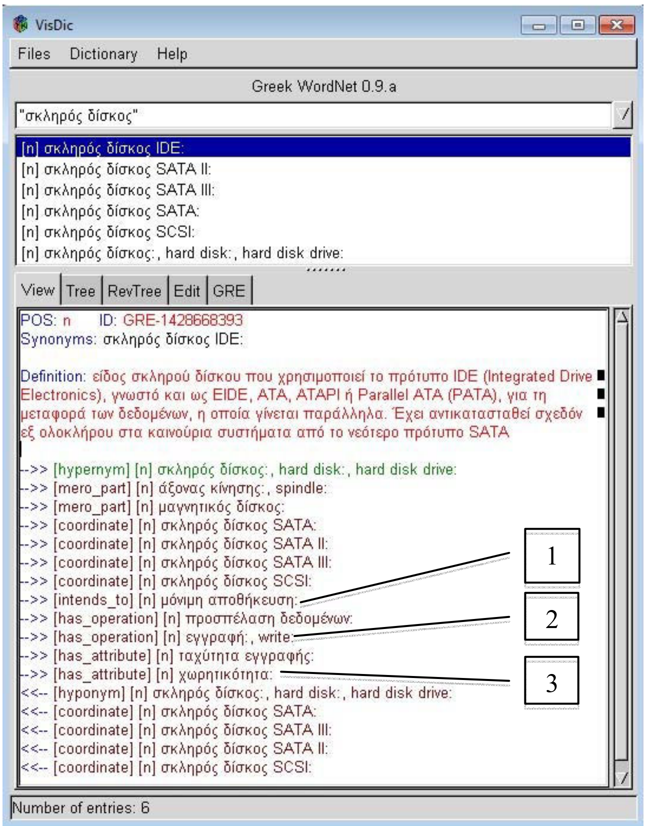


Fig. 4. Search entry “σκληρός δίσκος (hard disc)”

Figure 5 shows the result of the search entry “προσπέλαση δεδομένων (data access)”, a term derived from the Computer Memory domain, under VisDic Editor “Tree tab”. Example 1 describes the transformational structure relation “σκληρός δίσκος SCSI (hard disc SCSI) – has_operation - προσπέλαση δεδομένων (data access)”. Example 2 describes the transformational structure meronymical relation “αναζήτηση (searching) – is holo_event - προσπέλαση δεδομένων (data access)”. Example 3 describes the transformational structure relation “μεταφορά μπλοκ (block transfer) – comes_after & caused_by – αναζήτηση (searching)”.

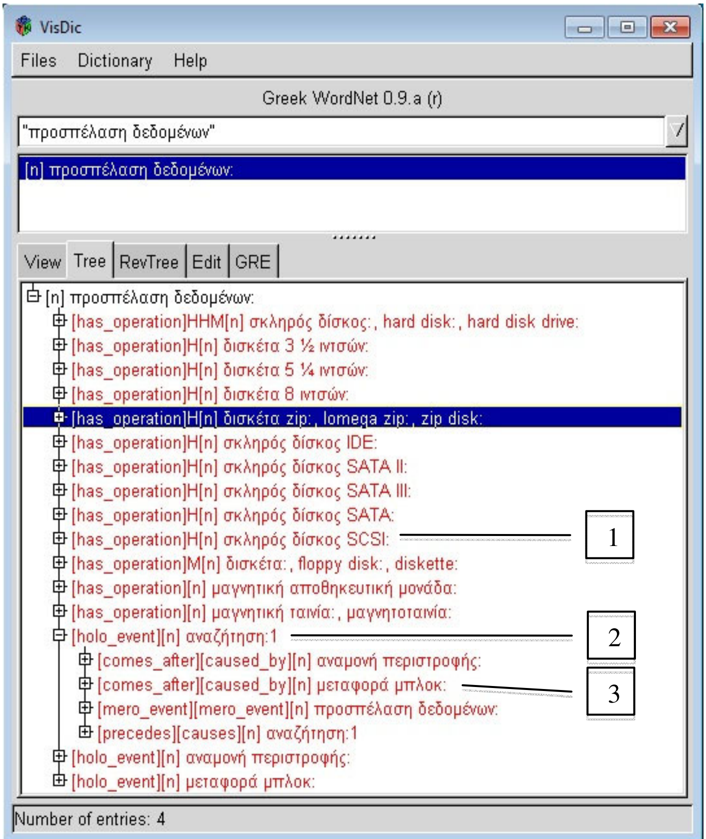


Fig. 5. Search entry "προσπέλαση δεδομένων (data access)"

4 Discussion

In many cases, during the extension of the Greek WordNet with new relations, the production of circular paths within the graph of the terms associations was observed. This kind of semantic cyclic reference would be considered as a semantic equivalent to the phenomenon of the cyclic definitions [7].

After elaborating the Greek WordNet through VisDic Editor we could propose some improvements concerning functions and usability, as well as some general suggestions concerning the semantic analysis of the conceptual content for extending the computational dictionary. We noticed that there is a need of:

– Defining more categories under the concept of "meronymy". For example, there is a different way of «inheritance» when X just extends the concept Y, from the situation when X provokes a behavioral change to Y
– A clear discrimination between terms expressing abstract and physical entities

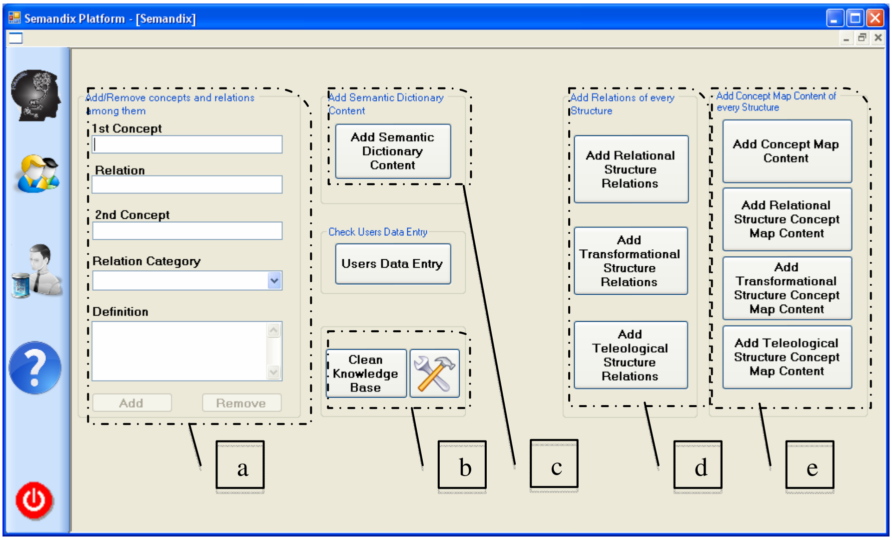


Fig. 6. Semandix enrichment administration

– A friendlier and more configurable extension tool for extending Greek WordNet, especially with terms of the Computer Science domain

One of the basic purposes of the extension of the Greek WordNet and its adjustment, according to the Baudet & Denhière text comprehension model was the need of using the XML file of the extended WordNet to a semantic tool called Semandix [8]. Semandix (Seman-tic dix-ionary) is a tool constructing a knowledge base by using, as a basis, the Baudet & Denhière text comprehension model too (Figure 7). Semandix allows the investigation of concepts and relations among them within a free text. Its knowledge base can be enriched with content of concept maps [9] and WordNets by XML format files extracted from CMapTools and VisDic editor, respectively. On the Semandix administration tab there are the following options:

– Adding individual terms (concepts) and relation with another terms (Figure 6a)
– Cleaning the whole knowledge base (Figure 6b)
– Enriching the knowledge base with WordNet content by adding XML format files extracted by VisDic (Figure 6c)
– Parallel enrichment of the system knowledge base with relations referring to each structure of the text comprehension model (Figure 6d)
– Content enrichment of the knowledge base with concept maps by adding XML format files extracted from CMapTools (Figure 6e)

We used the extended Greek WordNet to enrich Semandix by “Add Semantic Dictionary Content” (option in figure 6c). Our work on Semandix is being continued, in order for it to reach its ultimate goal, which is to explore alternative conceptions appearing within free text responses, according to the Baudet and Denhière text comprehension model [10].

Acknowledgments. The work described in this paper was partially supported by a grant from the project “Implementation of the Trainers’ Education” of the act “Training Teachers on ICT in Education" of the operational program "Education and Lifelong Learning”, which is funded by the Greek Government and the European Union (European Social Fund).

References

1. Gruber, T.: Toward Principles for the Design of Ontologies Used for Knowledge Sharing. Technical Report KSL 93-04. Stanford University, Knowledge Systems Laboratory. Revision (1993)
2. Horák, A., Smrž, P.: VisDic-WordNet Browsing and Editing Tool. In: Proceedings of the 2nd International WordNet Conference (GWC 2004), pp. 136–141. Masaryk University, Brno (2004) ISBN 80-210-3302-9
3. Tsaganou, G., Grigoriadou, M.: Authoring with ReTuDiSAuth for Adaptive Learning from Text. The International Journal of Learning 16(10), 1–10 (2009) ISSN 1447-9494
4. Kremizis, A., Konstantinidi, I., Papadaki, M., Keramidas, G., Grigoriadou, M.: Greek WordNet extension in the domain of Psychology and Computer Science. In: Proceedings of the 8th Hellenic European Research Computer Mathematics and its Applications Conference (HERCMA 2007). Economical University, Athens (2007), http://www.aueb.gr/pympe/hercma/proceedings2007/
5. Baudet, S., Denhière, G.: Lecture, compréhension de texte et science cognitive. Presses Universitaires de France, Paris (1992)
6. Brookshear, G.: Computer Science: An Overview, 9th edn. Pearson Addison Wesley (2006)
7. Namjoshi, K., Kurshan, R.: Efficient Analysis of Cyclic Definitions. In: Halbwachs, N., Peled, D.A. (eds.) CAV 1999. LNCS, vol. 1633, pp. 394–405. Springer, Heidelberg (1999)
8. Blitsas, P., Grigoriadou, M., Mitsis, C.: Constructing a Knowledge Base according to a Text Comprehension Model. In: Ifenthaler, D., Isaias, P. (eds.) Multiple Perspectives on Problem Solving and Learning in the Digital Age, 1st edn., pp. 67–88 (2010) ISBN: 978-1-4419-7611-6
9. Blitsas, P., Papadopoulos, G., Grigoriadou, M.: How Concept Mapping Can Support Technical Systems Understanding Based on Denhière-Baudet Text Comprehension Model. In: Proceedings of the 9th IEEE International Conference on Advanced Learning Technologies (ICALT 2009), Riga, Latvia, July 14-18, pp. 352–354 (2009)
10. Blitsas, P., Grigoriadou, M.: Towards A Knowledge-Based Free-Text Response Assessment System. In: Proceedings of the IADIS International Conference in Cognition and Exploratory Learning in Digital Age (CELDA 2008), Freiburg, Germany, October 13-15, pp. 37–44 (2008)

An Experimental Study of the Use of Multiple Humanoid Robots as a Social Communication Medium

Kotaro Hayashi[1,2], Takayuki Kanda[2], Hiroshi Ishiguro[2,3],
Tsukasa Ogasawara[1], and Norihiro Hagita[1,2]

[1] NAIST, Nara, Japan
[2] ATR Intelligent Robotics and Communication Labs, Kyoto, Japan
[3] Osaka University, Osaka, Japan
Kotaro-h@is.naist.jp, kanda@atr.jp,
ishiguro@ams.eng.osaka-u.ac.jp,
ogasawar@is.naist.jp, hagita@atr.jp

Abstract. This paper reports on an experimental investigation into the use of humanoid robots as a communication medium. Many social robots have been developed and tried for use in urban environments, but due to their limited perception, their degree of interactivity is still far poorer than that of humans. In this study, our approach used the robots as a non-interactive medium. We propose using robots as a passive-social medium, in which multiple robots converse with each other.

1 Introduction

Over the past several years, many social robots that work in daily environments have been developed such as assistive robots for elderly people (Heerink et al., 2008; Scopelliti et al., 2005; Dario et al., 2001). Researchers have gone further by exploring communicative tasks that mainly involve interaction with people, such as a shopping assistant (Gross et al., 2008), collecting information in a city (Weiss et al., 2008), and delivering snacks at an office (Lee et al., 2009). Robots used for such research can typically make sophisticated human-like expressions. Through their human-like bodies and gestures, such robots are suited to communicate with humans to the extent that humans in these situations sometimes unconsciously behave as if they were communicating with peers. Such abilities could allow robots to perform tasks in human society involving communication such as guiding people along a route (Sakamoto et al., 2005).

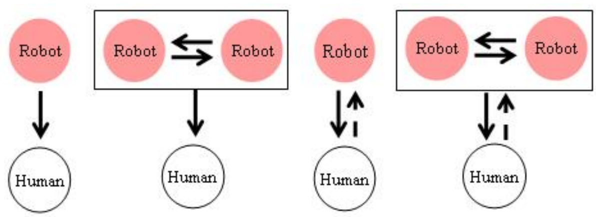
On the other hand, the interaction capability of robots is still under development and they have limited sensing capability. These shortcomings are particularly noticeable when we introduce robots into our daily lives. Although the appearance of a humanoid robot often makes people believe that it is capable of human-like communication, it cannot currently engage in such sophisticated communication. The pursuit of more capable sensing and recognition remains at the forefront of robotics research. The results of such research should eventually be integrated into robots so that they can behave as ideal interaction partners that are capable of human-like communication. Pioneering research work in human-robot interaction (HRI) has revealed what robots can accomplish, such as museum guidance (Siegwart et al.,

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 32 – 41, 2011.

2003; Shiomi et al., 2006), perspective-taking (Trafton et al., 2006), operation support (Breazeal et al., 2005), behaving as a well-mannered servant, and support for language study (Kanda et al., 2004). However, a robot's ability to inform humans is still quite limited.

On the other hand, recent research in HCI (human-computer interaction) has highlighted the importance of robots as a new interface medium. Reaves and Nass researched the role of computers as a new interface medium in the manner of previous media, such as television and radio, and they proved that humans act toward computer interfaces (even a simple text-based interface) as if they were communicating with other humans (Reeves and Nass, 1996). Cassell et al. demonstrated the importance of anthropomorphic expressions, such as arms and heads on embodied agents, for effective communication with humans (Cassell et al., 1999). Cory and Cynthia compared a robot and a computer-graphic agent and found that the robot was suitable for communication about real-world objects (Kidd et al., 2004).

Users are under no obligation to take part in the conversation. In other words, borrowing terms from Clark's linguistic literature (Clark, 1996), the user is placed in a bystander position (free from responsibility for the conversation), where robots are speaking and listening as participants (responsible for it). This characteristic makes it similar to previous media: people do not have to respond to what the medium (the person in the medium) says, or to return a greeting when the medium greets, or feel uncomfortable about leaving in front of the medium.



(a) passive (b) passive-social (c) interactive (d) interactive-social

Fig. 1. Robot(s) as medium

Fig. 1 shows the difference of this type of medium compared to other forms of human-robot interaction. At times robots have been used for merely presenting information to people, which we named a passive medium. This is the same as a news program on TV where one announcer reads the news. On the other hand, many researchers have been struggling to realize robots that act as an interactive medium, which intends to accept requests from people as well as present information to people. The robot-conversation-type medium, on which we focus in this paper, is named a passive-social medium. It does not accept requests from people, that is as same as the passive medium; however, it does present more information than the passive medium, through its social ability: expression of conversation. It is rather similar to a news program on TV where announcers discuss comments spoken by others.

We are exploring an alternate approach to maximizing the information that a robot system can offer to people, focusing on attracting ordinary people's interest to the information. This new strategy is based on letting users observe a conversation between robots. For example, Kanda et al. proved that users understand a robot's speech more easily and more actively respond to it after observing a conversation between two robots (Kanda et al., 2002). We named this kind of medium the "passive-social medium." Figure 1 illustrates the difference between this medium and other forms of human-robot interaction.

In this paper, we focus on the robot-conversation-type medium, a passive-social medium (Fig. 1, b). In this case, the robots do not accept requests from people, similar to a passive medium, but this case still attracts people's interest to information more than does a passive medium through its social ability, i.e. the expression of conversation. We believe that a "passive-social medium" is a more natural way to offer information to people than a simple passive medium. This is similar to a news program on television where announcers discuss comments told by others. Figure 1, d shows what we call an interactive-social medium, but such a medium has a weakness in its interactivity, just as in the case of a conventional interactive robot medium.

2 Multi-robot Communication System

For Passive-Social medium, we developed a system consisted of a sensor and humanoid robot(s).A scenario-controlling system, described below, controlled the robots' behavior. The robots' behavior was written in a simple scripting language that is easy to prepare.

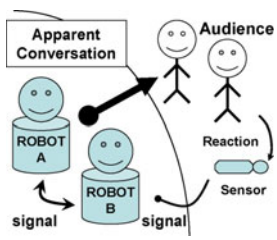


Fig. 2. Schematic of multi-robot communication

2.1 Design Policy

This system implements social expression capabilities and interactivity that perform reliably in a real environment. The social expression capability is based on a system we had developed that allows precise control of conversation timing and easy development. Regarding the interactivity, we limited it to be very simple but robust. The system immediately responds when a person comes close to the robot(s) by making the robot bow to the person. This limited-realistic interactivity is accomplished with a laser range-finder placed in front of the robot. We did not use any other sensors such as audible or visible, because outputs from such sensors are uncertain in a real environment. Thus, what we refer to as "limited-realistic interactivity" is very different from that in some interactive robots, such as Robovie (Kanda et al., 2004) where people may enjoy the unpredictability of unstable sensing capability. We decided on this implementation because unstable interactivity does not work when the purpose is to inform people. Users would be frustrated if they could not retrieve the information they needed.

2.2 Humanoid Robot

We used the humanoid robot Robovie for this system. “Robovie” is an interactive humanoid robot characterized by its human-like physical expressions and its various sensors. We used humanoid robots because a human-like body is useful in naturally catching people’s attention. The human-like body consists of a body, a head, a pair of eyes, and two arms. When combined, these parts can generate the complex body movements required for communication. Its height is 120 cm and its diameter is 40 cm. The robot has two 4x2 degrees of freedom in its arms, 3 degrees of freedom in its head, and a mobile platform. It can synthesize and produce a voice through a speaker.

2.3 Sensor

To sense a human approaching a robot, we used a laser range-finder or sound level meter.

The laser range-finder that we used is the LMS200 made by SICK. This sensor can scan 180˚ degrees horizontally, and it measures this range within a 5-m distance with a minimum resolution limit of 0.25˚ and 10 mm.The output is used to make the robots "look at" or turn their heads toward the human passing by them.

The sound-level meter continually receives the volume data [dB] from it.

2.4 Scenario-Controlling System

The system is based on our scripting language for multiple robots, with a new function for changing the scenario that the robots act out when a human presence is detected. The scripting language has adequate capabilities for describing multiple robot communication and is simple enough for a developer to easily use it to control the robots’ behavior.

In this system, a set of robots interprets script files and executes scripts written in this language. One robot becomes the master. The master robot receives signals from the sensor, decides which scenario to execute, and informs its partner robot about it.

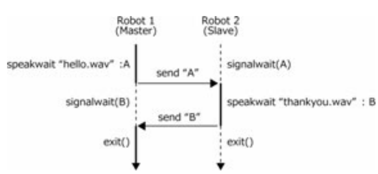


Fig. 3. Example of signal exchanges

Figure 3 shows an example of the scripting language. In Fig. 3, after Robot1 finishes playing the voice file “Hello.wav,” it sends signal “A” to Robot2. At that time, Robot2 is waiting for signal “A.” After receiving signal “A,” Robot2 plays the “thankyou.wav” file and sends signal “B.” When Robot1 receives it, this scenario is finished.

Figure 4 shows how the system works in the interactive condition. If Robot1 (master robot) notices there is a person nearby, it decides which scenario to execute and sends the corresponding signal to its partner. When there is no human around, the robots play the idling scenario.

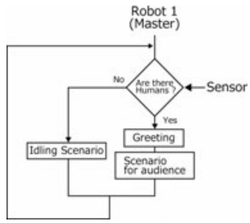


Fig. 4. Example of restrictive reaction

In this way, we constructed a system that is capable of interpreting its environment and executing scenarios accordingly.

3 Experiments

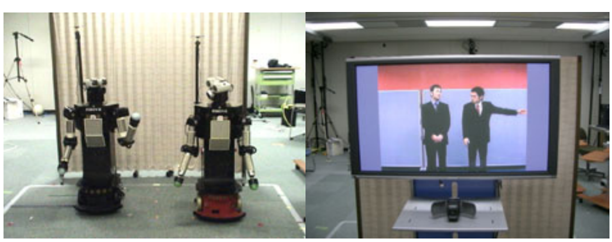
We conducted an experiment to investigate the utility of robots as a passive-social medium in this application by comparing it with television in a field experiment at a train station for eight days to investigate the effects of a passive-social medium.

3.1 Experiment Comparing it with Television

- **Method.** Each subject watched a performance by either robots (Fig. 5, a) or humans (Fig. 5, b) in front of them (Fig. 6):

Robot condition. Two robots performed in front of the subject. Please refer to the movie in the proceedings CD for a scene of the robot performance.

Human condition. Two amateurs, who have practiced performance for three years, performed on a TV screen in front of the subject.



(a) Robot condition (b) Human condition

Fig. 5. Conditions

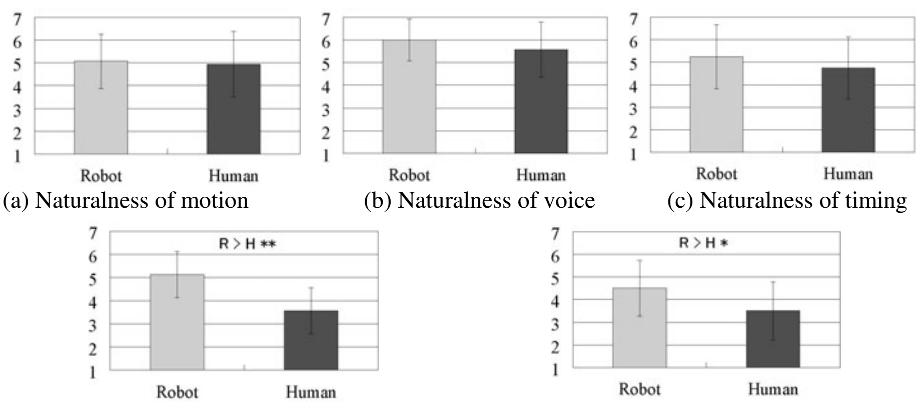
- **Measurement.** Once the experiment was finished, we distributed questionnaires to the subjects who rated the performances on a 1-to-7 scale for each question where

7 is the most positive. There were five questions: naturalness of motion, naturalness of voice, naturalness of timing, presence, and overall impression.

- **Result**

Naturalness of motion, voice, and timing. Figure 6 a-c shows a comparison of questionnaire results for the naturalness scores. From the results of a one-way between-groups design ANOVA (analysis of variance), there was no significant difference between the robot performance and the human performance in the average scores for naturalness.

Presence and overall impression. A one-way between-groups design ANOVA did show a significant difference between the robot performance and the human performance in the average scores when we consider presence and overall impression (Presence: $F(1,31) = 18.49$, $p<.01$; Overall impression: $F(1,31) = 5.71$, $p<.05$). That is, the performance by two robots had higher scores for presence and overall impressions than the performance by humans shown in the video. (Figure d,e)



(a) Naturalness of motion (b) Naturalness of voice (c) Naturalness of timing

(d) Presence (e) Overall impression

Fig. 6. Result of experiment comparing it with television

3.2 Experiment in the Real Environment

- **Method**

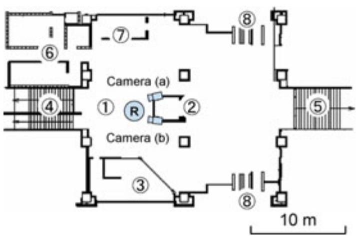
Gakken Nara-Tomigaoka Station was opened in March 2006 as the terminal station of the Keihanna New Line, belonging to the Kintetsu Railway. The Keihanna New Line connects residential districts with the center of Osaka (Fig. 7 b). Station users are mainly commuters and students. There are usually four trains per hour, but in the morning and evening rush hours there are seven trains per hour. Figure 8 shows the experiment's environment. Most users go down the stairs from the platform after they exit a train. We set the robot(s) in front of the left stairway (Fig. 8). The robot(s) announced information toward users mainly coming from the left stairway.

We observed how the users reacted to the behaviors of the robot(s). For this observation, we set cameras on the ceiling nearby (Fig. 8, cameras (a) (b)).

All station users who passed by the robot(s) were assumed to be participants. Their behavior was observed by video. We requested that users who stopped to watch the robot(s) answer a voluntary questionnaire. We obtained permission to record video

(a) Robots as medium in station (b) View of Station

Fig. 7. Scenes of the experiment



1 : Experimental field
2 : Elevator
3 : Shop
4 : Left stairway
5 : Right stairway
6 : Toilets
7 : Vending machines
8 : Ticket gates

Fig. 8. Station map

from the responsible authorities of the station, and a notice was displayed in the station about the video recording.

- **Mesurement**

We requested station users who stopped to watch the robot(s) to answer a questionnaire. We obtained answers from 163 station users. The questionnaire had three questions as follows in which they rated items on a scale of 1 to 7, where 7 is the most positive:

  - Feeling of being addressed by the robot
  - Interest in the content of the information the robot(s) is announcing
  - Enjoyment

- **Conditions**

Passive condition (P condition). In this condition, one humanoid robot was installed (Fig. 7(a)). The robot had a sensor in front of it, although the sensor was not used. The robot randomly played the five scenarios announcing station and travel information continuously.

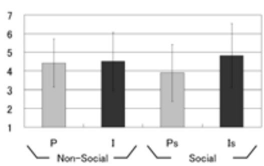
Interactive condition (I condition). One humanoid robot was installed as in the P condition, but the robot had limited-realistic interactivity. That is, it had a sensor (laser range-finder) in front of it and changed the scenario according to the position of the human. Concretely, if there was no person near the robot, the robot played the idling scenario. When the sensor detected a person within a semicircle of 3.5 meters, the robot stopped playing the idling scenario, looked at the person, bowed and said "Hello." After that, while one or more persons were within the range of 3.5 meters, the robot randomly played the five scenarios announcing station and travel information.

Passive-social condition (Ps condition). Two humanoid robots were installed (Fig. 7(a)). The robots had a sensor in front of them, but the sensor was not used. The robots randomly played the five scenarios announcing station and travel information by communicating with each other.

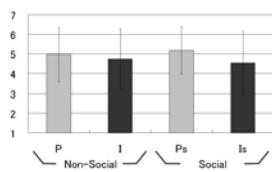
Interactive-social condition (Is condition). Two humanoid robots were installed as in the Ps condition. The robots had limited-realistic interactivity: the robots had an operating sensor in front of them and changed the scenario according to the position of the human. Concretely, if there was no person near the robots, the robots played the idling scenario. (In this scenario, robots chat with each other.) When the sensor detected a person within a 3.5m-radius semicircle, the robots stopped playing the idling scenario, looked in the direction of the person, bowed and said "Hello." After that, as long as one or more persons was within a range of 3.5 meters, the robots randomly played the scenarios announcing station and travel information by communicating with each other.

- **Summary of Results**

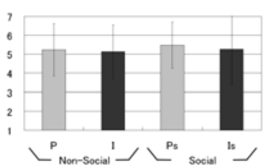
The results indicated that limited-realistic interactivity of the robot gives people the feeling of being addressed by the robot(s). On the other hand, it makes people lose interest in the information. From this result, we believe that using "interactive" as a medium does not necessarily provide a good result in its current form, since such performance has limited realistic use in a real environment. However, the non-social conditions had a lower chance of making people stop at the robot. These findings indicate that the passive-social medium is promising because the system has a better chance of getting people to stop and become interested in the information announced by the robot. (Fig9 a-d)



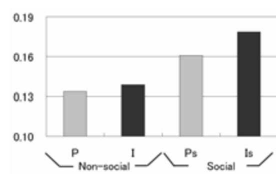
(a) Feeling of being addressed by the robot



(b) Interest in information robot(s) announcing



(c) Enjoyment



(d) Rate of "Stopping to watch"

Fig. 9. Results of experiment in the real environment

4 Conclusion

Although we have used robots as a passive-social medium, we have not truly investigated the effects of robots as a passive-social medium in comparison with other forms. In both trials, robots got people's attention so that they crowded around to see the robots. One of the difficulties has been that when people have a strong interest in robots, it is difficult to identify the effects of the passive-social medium because people appreciate any encounter with a robot due to its novelty.

The experimental results revealed that a two-robot condition (passive-social and interactive-social conditions) was better than a one-robot condition in terms of getting people to stop at the robots. Once people stopped, these conditions did not make any difference. Instead, a lack of interactivity (passive-social and passive conditions) had the advantage of attracting people's interest in the contents of the utterances. Thus, the passive-social condition proved to be the best for this purpose among the conditions tested in the experiment.

Although the experiment revealed the positive aspect of a passive-social medium on the "interest" aspect, it is not clear how naturally the passive-social medium offers information compared with other types of medium. The experimental results revealed effects when people glanced at the robot to decide whether to stop; however, the results did not reveal effects after stopping at the robots. The difficulty is in experimental control. In this experiment, we controlled the contents of the information that the robots said. Two robots (passive-social condition) could enable construction of a bigger variety of scenarios than is possible with a single robot. For example, one robot might ask a question to another, after which the other would make a response. Use of such a staged effect, however, could cause differences not only due to the conditions (passive-social vs. passive) but also due to the different contents of the utterances. Thus, we did not implement such techniques in this experiment. Probably adding such a feature would make robots more enjoyable and make interaction with people more natural. Demonstrating such effects will be one of our future studies.

References

1. Heerink, M., Kröse, B., Wielinga, B., Evers, V.: Enjoyment, Intention to Use and Actual Use of a Conversational Robot by Elderly People. In: ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI 2008), pp. 113–120. ACM press, Amsterdam (2008)
2. Scopelliti, M., Giuliani, M.V., Fornara, F.: Robots in a Domestic Setting: A Psychological Approach. Universal Access in the Information Society 4, 146–155 (2005)
3. Dario, P., Dario Guglielmelli, E., Laschi, C.: Humanoids and Personal Robots: Design and Experiments. Journal of Robotic Systems 18, 673–690 (2001)
4. Gross, H.M., Böhme, H.J., Schröter, C., Müller, S., König, A.C., Martin, M., Bley, A.: ShopBot: Progress in Developing an Interactive Mobile Shopping Assistant for Everyday Use. In: International Conference on Systems, Man and Cybernetics, pp. 3471–378 (2008)
5. Weiss, A., Bernhaupt, R., Tscheligi, M., Wollherr, D., Kuehnlenz, K., Buss, M.A.: Methodological variation for acceptance evaluation of human-robot interaction in public places. In: IEEE Int. Symposium on Robot and Human Interactive Communication (RO-MAN 2008), pp. 713–718. IEEE Press, Munich (2008)
6. Lee, M.K., Forlizzi, J., Rybski, P.E., Crabbe, F., Chung, W., Finkle, J., Glaser, E., Kiesler, S.: The Snackbot: Documenting the design of a robot for long-term human-robot interaction. In: ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI 2009), pp. 7–14. ACM Press, San Diego (2009)

7. Sakamoto, D., Kanda, T., Ono, T., Kamashima, M., Imai, M., Ishiguro, H.: Cooperative embodied communication emerged by interactive humanoid robots. International Journal of Human-Computer Studies 62, 247–265 (2005)
8. Scassellati, B.: Investigating Models of Social Development Using a Humanoid Robot. Biorobotics (2000)
9. Kozima, H., Vatikiotis-Bateson, E.: Communicative criteria for processing time/space-varying information. In: Proc. IEEE Int. Workshop on Robot and Human Communication (2001)
10. Breazeall, C., Scassellati, B.: A context-dependent attention system for a social robot. In: Proc. Int. Joint Conf. on Artificial Intelligence (IJCAI 1999), pp. 1146–1151. IJCAII Press, Stockholm (1999)
11. Siegwart, R., et al.: Robox at Expo.02: A Large Scale Installation of Personal Robots. Robotics and Autonomous Systems 42, 203–222 (2003)
12. Shiomi, M., Kanda, T., Ishiguro, H., Hagita, N.: Interactive Humanoid Robots for a Science Museum. In: ACM 1st Annual Conference on Human-Robot Interaction (HRI 2006), pp. 305–312. ACM Press, Salt Lake City (2006)
13. Trafton, J.G., Schultz, A.C., Perznowski, D., Bugajska, M.D., Adams, W., Cassimatis, N.L., Brock, D.P.: Children and robots learning to play hide and seek. In: ACM 1st Annual Conference on Human-Robot Interaction (HRI 2006), pp. 242–249. ACM Press, Salt Lake City (2006)
14. Breazeal, C., Kidd, C., Thomaz, A.L., Hoffman, G., Berlin, M.: Effects of Nonverbal Communication on Efficiency and Robustness in Human-Robot Teamwork. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robotis and Systems (IROS 2005), pp. 708–713. IEEE Press, Edmonton (2005)
15. Kanda, T., Hirano, T., Eaton, D., Ishiguro, H.: Interactive Robots as Social Partners and Peer Tutors for Children: A Field Trial. Human Computer Interaction 19, 61–84 (2004)
16. Reeves, B., Nass, C.: The media equation (1996)
17. Cassell, J., Bickmore, T., Billinghurst, M., Campbell, L., Chang, K., Vilhjalmsson, H., Yan, H.: Embodiment in Conversational Interfaces. In: Rea. Conf. on Human Factors in Computing Systems (CHI 1999), pp. 520–527. ACM Press, Pittsburgh (1999)
18. Kidd, C., Breazeal, C.: Effect of a Robot on User Perceptions. In: Int. Conf. on Intelligent Robots and Systems (IROS 2004), pp. 3559–3564. IEEE Press, Sendai International Center (2004)
19. Clark, H.H.: Using Language. Cambridge University Press, Cambridge (1996)
20. Kanda, T., Ishiguro, H., Ono, T., Imai, M., Mase, K.: Multi-robot Cooperation for Human-Robot Communication. In: IEEE Int. Workshop on Robot and Human Communication (ROMAN 2002), pp. 271–276. IEEE Press, Berlin (2002)
21. Kanda, T., Ishiguro, H., Imai, M., Ono, T.: Development and Evaluation of Interactive Humanoid Robots. Proceedings of the IEEE 11, 1839–1850 (2004)

A Multitasking Approach to Adaptive Spoken Dialogue Management

Tobias Heinroth, Dan Denich, and Wolfgang Minker

Institute of Information Technology
Ulm University, 89081 Ulm, Germany
{tobias.heinroth,dan.denich,wolfgang.minker}@uni-ulm.de

Abstract. Undoubtedly one of the key factors of a computed world, are the interfaces users ought to use. In this paper we present the adaptive spoken dialogue manager OwlSpeak to provide a spoken interface to a computed world, in our case to an Intelligent Environment. The most important feature of the dialogue manager is its ability to pause, resume, and to switch between more than one interactive task, which is a prerequisite to provide adaptive spoken dialogues. Especially within Intelligent Environments it is necessary to modify the status of an interface depending on the changing contexts of the environment and on the actual requirements the user may have. We present the implementation and evaluation of OwlSpeak as part of an existing Intelligent Environment that can be used by real subjects and show how multitasking can be utilised to cope with an adaptive speech interface.

Keywords: HCI, Intelligent Environments.

1 Introduction

Nowadays users have to deal with a variety of modalities that allow interaction with a computer. Computers and similar devices provide – besides keyboards and mice – visual, tactual, gesture-, and of course speech-based interfaces, to name but a few. The last-mentioned modality undoubtedly is the most natural communication medium for humans. The vision of talking naturally to a computer is still not realised. However, in several domains such as in-car tasks or telephony-based services Spoken Dialogue Systems (SDS) are getting more and more common. There are various requirements for SDS depending on the desired way of spoken interaction and on the handling of commands and controls the system provides. When using a mobile phone, for example, free text inputs such as negotiating or discussing are usually not necessary: a phone should merely understand commands like "Call Peter in the office!". However even such commands are not as simple as they seem to be. For example, the mobile phone shouldn't start calling Peter if someone says "… you might call Peter in the office and ask him…" during a conversation, for example. Even so such examples are apparently artificial; the necessity of an SDS used as interface for a mobile phone is definitely a matter under discussion. Its primary functionality – establishing calls and dialling – can usually be performed by using a keypad and a display. In general for most SDS scenarios a graphical interface could also be utilised.

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 42 – 51, 2011.

However, besides the issue of necessity in general there are several arguments that make the case for SDS: one of their main benefits is their advance in efficiency within many scenarios (especially related to Intelligent Environments). If a user has to control 12 lights, for example, a graphical interface would intuitively consist of 12 buttons. If the same task has to be handled by an SDS fewer commands would be necessary. The user would just ask the system to "switch on the light" and would not have to cope with a graphical control panel. Furthermore, in case of ambiguities the SDS could check back which light the user wants to switch on. Within the scientific area of Intelligent Environments (IEs) SDS technologies thus result in one of the most efficient and natural interfaces between humans and computer-based systems. In this context we refer to IEs as networks of various different components such as sensors, actuators, and processors that are able to automatically exchange information about themselves and their surrounding without any human intervention. Thus a user can provide input to *component A* that analyses the request and provides the new information to *component B* that might process the request and/or route the information to a further *component C*. Within the context of IEs, proactive behaviour (warning, information, etc.), and negotiative dialogues speech is a promising modality, for many tasks such as the mentioned command-and-control of devices or services. In particular for elderly people, for disabled persons or people with serious diseases who cannot stand up without further ado SDSs are extremely useful since such a system can provide a centralised and at once natural interface that can easily be accessed.

The remainder of this paper is structured as follows: The next section provides an overview on the scientific field of SDS and of the related work. Section 2 presents some use cases and their related requirements with respect to multitasking by speech. In Section 3 we provide details on the implemented prototype and show how multitasking over several disjoint dialogue domains can be realised. Section 5 shows the results of the evaluation of the proposed approach. The paper concludes and provides some future work in Section 6.

2 Related Work

An SDS is a computer-based system that enables a user to bilaterally communicate with a machine via spoken language. The three most important layers of an SDS are the acoustic front-end, the semantic layer, and the logical layer, which is constituted by the Spoken Dialogue Manager (SDM). A speech recognition module and a speech synthesis module constitute the acoustic front-end that is usually accessed by the user via microphone(s) and speaker(s).

One of the challenges regarding the realisation of real-life SDSs is their complexity. How much information must be taken into account in order to allow for a meaningful dialogue between the user and the system? How should this information be modelled in order to be both computer-readable and on the other hand easily assimilable into a probably on-going dialogue? Obviously the SDS and most notably the SDM must cope, depending on the scenario and the task to be tackled, with understanding and interpreting a maximum amount of information whilst keeping the complexity moderate. In order to realise adaptive and therefore intelligent spoken dialogue it is not only necessary to provide advanced voice recognition and speech synthesis but also to incorporate an adaptive SDM residing at the core of such an SDS. Nowadays one of the most widespread technologies to implement SDMs is the W3C standardized VoiceXML description language. The idea behind this approach is to simplify the development of dialogues by providing a model description of the dialogue to be expressed. However, VoiceXML is not able to persistently store and therefore describe a specific state of a dialogue. Thus dialogue strategies that need functionalities such as pausing and resuming of parallel tasks can only be implemented difficultly.

Within the scientific field of IEs several international research projects have been concerned with spoken dialogue interaction for quite a long time [7, 3]. Three directions can be recognized that have been discussed in the recent past: heavyweight rule-based frameworks such as the TrindiKit [5], agent-based systems such as RavenClaw [1] and statistical approaches such as the Bayes Net Prototype implemented within the TALK Project [10]. The two former approaches require strong assumptions regarding the set-up and adjustment. In RavenClaw a dedicated agent would be needed for each task. This agent has to match the assumptions of the language recognition and logic-processing components of RavenClaw, hence it is very difficult to create agents that can cooperate with various other agents. In TrindiKit the dialogue flow is indirectly described by a rule-base. These rules have to be kept in a coherent state – if a new rule affects older ones the complete rule set must be updated. Once the rule-base is implemented the system performs well but the more complex the dialogue, the more complex the rule-base. All statistical approaches rely on the availability of training data, which appears to be a significant disadvantage as it is costly to collect corpora and to train the statistical models. Furthermore all these approaches are domain-dependent and it does not seem to be trivial to introduce new dialogues that may handle new domains.

3 Use Cases and Requirements

In the following we present a short example to describe our approach. We assume that in a fictive world where the user lives together with his IE a typical situation is the arrival at home. Besides a “greeting” task there are several other interactive tasks running in parallel – each providing the possibility of spoken interaction. Since one of the main duties an IE should handle is the control of tasks, i.e., the system should provide possibilities to facilitate the user's access to various functionalities, it is necessary to provide a (probably varying) set of spoken commands the system can interpret and execute. An example for such behaviour could be a user, after entering the apartment, telling the system to switch the lights on. Fig. 1 shows a set of three interactive tasks that may form an exemplary interactive situation.

Since the SDM adapts to the context it needs to be able to receive triggers from the outside world to change its state. The initial phase therefore is triggered by a “user enters room” event. This event might happen only once a day and/or when the user has left the room for a specified period depending on the configuration of the IE. In our example the SDM is set up to wait until the user greets the system (Task 1). It further activates a control task that listens to possible lighting control commands the user may utter (Task 3). Initial system studies in the iSpace at the University of Essex revealed that the subjects preferred the SDS to be as unobtrusive as possible [4]. Thus we have designed the system to behave rather passively and not to proactively initialise a conversion if this can be avoided. By default, a control task such as Task 3 waits for user input and therefore behaves passively. However if the user initialises talking to the system by uttering a spoken command the system could take this opportunity to start dialogues that otherwise would have to be proactively initiated.

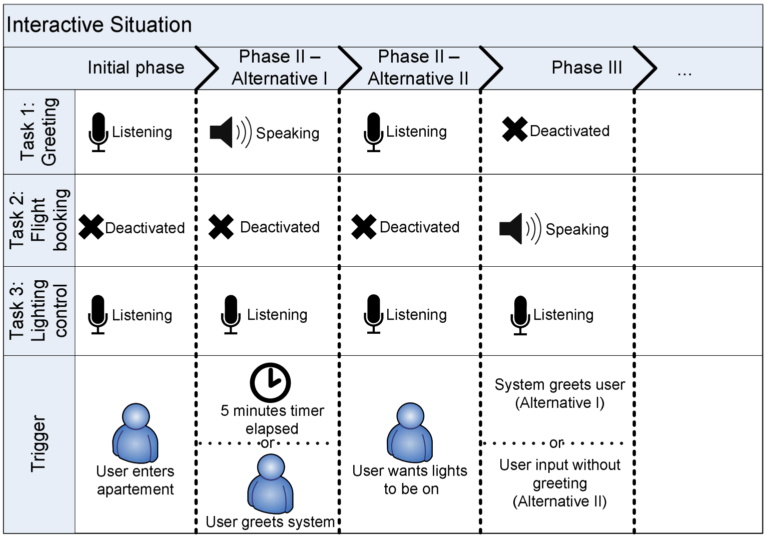


Fig. 1. An interactive situation that may occur with two alternatives

Fig. 1 presents two alternatives showing how the situation could proceed in Phase II: Alternative I contains two triggers that might allow the system to perform Task 1; the five-minutes-timer elapsed since the user entered the room or – probably the more usual case – the user greets the system. As mentioned above the reason for such a five-minutes-timer is that the system should act as unobtrusively as possible. Note that Task 3 is still active since the system is meant for handling more than one interactive task in parallel. If one of the two triggers is actuated the system would greet the user and add a semantic value such as "userInitiatedConversation" to the knowledge base (see Section 4). This would allow Phase III to start. Table 1 shows a possible conversation that might occur when making use of the proposed set of dialogues.

Alternatively Phase II could pass off conditioned by the user telling the system to switch the lights on. This would make Task I obsolete – the system shouldn't greet the user in response to such a spoken command. It would be more natural if the system skips the greeting task and activates the proactive Task 2 "Flight booking" instead. Fig. 1 shows Phase III constituted by the additionally activated Task 2 and the still running Task 3. The preceding greeting task has either become obsolete or has already been processed. Since the user (or the IE) can dynamically activate or deactivate the tasks the SDM may perform, it is possible at any time to end a conversation with the system or to start a dialogue the user respectively the system has not been aware of.

Table 1. A dialogue snipped that might occur during an interactive situation

Speaker	Utterance
Suki	Hello Julia!
Julia	Hi Suki!
Suki	Switch the lights on!
Julia	Do you want to start booking the flight now?

The example above describes a typical situation that may consist of more than one task in parallel. Obviously multitasking is a main source of and a main reason for adapting spoken dialogues in general. In [2] we have introduced three classes of adaptation: Device Adaptation, Event Adaptation, and Task Adaptation. In this paper we focus on two classes that influence the proposed multitasking approach. Herewith the first important class is *Device Adaptation.* Environmental changes may vary, depending on the surrounding and the situation of the user (kitchen, living room, car, etc.) and the availability of devices and services. This requires the capability to continuously change grammars, utterances, and system commands to a changing device population and changing user focus within the IE.

Event Adaptation also relates to environmental changes. Since various tasks within IEs are to be accomplished it is necessary to move the actual focus of an on-going dialogue to other (contingently more urgent) dialogues. These may consist of informative system utterances, alerts, or short yes-no-questions. Afterwards the on-going dialogue would have to be resumed. We have recognised two types of events that require adaptation: external events and internal events. While the former ones always need an entity that throws the specific event, the latter ones may be initiated by the dialogue manager itself. Reasons for initiating an internal event can be various and sundry: fixed priorities, dynamic priorities (i.e., changing over time), semantics, and depending on the progress (positive or negative) made within an on-going dialogue.

In the next section we will detail how the modular architecture of the prototype and the idea of a unified knowledge base describing the spoken dialogue and its state provide a fertile ground to realise adaptive behaviour in the described manner. One focus of the conducted research has been set on Event Adaptation and its two-sided mode of execution in practice. As mentioned above there are two kinds of events the SDM has to react on: external and internal events. The simplest case would be an external entity sending an event (e.g., an alert message) to the SDM, which immediately reacts and therefore interrupts any on-going dialogue in order to utter the alert message. However such behaviour is not always comfortable and comprehensible for the user. Thus we propose to not only react on external events but to incorporate also internal events. In case the alert message mentioned above is not time-critical it should not directly interrupt an on-going dialogue but should be suppressed until an internal event triggers the system to utter to message. For example, a trigger the system could use is the time that elapses since the external event occurred. If a specific threshold is reached (i.e., the alert gets more urgent during time) there is either no further need to suppress the message since the on-going dialogue has already been terminated or finally there is no other way than to interrupt the on-going dialogue.

4 The OwlSpeak Prototype

In order to meet the main requirements for multitasking mentioned in the previous section we have implemented the proposed Spoken Dialogue Manager OwlSpeak[1]. The Passive View variation of the Model-View-Presenter (MVP) pattern [8] has been used. This architectural approach allows for a maximum of flexibility regarding the switching between independent or interrelated tasks. The underlying idea of MVP is that an application or system should be divided into three logical parts, the Model, the View and the Presenter. The user only interacts with the View layer. Contrary to MVC the Presenter mediates between Model and View – the Model conveys no functionality, i.e., it is not an application but solely encodes the knowledge that is used by the Presenter. The term Model in this case refers to a *Domain Model*. Therefore especially for multitasking systems that provide a direct interface to the user (i.e., user interfaces) MVP is perfectly suitable.

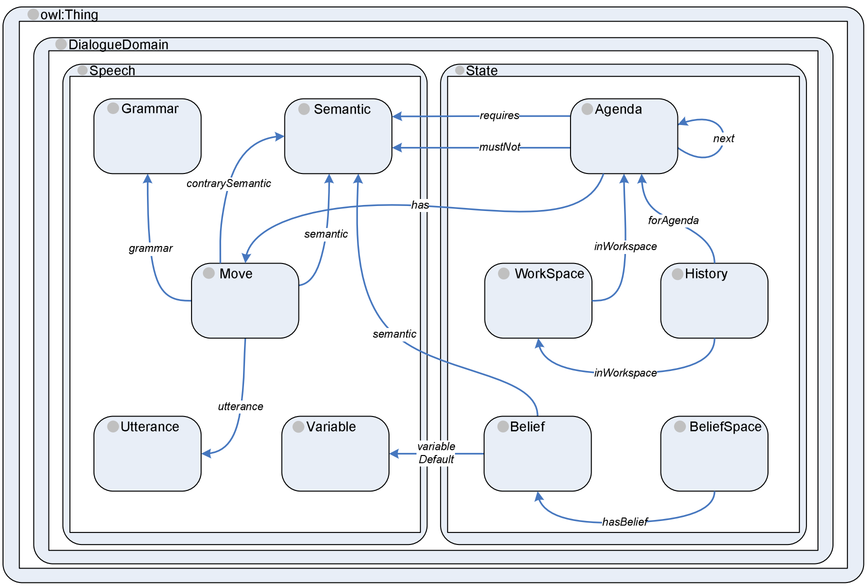


Fig. 2. Overview of the classes and main relations of the Spoken Dialogue Ontology

To be able to communicate with a user or with other external entities the application needs a knowledgebase that describes facts and the relation between such facts. A fact could, for example, be the name of a person or an ID number. A relation could be "has", which could be used for person "A" has ID number "4711". Without such knowledge a system would not be able to generate useful output, i.e., act as *knowledge source* nor to understand input that is provided by external entities, thus acting as *knowledge sink*. The term Domain Model could therefore be specified as the knowledge a system needs in order to be able to interact with the context and the user

[1] http://sourceforge.net/projects/owlspeak/

in a meaningful way. There are many ways to establish such a knowledgebase; SQL databases or XML files could be utilised, to name but a few. A more sophisticated option is to make use of ontologies to provide a common understandable knowledge base [6].

The underlying knowledge base of OwlSpeak is modelled using OWL ontologies, so called Spoken Dialogue Ontologies (SDOs). We have implemented a tree shaped structure to arrange the data-bearing individuals using a defined set of classes. The root of each knowledge base is DialogueDomain, which has the two subclasses Speech and State. We divide the ontology into these two main branches since we want to distinguish between knowledge that corresponds to the static structure and knowledge that corresponds to the dynamic state of the actual dialogue. Fig. 2 shows an overview of all classes populating the SDO together with the relations interlinking them. OwlSpeak makes use of a specific number of dialogue representations. These representations serve as Domain Models. Each representation provides knowledge about both dialogue flow and state of a specific spoken conversation. Depending on contextual information various sets of SDOs can be activated or deactivated. It is furthermore possible to add new representations for dialogues during runtime and therefore extend the knowledgebase, i.e., the Model.

5 Evaluation

The main question of the evaluation was *how do users cope with the multitasking capabilities of the prototype*? A challenge herewith was that it is exceedingly difficult to evaluate an SDM without evaluating the SDS that provides ASR and TTS, which certainly are both strongly perceived by the user. To solve this issue we divided the subjects into two groups each of them conducting the same spoken dialogue but using different multitasking strategies. All in all 26 mixed-gender subjects in the age between 17 and 59 years participated in the evaluation. The first group (Group A) conducted the main dialogue, a travel booking task, and received several reminders afterwards. The results of this group were used as baseline. The second Group (Group B) received the reminders dynamically during the on-going dialogue and therefore has been engaged in a more comprehensive multitasking conversation. The main issue of the evaluation setup was how to measure which of the approaches performs better and what does "better" mean in this context. Established SDS evaluation approaches such as PARADISE [9] utilise metrics such as "task completion", "repetition rate", and "error rate", which are only partly useful for rating the SDM itself. We have decided that a proper way of evaluating the multitasking capabilities of the prototype is to find out if it is easier for Group A or for Group B to recognize the reminders and to remind them after the dialogue has been concluded. During the evaluation the subjects had to imagine that they are talking to their IE, which is able to correspond with the travel agency in order to book a flight and a hotel for the next holiday. They furthermore received the following reminders:

- "Your friend Oliver has his birthday on the 15th of August. You might buy him a present."
- "I should remind you to rent a movie for tonight."
- "According to the weather report there will be heavy rain today."

All participants concluded the dialogues successfully. The overall performance of the system was rated positively. Both groups rated the system nearly equally on a scale from 1 (very bad) to 10 (very good) as good (6,35) in average. However as mentioned above these numbers are not sufficient to measure the performance, usability or functionality of the SDM. The user perceives the whole SDS and therefore merely rates it. However these numbers cannot be left aside: If the subjects had rated the whole system as bad it wouldn't make any sense to look for results that help us rating the SDM. A bad SDS would inevitably lead to a bad rating of the included SDM. Thus in our case an SDS rated as good serves as a fertile ground to measure the performance of the SDM, which was the main topic of the experiment.

Table 2. The number of subjects per group who gave right answers per category

Information	A	B	A+B
Friend's name	0	0	0
Day of Birth	8	5	13
Rent movie	6	2	8
Buying present	9	3	12
Watching movie	6	2	8
Rainy weather	8	9	17

Since the subjects were not explicitly told to take care of any additional information but only to complete the “travel booking” task, it was expected that only a few test persons would retain all information the system provides. Since Group A received the three reminders after the main tasks have been concluded we furthermore expected that Group A would experience a slight advantage. To be able to rate the outcome of the questionnaire we counted a point for each information the specific subjects kept in mind. Thus a subject who didn't remind any of the additional information would gain 0 points and a test person who reminded all of the provided information would gain 6 points: for the right name of the friend, the right date, renting a movie, buying a present, watching a movie, and the rainy weather. Table 2 shows the number of subjects per group who gave the right answers to the specific questions. We assume that the friend's name – Oliver – was only poorly synthesised, thus no one was able to remind it. It is obvious that the static group outperforms the dynamic group in all but one category. The information about the weather came up at the end of both static and dynamic dialogue. The subjects of Group B had much more problems perceiving the information that was provided during the on-going main booking dialogues. A reason for this could be that the test persons were nearly totally occupied by managing the travel booking task. Since the system did not provide any assistance in switching to a “reminder” task the users blanked out all information that was not related to the main aim: concluding the booking.

During a free discussion that followed each experiment several participants stated that they had consciously ignored all information that didn’t directly relate to the main task. However if the subjects had been informed beforehand that they will have to answer specific questions about the dialogue history we would have gain a totally different result since it is far from practice that a user would not be confronted with such reminders spontaneously. A main reason for the overall bad result – only the rainy weather was reminded by nearly all participants – could be that we had to use an English-language SDS. This could have been problematic for the German native speakers that participated in the experiment. However we expect that even with a German speaking system the discrepancy between the two groups approach would still be observable.

5 Conclusion

The current version of the OwlSpeak Spoken Dialogue Manager already fulfils several requirements that arise from the multiple task-based situations that occur within IEs. It is able to pause and resume active tasks, add and remove dialogue domains, permanently save the state of a dialogue, and furthermore it provides more than one active spoken dialogue in parallel.

Those functionalities allow for adaptive spoken dialogues that cover several disjoint or partly overlapping domains. However it seems to be obvious that it is a different task for the user to use the spoken interface to solve a *single* task (e.g., book a flight) with the help of an SDS or to use an SDS to interface with an IE that provides a variety of different controls and tasks. Thus the main question that has to be answered before multitasking SDSs will be used in everyday life is *how do users cope with the multitasking capabilities of a system such as the proposed prototype?*

Compared to GUIs users are already accustomed to their multitasking capabilities: techniques such as taskbars or widgets are totally adopted by the users. However when it comes to spoken interaction we are far away from such a wide spread user acceptance. Enhancing the common usage of SDSs by adding the functionality of multitasking could be an important step towards wider application of spoken interfaces. The evaluation revealed several questions that have to be answered before. How can a change of focus from one task to another be signalised by the system or by the user? How can the system distinguish between user inputs that might relate to different tasks? How can the user distinguish between system outputs that might relate to different tasks? Future work would be to find and evaluate technical solutions to answer these questions. We plan to compare different dialogue strategies that might be used to signalise a task change: for example, remarks before changes occurs, auditory icons, or more complex sub-dialogues may proof useful for the users. Furthermore we are currently implementing sophisticated functionalities that enable OwlSpeak to detect task changes that are initialised by the user.

Acknowledgment. The research leading to these results has received funding from the European Community's 7th Framework Programme (FP7/2007-2013) under grant agreement n° 216837.

References

[1] Bohus, D., Rudnicky, A.I.: The ravenclaw dialog management framework: Architecture and systems. Computer Speech & Language 23, 332–361 (2009)

[2] Heinroth, T., Denich, D., Schmitt, A.: Owlspeak - adaptive spoken dialogue within intelligent environments. In: 8th IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops), pp. 666–671 (March 2010)

[3] Heinroth, T., Kameas, A., Pruvost, G., Seremeti, L., Bellik, Y., Minker, W.: Human-Computer Interaction in Next Generation Ambient Intelligent Environments. Intelligent Decision Technologies 5(1) (2011)
[4] van Helvert, J., Hagras, H., Kameas, A.: D27 - prototype testing and validation. Tech. rep., The ATRACO Project (GA no 216837, 7th FP) (2009)
[5] Larsson, S., Traum, D.: Information state and dialogue management in the trindi dialogue move engine. Natural Language Engineering, 323–340 (2000)
[6] McGuinness, D.L., van Harmelen, F.: Owl web ontology language. W3C Recommendation (2004)
[7] Minker, W., López-Cózar, R., McTear, M.: The role of spoken language dialogue interaction in intelligent environments. Journal of Ambient Intelligence and Smart Environments 1(1), 31–36 (2009)
[8] Potel, M.: MVP: Model-View-Presenter The Taligent Programming Model for C++ and Java. Tech. rep., Taligent Inc (1996)
[9] Walker, M.A., Litman, D.J., Kamm, C.A., Abella, A.: Paradise: a framework for evaluating spoken dialogue agents. In: Proceedings of the Eighth Conference on European Chapter of the Association for Computational Linguistics (1997)
[10] Young, S., Williams, J., Schatzmann, J., Stuttle, M., Weilhammer, K.: D4.3: Bayes net prototype - the hidden information state dialogue manager. Tech. rep., TALK - Talk and Look: Tools for Ambient Linguistic Knowledge, IST-507802, 6th FP (2006)

From Clouds to Rain: Consolidating and Simplifying Online Communication Services with Easy One Communicator

Jeffery Hoehl[1] and Gregg Vanderheiden[2]

[1] Coleman Institute for Cognitive Disabilities and Department of Computer Science
University of Colorado, Boulder, Colorado USA 80309
jeffery.hoehl@colorado.edu
[2] Trace R&D Center, University of Wisconsin-Madison
Madison, WI USA 53706
gv@trace.wisc.edu

Abstract. The growth of cloud computing has encouraged an ecosystem where diverse, task-specific web and mobile applications have flourished. However, this fragmentation of tasks across disparate applications has introduced a level of complexity and separation that can be time-consuming and difficult to navigate for individuals who are older, have cognitive disabilities, or are unfamiliar with web and mobile interfaces. Each new application represents an additional interface that must be learned, and not necessarily easily or intuitively. Easy One Communicator is a web application concept that addresses the complexity of online communication tools by consolidating and simplifying the management of online communication. Easy One Communicator provides a simple interface for users to read and send emails, text and video chat, share photos and maintain contact information. The interface is designed to be obvious to use and personalized to each user's changing cognitive abilities and needs.

Keywords: Cognitive disabilities, universal design, cloud computing, computer-mediated communication.

1 Introduction

The number of Internet users has steadily risen over the past decade and continues to do so both in the United States and worldwide [1]. More users are not only online, but are online more frequently, use an increasing number of online services, access the Internet across more types of devices and are communicating with each other in more ways than before [2]. Online communication services and social media have become a significant portion of how people are communicating with each other and are often used to supplement offline, face-to-face relationships [3].

The growth of online services and the creation of new services have largely been supported by cloud computing technologies. Cloud computing has allowed for a lower barrier of entry for new services that have the illusion of infinite resources [4] while allowing users to more flexibly interact with software over the Internet with

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 52 – 61, 2011.

little more than a computing device, an Internet connection and a web browser. The growth of cloud computing has not only allowed for an increased number of online communication services to develop as a whole, but has encouraged an ecosystem of task-specific web and mobile applications to flourish, in contrast to the large, integrated desktop applications of the past. Users often routinely switch between several applications for emailing, chatting, photo sharing, blogging, micro-blogging, social networking and video-chatting tasks.

However, this fragmentation of tasks across disparate applications has introduced a level of complexity that can be time-consuming and difficult for many users to navigate. Each new web and mobile application is an additional user interface that must be learned and familiarized. Though some user interface elements may be similar, most applications have a unique layout and workflow that is not always easy to learn or intuitive to navigate. This can be especially problematic for individuals who are older, for individuals with cognitive disabilities, or for anyone unfamiliar with web and mobile interfaces.

The fragmented social media application ecosystem has created a need for a simple alternative web application that provides online communication tools for users that find the traditional interfaces difficult to use. Easy One Communicator is a web application that addresses the complexity of online communication tools, and the ecosystem as a whole, by consolidating and simplifying online communication tasks. Easy One Communicator provides a simple interface for users to read and send emails, participate in text and video chats, share photos and maintain contact information. The interface is designed to be as obvious to use as possible and easily personalized to each user's changing cognitive abilities and needs. The interface can be adjusted to add features as a user becomes more proficient with the application over time or can be adjusted to remove complexity as a user becomes less proficient. The two-way adjustment is particularly important for aging users and users with cognitive disabilities that may not become more proficient with increased application use and may even become less proficient over time.

Although other web applications are available for aggregating communication streams, they often present data in a dashboard fashion that addresses convenience at a cost of increased complexity and decreased ease of use. Easy One Communicator focuses largely on providing a simple, easy to use and easy to learn interface rather than solely providing data aggregation.

2 Cloud Computing and Social Media

2.1 Cloudy Definitions

As discussed by Armbrust, et al., in [4] the term "cloud computing" has been given many meanings and definitions. It has been used to describe everything from the Internet as a whole to distributed systems, grid computing, web applications and utility computing, amongst many other computing paradigms. This paper will follow the distinctions made by Armbrust, et al., and generally use "cloud computing" to refer to "both the applications delivered as services over the Internet and the hardware and systems software in the data centers that provide those services" [4].

A distinction can still be made within the definition of cloud computing between the "applications delivered as services over the Internet" and the "hardware and systems software in the data centers." Specifically, "Software as a Service" (SaaS) refers to the delivered applications, while the hardware and system software used are referred to as a "cloud" [4]. Thus, cloud computing describes the larger architecture by which hardware and software is used to provide large-scale, on-demand computing resources while SaaS generally describes the actual applications and services provided to end-users that are often built on top of cloud computing architectures.

Adding to the confusion between cloud computing and SaaS is that "[i]n some cases, the same actor can play multiple roles. For instance, a cloud provider might also host its own customer-facing services on cloud infrastructure" [4]. This is often the case in many of the most popular, free web applications available to Internet users. For instance, Amazon, Facebook, Google and Microsoft all maintain cloud computing architectures that they provide access to for SaaS providers in a typical "cloud provider" role, however, each company also provides their own SaaS solutions directly to end-users in the form of web applications.

2.2 Benefits of Social Media Clouds

To many end-users, the distinctions between the logical roles of a cloud provider and a SaaS provider are unimportant and thus conflated and understood simply as an abstracted role of a web application provider. A user of a social networking service (SNS), for instance, may only interact with the web interface provided through a typical browser and not be interested in other methods for accessing the social network.

However, one of the greatest benefits of SaaS providers are the APIs and web services that are provided to access the underlying data and services. By providing public APIs, many SNS providers allow for access to the features and services of a particular cloud through alternative interfaces and devices.

Expanding upon the SNS example above, APIs and web services allow for multiple interfaces to a social network to be created. For instance, one can access many social networks through not only a standard desktop web browser but also through cell phones, smartphones, tablet computers, client-side applications and even consumer electronic devices such as web-connected televisions, game consoles, etc. These interfaces can be provided by the SNS provider or by third-party developers. Furthermore, SaaS providers often allow developers to send and receive data across other cloud computing services; data from one social network can be streamed to another and vice versa. Thus, social network interfaces can be tailored to the devices they are run on as well as the contexts they are used in.

The wide availability and use of cloud computing APIs has led to the main promise and selling point of cloud computing: from an end-user's perspective, cloud computing allows access to their information from anywhere in the world, at anytime, and from any device as long as they have an available Internet connection.

2.3 Drawbacks of Social Media Clouds

The flexibility and decentralization in the cloud computing architecture inherently has a complexity to its use that has been exacerbated by the growing popularity of mobile devices and web applications. The increased use of smartphones and web applications has increased the use of applications that are smaller in scope and more task-specific than traditional applications. Rather than having a relatively few number of applications that access multiple clouds and services, recent trends have been to create a larger number of applications that access a fewer number of clouds and fewer features in each API. Thus, end-users are exposed to, rather than insulated from, the increasing number of services being made available. In 2009, 36% of adults managed two SNS profiles while 16% managed three or more. These numbers are rising as well indicating that adults are accumulating and managing an increasing amount of SNS profiles [2]. In other words, managing one's online social communication requires an increasingly greater number of discrete applications rather than using fewer, multifaceted applications.

At the application level, this approach seems beneficial. Each mobile and web application is inherently "simpler" due to a limited scope and feature set. However, the complexity lies not in each individual application, but in the volume and disparity between applications. Rather than having a single application with a consistent interface, users are exposed to an increasing number of interfaces with varying layouts, workflows, designs and features. Furthermore, as more and more communication moves online, the learning and use of varied applications is required rather than learning new features within a familiar application.

This approach is especially problematic for older users, users unfamiliar with current mobile and web interfaces, and users with certain cognitive disabilities that may have difficulties learning and keeping track of different interfaces. Various desktop, mobile and web applications are required to accomplish common communication tasks including managing email, text chat, video chat and photo sharing.

2.4 Connecting Clouds

By leveraging the APIs provided by cloud computing providers, the various services of widely used web applications can be combined in a simpler and easier to use interface. Consolidating and centralizing access to various online communication channels can avoid the inherent complexity caused by disparate applications and can provide a single point of access to numerous social media services. This reduces the number of required applications needed to complete a task and immediately reduces the complexity of using online communication tools.

Providing a custom interface also allows the tools and services to be customized for specific needs and requirements of a target population. Standard interfaces to cloud-based services often do not take into consideration the needs of users with cognitive disabilities or the needs of many aging populations. By creating a custom interface, the needs of specific populations can be directly addressed and handled more effectively.

A significant benefit of leveraging APIs from cloud computing providers rather than providing one's own underlying services is that most of the benefits of cloud computing are retained. Since the underlying platforms are still cloud-based, the robustness, scalability, security and performance of the cloud-based architecture are retained without the hardware costs or maintenance overhead. In many ways, users have come to expect cloud computing levels of performance from web applications and not utilizing cloud computing services could lead to downtime, performance issues and a general user experience deemed unacceptable by users.

3 Easy One Communicator

The Easy One Communicator application adopts the idea of bridging several disparate cloud-based communication services into one web-based application. The web application is designed with an emphasis on simplicity and ease of use.

3.1 Target Audience

Although online adults still generally lag behind their teen and young adult counterparts in social media use, they do use certain services more often. For instance, Twitter usage and blogging among adults is increasing while decreasing among teenagers [2]. This suggests that adults do care about and use social media software, especially those applications intentionally designed to be easy to use and that have simple, straight-forward interfaces and presentations.

Easy One Communicator extends the notion of providing easy to use social media software to create a very simple, layered communication application for use by individuals who are older, for individuals with cognitive disabilities, and for anyone that finds current email, instant messaging and computer communication programs too confusing.

It is estimated that there are over 20 million persons with a cognitive disability in the United States alone [5]. This number is expected to increase as the population ages, and the target audience is even larger when one includes users who simply prefer a single interface to the separated interfaces of most online applications.

3.2 Communication Capabilities

Since Easy One Communicator leverages APIs from cloud computing providers, its capabilities are theoretically bounded only by the available APIs and features provided by social media cloud providers. However, to effectively provide a useful, unified communication application, several major capabilities were identified as minimum requirements to a simple online communication platform:

- To receive and read email
- To write and send email
- To receive and acknowledge an invitation to chat by text, voice and/or video
- To send an invitation to chat by text, voice and/or video
- To receive and view pictures
- To use a contact book that integrates with the email and chat capabilities (and, optionally, to maintain it)

Each capability should initially be provided with a limited set of functionality to promote the simplicity and ease of use of the application. For instance, advanced email features (i.e. BCC, rich-text, embedded images) need not be supported to instead promote simplicity of the user interface [6].

To further simplify this for an initial user, the contact book and photo album can be maintained by another person who acts as assistant using a separate interface. The album can also be setup to allow a set of people to submit to it so it can continually be refreshed by say family members without the complexity of maintenance falling on the user.

3.3 Simple and Adaptive Design

The Easy One Communicator application is intended to be as simple and obvious to use as possible. Ideally, it should take no instruction to use, even for someone who has never used a program of this type before (and thus be usable even by someone who cannot remember instructions from one day to the next). Ideally, it could be used by someone who has never used a computer before. The interface is intended to be simple and self-cuing enough that simply knowing that they would like to send an email or to look at an email or to chat should be enough for them to see immediately what they should do based on previous life experiences.

This task is particularly challenging because so many things that programmers take for granted are not necessarily true for the target user population. Windows and modes, menus and hierarchies, icons that “launch” programs or features, pages that disappear, etc., can prove to be completely foreign and baffling. And one cannot count on users “figuring it out” or “learning the system over time.” What some target users learn one day may be forgotten the next. Although some of these users may be able to remember complicated recipes and procedures that they learned earlier in their lives, they may not be able to remember something that they learned (or were taught) today when tomorrow comes.

Because the target population will have differing levels of skill and because some individuals will be able to learn over time and add small amounts of complexity, the application benefits from being designed in a layered fashion. It has one mode that is ultra simple with just the basic features. There are then additional, configurable, modes that are almost identical, but add a couple extra features or capabilities. The available features can be chosen for each user based on his/her abilities and progression with the system. The modes are all as consistent and identical as possible to facilitate progression up and down the levels. It should be noted that some individuals might also regress (e.g. need features removed from the modes) over time.

One version of what an Easy One Communicator might look like is show in figures 1 and 2. The main communication functions are always visible on the screen so that a user can easily switch between tasks, and available actions are readily identifiable. Multimodal presentations of information are used when possible by combining text and graphics for functions. The application itself also has very few distracting visual elements with all displayed items being actionable and a limited number of options being displayed at one time.

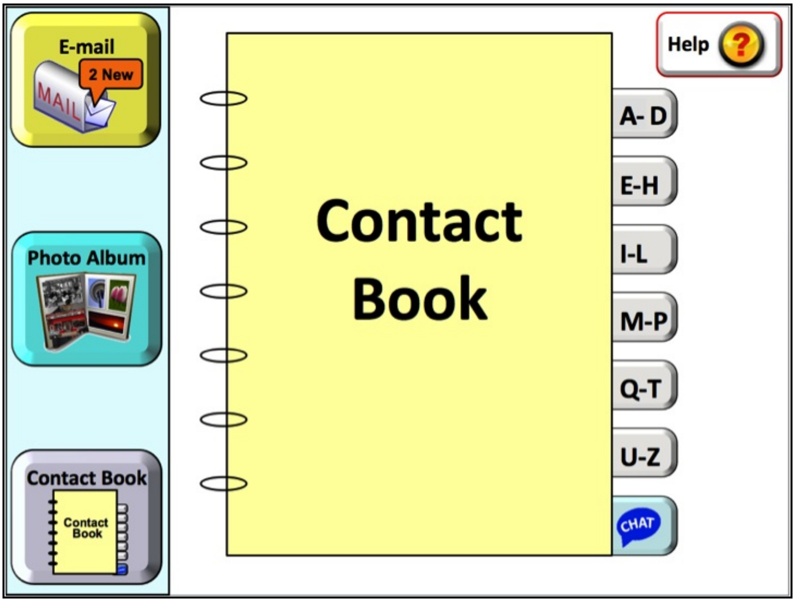


Fig. 1. An example of what a contact book might look like for the Easy One Communicator with buttons shown for browsing the user's contacts by name

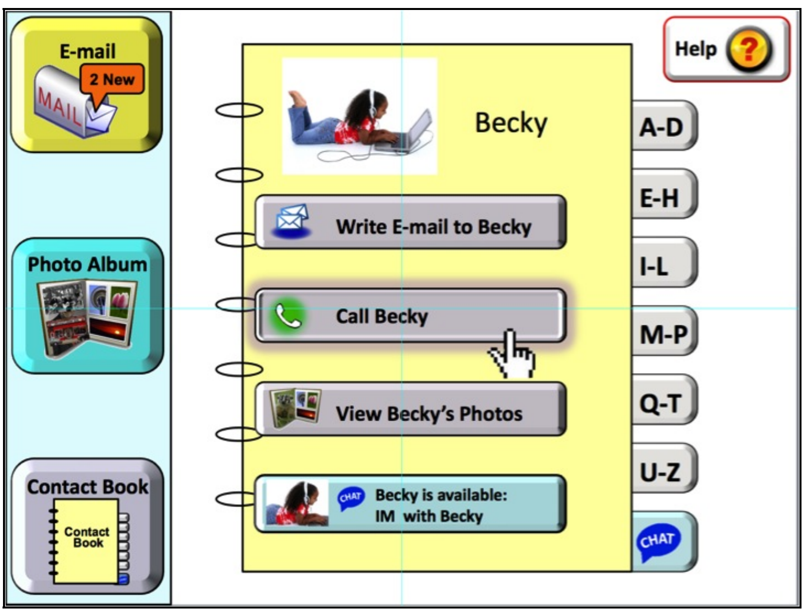


Fig. 2. An example of what a contact book entry might look like for the Easy One Communicator with options shown for emailing, calling, viewing photos and chatting with another user

3.4 Minimal Technical Requirements for Use

To make Easy One Communicator available to the broadest range of users, a focus is made on requiring very few technical requirements and as little technical knowledge as possible. The use of the application should require only a web browser, with HTML 5 and JavaScript support, and an Internet connection. By keeping these requirements simple, additional software installation and configuration is avoided, a common trouble spot for many client-based and desktop applications. Furthermore, since Easy One Communicator runs as a web application, maintenance (such as software updating) is unnecessary and less technical knowledge is required to begin using the application. Many users can begin using Easy One Communicator by simply entering a URL into a web browser.

To facilitate use by someone with a memory problem, a USB "keytyper" fob was developed that can launch a browser in kiosk mode when inserted into a computer, putting the person on an Easy One Communicator web page while completely hiding the operating system, browser toolbars and all of their complexity (figure 3).

The technical requirements are intentionally designed to meet those that are available in many public spaces so end users are not required to own or have full permissions to a personal computer. Easy One Communicator should be able to run from Internet-connected computers, such as those found in libraries, coffee shops and hotels, which often do not allow the installation of new software due to security restrictions but do allow access to the Internet and web-based applications. An elder could plug the key fob in and instantly turn any computer into their simple computer including the one in the library, nursing home, children's home or anywhere.

Fig. 3. A USB "keytyper" fob that, when inserted into a computer, will automatically launch Easy One Communicator in a kiosk mode

3.5 Offloading the Complexity

Although Easy One Communicator is designed to abstract the user away from many of the technical complexities of current social media applications, its design inherently involves some potentially complex administration.

Easy One Communicator involves combining resources from cloud computing service providers and a user account must be managed for each service used. For basic functionality user accounts are needed for an email service, photo sharing service and chat service. In the case of chat, several services require additional user accounts. For instance, if an Easy One Communicator user would like to chat with contacts on the AIM or MSN networks, they must have an account on those specific services. Like a user's photos and contact book, the administration of the ancillary accounts would be handled by another user that acts as an assistant, likely a family member, close friend or service, and will need a separate interface for administration.

Like most web-based applications, security and authentication are major considerations needed to protect users' information and data. These concerns are especially important for the target audience since some users may not have the knowledge to protect themselves while using traditional social media applications. Some security can be provided without requiring the user's involvement or additional technical knowledge. For instance, access to Easy One Communicator can be limited to secure protocols (such as HTTPS) to ensure data is encrypted in transport. Due to its simple interface and walled-garden approach, Easy One Communicator is also able to offer some inherent security protection at the application level. For instance, the email addresses and IM aliases of a user's contacts are not presented to most end users directly; only the names of the contacts are displayed. As such, it is difficult for an unauthorized user, to accidentally obtain this information and provide it to untrustworthy sources. The system can be configured to only accept incoming mail or calls from registered individuals preventing spam and scams. Authentication can also be handled through alternate means. Since many users find it difficult to not only remember but also routinely update and change their password, alternate authentication methods can be provided. For instance, users could potentially be asked to provide voice-based authentication or picture-based authentication.

3.6 Future Work

As social media technology grows and hardware continues to advance, many new features and capabilities will continually be provided to users of traditional social media applications. Easy One Communicator is designed to allow these capabilities to be included, on a configurable basis, to its user base as well. The layered and configurable modes could be enhanced to include not only advanced features but also newer communication modes of social media.

For instance, status updates, micro-blogging and location services are becoming increasingly popular across many social media applications. A possible enhancement to Easy One Communicator could be the addition of a simplified news feed or a method to add status updates. As features become more useful to users of traditional social media, they can also be made available to Easy One Communicator's users.

4 Conclusion

As social media becomes increasing popular its value increases to a growing number of users. As is often the case, many social media service providers have not adequately responded to the needs of many users that find the current tools too complicated to use. Easy One Communicator integrates the most commonly used social media tools into one web-based application with a focus on simplicity and a layered design that adapts to a user's cognitive abilities and needs. By providing an easy to use solution, the application can allow aging users, users with cognitive disabilities and anyone that finds current applications too confusing a method for continued participation in online social media.

Acknowledgments. The contents of this paper were developed with funding from the National Institute on Disability and Rehabilitation Research, U.S. Department of Education, grant number H133E080022 (RERC on Universal Interface and IT Access) and the Coleman Institute for Cognitive Disabilities (http://www.colemaninstitute.org). However, the contents do not necessarily represent the policy of the Department of Education, and you should not assume endorsement by the Federal Government.

Many of the ideas and thoughts presented have also been the result of discussions and feedback provided by Clayton Lewis, Jim Sullivan and Antranig Basman.

Graphics of Easy One Communicator courtesy of Team 1 ISyE 662 - Design for Disability and Aging Class at the University of Wisconsin- Madison: Laura Bagley, Steve Baran, Daniel Burkhardt, Maia Jacobs, Sara Karle, Matthew Kopetsky, Emily Maslonkowski, Josh Mayer, Ryan McAsey, Sarah Offutt and Andrea Walk.

References

1. World Internet Usage Statistics News and World Population Stats, `http://www.internetworldstats.com/stats.htm`
2. Lenhart, A., Purcell, K., Smith, A., Zickuhr, K.: Social Media & Mobile Internet Use Among Teens and Young Adults. Pew Internet & American Life Project (2010), `http://www.pewinternet.org/Reports/2010/Social-Media-and-Young-Adults.aspx`
3. Lampe, C., Ellison, N., Steinfield, C.: A face(book) in the crowd: social searching vs. social browsing. In: Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work, pp. 167–170 (2006)
4. Armbrust, M., Fox, A., Griffith, R., Joseph, A.D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I., Zaharia, M.: A view of cloud computing. Commun. ACM. 53, 50–58 (2010)
5. Braddock, D., Rizzolo, M.C., Thompson, M., Bell, R.: Emerging Technologies and Cognitive Disability. J. Special Education Technology. 19, 49–56 (2004)
6. Lewis, C.: Simplicity in cognitive assistive technology: a framework and agenda for research. Univers. Access Inf. Soc. 5, 351–361 (2007)

Use of Speech Technology in Real Life Environment

Ruimin Hu[1], Shaojian Zhu[2], Jinjuan Feng[1], and Andrew Sears[2]

[1] Computer and Information Sciences Department, Towson University, Towson, MD 21252
{jfeng,rhu1}@towson.edu
[2] Department of Information Systems, Interactive Systems Research Center UMBC, Baltimore, MD 21250
{szhu1,asears}@umbc.edu

Abstract. This paper reports results from two six-month field studies investigating the use of speech-based interactions in real world environments. The first study focused on the use of speech-based dictation/word processing tasks that allow users to generate text such as reports or emails. The second study took a broader view and examined two types of interactions: speech-based dictation for word processing and speech-based command and control supporting interactions with the graphical user interface. The results suggest that user requirements for speech-based interactions have evolved as the technology matured to include better support for formatting text documents as well as more effective support for other applications. While significant research has focused on the use of speech recognition to generate text, our participants spent substantially more time using speech recognition for other, often unexpected tasks such as instant messaging and web browsing. The use of speech recognition to support informal communication is particularly interesting as accuracy may be less critical in this context. Further research is needed to address these emerging requirements for speech technologies.

Keywords: Speech-based Application, Speech Interaction, Speech Technology, Physical Impairment, Field Study

1 Introduction

Speech-based applications allow users to communicate with computers or computer-related devices without the use of traditional input devices such as the keyboard and mouse. As a result, speech technologies are of particular importance for some individuals with physical disabilities. While multiple studies have investigated the use of speech-based input by individuals with physical disabilities, most were lab-based studies that focused on pre-defined tasks that were somewhat, but not completely, representative of the everyday tasks and environments [20]. In addition, these studies have typically been quite short in duration, ranging from one to ten hours [6]. As a result, we have limited knowledge of how people may use speech recognition as they

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 62–71, 2011.

interact with computers as part of their daily activities in realistic environments over an extended period of time.

To begin filling this gap, we conducted two six-month field studies investigating how individuals with and without physical disabilities use speech technologies when interacting with desktop computers in their home environment. The first study focused on the use of speech recognition for dictation/word processing tasks such as writing reports or emails. The second study took a broader view and examined two types of interactions: speech-based dictation for word processing and speech-based command and control to support interactions with the graphical user interface.

2 Related Research

Automatic Speech Recognition (ASR) has experienced significant commercial success in recent decades [13]. Traditional dictation software such as Dragon Natually-Speaking offers an improved user interface and higher recognition accuracy than in the past. The integration of speech recognition engines into commercial operating systems makes ASR readily available for the general public. Simultaneously, potential applications of ASR have become more numerous, covering a variety of tasks beyond the traditional domain of text documents. For example, both Dragon NatuallySpeaking and built-in speech recognition support in Windows Vista™ both support web browsing and desktop navigation. With all of these recent developments, there is limited research examining if and how these new functions are being used.

Speech-based dictation systems have been the traditional focus of both academic research and industry. For example, numerous hands-free ASR error correction techniques have been explored (e.g. [2]). Halverson et al. [8] studied error correction in large vocabulary, continuous speech recognition systems and identified different error correction patterns depending on the users' experience with speech-based interactions. A number of studies have investigated speech-based navigation within the context of text documents. Among them, McNair and Waibel [17] investigated an early version of target-based navigation where incorrect words were selected via speech. Sears et al. [20, 21] investigated how individuals with high level spinal cord injuries (SCI) employed speech recognition systems to compose text documents. No differences were identified between participants with and without SCI in terms of recognition error rates, navigation command failure rates, or overall productivity. However, participants with SCI interrupted their dictation more frequently to correct errors and were more satisfied with their interactions. In addition, two long-term studies have investigated the use of speech-based dictation systems by individuals with motor and cognitive disabilities, but both studies included only one participant making it difficult to generalize the findings [16][19].

Speech-based command and control can support interactions with typical graphical user interfaces rather than text documents. Three techniques are currently available: target-based selection, speech-controlled cursors, and grid-based selection. Manaris and Harkreader [15] explored on the use of speech recognition to generate keystrokes and mouse events. Karimullah and Sears [14] focused explicitly on cursor control, evaluating the efficacy of a predictive cursor designed to help users compensate for processing delays associated with speech recognition. Many other approaches for controlling a cursor have been explored. For example, Mihara et al. [18] discussed a system in which multiple ‘ghost’ cursors are aligned vertically or horizontally with the actual cursor while de Mauro et al. [5] investigated a voice-controlled mouse that used individual vowels as commands. Harada et al. [9] explored a similar idea called a vocal joystick that allowed users to control the cursor by varying vocal parameters, i.e., vowel quality, loudness, and pitch. Finally, grid-based solutions position the cursor using recursive grids allowing users to ‘drill down’ until the cursor is in the desired location [12] with two variations of the grid-based solution being compared by Dai et al. [4].

A few studies examined speech-based interactions in different domains. For example, Christian et al. [3] investigated speech-based navigation in the context of the web while Arnold et al. [1] examined speech-based programming system that integrated speech recognition and a predefined syntax for the programming language. Similarly, Hubbell et al. [11] developed a syntax-directed graphical editor for programmers with physical impairments. Sporka et al. [22] explored the control of computer games using speech and non-verbal vocalizations while Harada et al. [10] developed Voicedraw, a completely hands-free speech application for generating free-form drawings. Both speech and non-speech voice-based interaction techniques were adopted.

These studies provided insights into the effective use of speech-based interactions. Clearly, speech-based solutions are more appropriate for certain types of tasks and users. An important limitation of most previous studies is that the evaluations were conducted in less than realistic environments with artificial tasks. While several studies have evaluated speech-based interaction in more realistic contexts, such as web navigation or gaming, the focus was on a single task, isolating users and ensuring that they would not interact with rest of the computing environment. Given the variety of tasks users perform on computers, and the variety of speech-based interactions that can be supported on personal computers, there are important questions to be answered regarding how users will interact with speech applications in realistic settings. In the current field studies, we investigate how users with and without physical disabilities use speech to interact with computers in their everyday lives.

3 First Field Study

3.1 Methodology

The first study focused on the use speech-based dictation to generate text documents. Five participants took part in this study (see table 1). All participants used computers provided by the researchers with exactly the same hardware and software specifications. Computers were provided for two reasons. First, some participants did not have their own computers at the time of the study. Second, some participants who had computers used different operating systems and applications. It is difficult to provide users with comparable speech-based interaction experiences if they are using different hardware or operating systems; thus it would be difficult to compare results among users. Providing computers for all participants addressed these problems. Several standard Windows Vista speech functions were modified for use in this study including one method of navigating within dictated text and a second technique that supported navigation within any application.

Dictation and Editing. The Vista speech system offers two ways of navigating within text documents. Target-based navigation allows users to select words by stating the navigation target (e.g., 'select book'). Direction-based navigation allows users to select targets by specifying the movement direction and units (e.g., 'move up five lines'). We added anchor-based navigation to provide more flexibility in error correction [7].

Desktop and Menu Interaction. Vista offers a grid-based navigation mechanism. Currently there is no limit to how many times a user can zoom in to a smaller portion of the screen, but after three or four levels it becomes difficult to identify either the target or grid numbers. We modified the existing grid-based solution to disable zooming after three levels at which point a magnification function was enabled to enlarge the selected grid. At this point, four simple commands ('Up', 'Down', 'Right' and 'Left') were enabled to allow users to fine-tune the cursor location. For detailed information on the effect of the magnification function see [23].

Table 1. Age and description of disabilities

Users	Age	Physical impairments
S1-P1	60	Arthritis in neck, spine, hands and wrists (carpal tunnel)
S1-P2	19	Duchene muscular dystrophy
S1-P3	34	Dexterity with M.S. and left hand
S1-P4	48	Stroke; Arthritis limits upper arm, hand & neck movements
S1-P5	41	Significant weakness in hands & arm

The study was conducted at the participants' homes and lasted for six months. During the first visit, the researchers demonstrated the speech functions offered in Microsoft Vista including the two modifications outlined above. Participants were instructed to use the computer and speech applications for whatever tasks they need to complete, but they were required to generate at least four pages of text with no restrictions on the content. Interviews were conducted at the end of each month to collect feedback and logged data.

3.2 Results

Overall, all participants increased how quickly they generated text in the first several months, indicating the positive learning effect. However, all participants also slowed down after the initial improvement and there were notable fluctuations in performance. For three participants (P1, P2, and P4), the average number of words generated is lower compared to results from earlier lab-based studies.

Quality of text generated the study was assessed based on the percentage of sentences containing errors (including both recognition and grammatical errors). The text documents generated by participants did not show a consistent trend with regard to quality. Fluctuation was very common. The quality of the documents produced by P1 and P2 improved during the observation period, but there was no clear trend for P3, P4, and P5.

Efficiency (WPM) does not appear to be a decisive factor for user satisfaction and future adoption of the technology. P1 and P4 had similar entry speed (3.7 vs. 5.1) with contrasting satisfaction and attitude. P1 was very negative and would not continue using the technology while P4 was very positive throughout the study. P5 was pretty fast (13.3) but held a neutral attitude towards adoption while both P2 and P3 (8.7 and 11.8 respectively) held a positive attitude towards adoption.

4 Second Field Study

4.1 Methodology

The second study focused on a broader context and investigated both speech-based dictation for word processing and speech-based command and control to support interactions with the graphical user interface. Ten participants took part in this study, five had no physical disabilities and five had disabilities (see table 2). The study employed the same procedure as the first field study except that the participants were instructed to use the computer and speech applications for whatever tasks they would like to complete.

Table 2. Age and description of disabilities

Users	Age	Physical impairments
S2-P1	57	Severe Carpal Tunnel Syndrome
S2-P2	57	Injury resulting in muscle weakness and lack of sensation in hands and arms
S2-P3	37	High level Spinal Cord Injury (C5, 6)
S2-P4	38	Stroke
S2-P5	58	Multiple Sclerosis

4.2 Results - Differences between Two Groups

Applications Used/Tasks Completed. The participants without disabilities used a larger variety of applications than those with disabilities. Although both groups used the speech functions in Word, Internet, Outlook, IM, and Desktop navigation, only Word was used by every participant. In addition to these applications, the participants without disabilities also used speech commands in the context of PowerPoint, programming, games, and Notepad. The users with disabilities focused more on using Word and Internet Explorer, which served their basic needs. Table 3 summarizes the applications used by the participants in both groups.

Table 3. Applications used by both groups

Applications Used	Users without Disabilities	Users with Disabilities
Word	X	X
Internet/email	X	X
Outlook/email	X	X
Desktop navigation	X	X
IM	X	X
Music and videos		X
PowerPoint	X	
Programming	X	
Game	X	
Notepad	X	

Frequency of Use. To our surprise, the participants without disabilities used the speech functions more frequently than those with disabilities. We observed that the infrequent usage of speech functions by the participants with disabilities seemed to be associated with a general lack of interest in computer usage rather than the speech functions themselves. Two of these participants did not have a computer at the time that the study started, so a computer was not an integral component of their daily lives. In general, these participants experienced difficulty generating text documents, so they had developed strategies to avoid those tasks.

Evolving Requirements for Speech Applications. This study confirmed a number of well-known challenges regarding speech-based applications, such as frequent recognition errors and inefficient navigation. More interestingly, the results revealed a number of emerging requirements that may serve as the focus for future research.

Effective Editing and Formatting Functions. In the past, recognition errors were widely acknowledged as a major problem for speech-based dictation applications. Consistent with these concerns, participants in our earlier lab studies used editing functions to fix recognition errors but they did not spend time worrying about the appearance of the resulting text. In contrast, the participants in the field study were very interested in fixing the appearance of the generated text documents. They spent substantial time trying to make the document ‘look right’. They cared about the details that participants in our lab studies would typically ignore. For example, participants frequently made changes to punctuation, fonts, and how text was aligned. During interviews, participants expressed a desire to learn more and to become proficient using the editing and formatting commands.

Most speech-based dictation systems offer editing and formatting functions. However, due to the large number of editing and formatting functions available, most commands are hidden in a lower level of the menu and are not directly visible on the screen. As a result, individuals needed to have substantial knowledge about the Word menu bar and where specific editing or formatting functions could be found. This created more problems for the participants with disabilities because they had not used Word as often. Two additional challenges created problems with editing and formatting: commands and dictation are still frequently confused and it is still too difficult to select text that needs to be formatted.

Integrated Desktop Interaction. Previous studies tended to examine speech-based interactions in a single context, such as text generation, drawing, or desktop navigation. During the field study, participants used speech in multiple applications and contexts and frequently switched between applications (see table 3). This pattern suggests that providing a consistent speech-based solution, which can be used regardless of the application, is critical. Consistent dialogue design among applications would also improve usability. Finally, participants expressed a need for additional functions, such as the ability to use speech to emulate short-cut keys.

Web Browsing. All participants with disabilities expressed great interest in using speech commands to access information on the Web, including those who had limited computing experience. These participants spent substantially more time browsing the Web than they did using Word and Outlook combined. Currently, the Vista speech environment offers limited web browsing functions, allowing users to move between links or to say the text associated with a link to open a target page. Entering a URL is a significant problem for the participants. Some participants used the keyboard to enter URLs, even though this was a rather slow process while others accessed web sites using the 'favorites' function or desktop icons.

Online Communication. Interestingly, multiple participants without disabilities and one with a disability used speech for online communication activities via IM. They commented that it was faster and that they experienced fewer problems regarding editing and formatting. They were more tolerant of recognition errors in IM than when using Word or Outlook. The participant with a disability who used speech for IM specifically commented that he would be reluctant to adopt speech for writing text documents but preferred using it for instant messaging because it was faster.

Data Entry in non-text Environment. Participants also expressed the need to use speech to enter data in non-text environments such as an online calendar or a spreadsheet. Currently, speech interactions are rather difficult under both circumstances. In Excel, users cannot directly dictate a word or a number. There are no effective methods for selecting specific columns or rows for formatting or analysis purposes. Participants also experienced problems positioning the cursor in a specific cell because the cursor frequently jumped to items in the menu bar. Calendar applications typically allow users to dictate event descriptions, but navigation was problematic. Data entry in a calendar requires the user to select specific time slots, which was difficult using speech.

Entertainment Applications. Multiple participants used speech to play games. Participants without disabilities were more interested in battle games with rich graphics and sound effects. They adopted multi-modal interaction strategies such as using the mouse to specify targets and speech to 'double click'. Participants with disabilities did not use games during the field study, but they expressed interests in using speech to play games such as poker or puzzles. One participant with disability did use speech to manage the songs on his iPhone and computer.

5 Discussion

The most noteworthy contribution of these field studies is highlighting the variety of applications used by participants and, as a consequence, newly emerging requirements for speech-based interactions. The findings confirm that interest in speech-based interactions extend beyond generating text. Within the traditional domain of text generation, participants demanded more powerful and comprehensive editing and formatting capabilities that not only allow them to correct errors, but also allow them to effectively manipulate the appearance of the document. This represents a very different scenario as compared to what has served as the focus of most related research. At the same time, this is a challenging task given the capabilities available in most systems.

Interestingly, a few participants (from both studies) developed similar multi-modal interaction strategies. Participants with disabilities leveraged the residual control they had of their hands and arms. These participants could control the mouse movement to some extent but clicking and double clicking were very difficult, so they would use the mouse to position the cursor on the target and speech for clicking. Consistent with the findings of Halverson et al. [8], the participants who developed this multi-modal strategy tended to have more experience using computers. The participants with limited computer experience may benefit from specific instructions or designs that explicitly encourage the adoption of multi-modal strategies when appropriate.

The fact that participants used multiple applications and frequently switched between applications highlights the importance of keeping the speech dialogue design consistent across applications. This is challenging considering the vast array of commands and tasks that an individual may encounter, but the benefits could be significant.

The first field study indicated that some participants failed to retain the skills they had initially learned with speech-based text entry, and that the adoption of speech technology did not directly correspond to productivity.

Compared to previously reported studies of speech applications, these field studies have the strength of being conducted in real environments. However, this introduced a number of challenges and constraints that made the environment less than ideal. The computers with the speech-based solutions were provided by the researchers. Some participants completed tasks with their own computers, and found it hard to switch to other computers. In some cases, switching computers was not practical due to the use of applications that were not available on the computers we provided. For example, in the first study, eight of the ten participants (all five without disabilities) used both the computer we provided and their own computer. This limited the amount of information that could be collected and also had a potential impact on the interaction styles.

6 Conclusions

The studies provided first hand data on how users with and without physical disabilities used speech applications with personal computers in real environments over a prolonged period of time. The results suggest that participants without disabilities used a greater variety of applications than participants with disabilities, but those with disabilities were more satisfied with the speech-based solutions they experienced. The results also suggest that user requirements for speech-based interactions have evolved as technology has matured to include better support for formatting text documents as well as more effective support for other applications. The use of speech recognition to support informal communication is particularly interesting. Further research should address these newly emerging requirements for speech technologies.

Acknowledgments. This material is based upon work supported by the National Science Foundation (NSF) under Grant No. CNS-0619379 and National Institute on Disabilities and Rehabilitation Research (NIDRR) under grant number H133G050354. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF and the NIDRR.

References

1. Arnold, S., Mark, L., Goldthwaite, J.: Programming by Voice, VocalProgramming. In: Proceedings of ASSETS 2000, pp. 149–155 (2000)
2. Baber, C., Hone, K.: Modeling Error Recovery and Repair in Automatic Speech Recognition. International Journal of Man-Machine Studies 39, 495–515 (1993)
3. Christian, K., Kules, B., Shneiderman, B., Youssef, A.: A Comparison of Voice Controlled and Mouse Controlled Web Browsing. In: Proceedings of Assets 2000, pp. 72–79 (2000)
4. Dai, L., Goldman, R., Sears, A., Lozier, J.: Speech-Based Cursor Control Using Grids: Modeling Performance and Comparisons with Other Solutions. Behaviour and Information Technology 24(3), 219–230 (2005)
5. de Mauro, C., Gori, M., Maggini, M., Martinelli, E.: Easy Access to Graphical Interfaces by Voice Mouse (2001), Available from the author at: demauro@dii.unisi.it
6. Feng, J., Karat, C.-M., Sears, A.: How Productivity Improves in Hands-Free Continuous Dictation Tasks: Lessons Learned from a Longitudinal Study. Interacting with Computers 17(3), 265–289 (2005)
7. Feng, J., Sears, A.: Using Confidence Scores to Improve Hands-Free Speech-Based Navigation in Continuous Dictation Systems. ACM Transactions on Computer-Human Interaction 11(4), 329–356 (2004)
8. Halverson, C., Horn, D., Karat, C.-M., Karat, J.: The Beauty of Errors: Patterns of Error Correction in Desktop Speech Systems. In: Proceedings of INTERACT 1999, pp. 133–140. IOS Press, Amsterdam (1999)
9. Harada, S., Landay, J., Malkin, J., Li, X., Bilmes, J.: The Vocal Joystick: Evaluation of Voice-Based Cursor Control Techniques. In: Proceedings of ASSETS 2006, Portland, Oregon, pp. 197–204 (2006)
10. Harada, S., Wobbrock, J.O., Landay, J.A.: VoiceDraw:A Hands-Free Voice-Driven Drawing Application for People with Motor Impairments. In: Proceedings of the 9th International ACM SIGACCESS Conference on Computers and Accessibility, pp. 27–34. ACM, New York (2007)
11. Hubbell, T., Langan, D., Hain, T.: A Voice Activated Syntax-Directed Editor for Manually Disabled Programmers. In: Proceedings of ASSETS 2006, pp. 205–212 (2006)
12. Kamel, H., Landay, J.: Sketching Images Eyes-Free: A Grid-Based Dynamic Drawing Tool for the Blind. In: Proceedings of ASSETS 2002, pp. 33–40 (2002)
13. Karat, C.-M., Vergo, J., Nahamoo, D.: Conversational Interface Technologies. In: Jacko, J., Sears, A. (eds.) The Human-Computer Interaction Handbook, pp. 169–186. Lawrence Erlbaum and Associates, Mahwah (2003)

14. Karimullah, A., Sears, A.: Speech-Based Cursor Control. In: Proceedings of ASSETS 2002, pp. 178–185 (2002)
15. Manaris, B., Harkreader, A.: SUITEKeys: A Speech Understanding Interface for the Motor-Control Challenged. In: Proceedings of the 3rd International ACM SIGCAPH Conference on Assistive Technologies (ASSETS 1998), pp. 108–115 (1998)
16. Manasse, B., Hux, K., Rankin-Erickson, J.: Speech Recognition Training for Enhancing Written Language Generation by a Traumatic Brain Injury Survivor. Brain Injury 14(11), 1015–1034 (2000)
17. McNair, A., Waibel, A.: Improving Recognizer Acceptance through Robust, Natural Speech Repair. In: Proceedings of the International Conference on Spoken Language Processing, pp. 1299–1302 (1994)
18. Mihara, Y., Shibayama, E., Takahashi, S.: The Migratory Cursor: Accurate Speech-Based Cursor Movement by Moving Multiple Ghost Cursors Using Non-Verbal Vocalizations. In: Proceedings of ASSETS 2005, pp. 76–83 (2005)
19. Pieper, M., Kobsa, A.: Talking to the Ceiling: An Interface for Bed-Ridden Manually Impaired Users. In: CHI 1999, Extended Abstract, pp. 9–10 (1999)
20. Sears, A., Karat, C.-M., Oseitutu, K., Karimullah, A., Feng, J.: Productivity, Satisfaction, and Interaction Strategies of Individual with Spinal Cord Injuries and Traditional Users Interacting with Speech Recognition Software. Universal Access in the Information Society 1, 4–15 (2001)
21. Sears, A., Feng, J., Oseitutu, K., Karat, C.-M.: Speech-Based Navigation During Dictation: Difficulties, Consequences, and Solutions. Human Computer Interaction 18(3), 229–257 (2003)
22. Sporka, A., Kurniawan, S., Mahmud, M., Slavik, P.: Non-speech Input and Speech Recognition for Real-Time Control of Computer Games. In: Proceedings of ASSETS 2006, pp. 213–220 (2006)
23. Zhu, S., Ma, Y., Feng, J., Sears, A.: Speech-Based Navigation: Improving Grid-Based Solutions. In: Gross, T., Gulliksen, J., Kotzé, P., Oestreicher, L., Palanque, P., Prates, R.O., Winckler, M. (eds.) INTERACT 2009. LNCS, vol. 5726, pp. 50–62. Springer, Heidelberg (2009)

Metagnostic Deductive Question Answering with Explanation from Texts

John Kontos, Joseph Armaos, and Ioanna Malagardi

National and Kapodistrian University of Athens
University Campus,
Ano Ilisia 15771 Athens, Hellas
ikontos2003@yahoo.com, iarmaos@otenet.gr,
imalagar@di.uoa.gr

Abstract. The present paper presents a system called AMYNTAS for "metagnostic" deductive question answering from texts. This system can logically combine information from texts and answer questions generating explanations for its operation exhibiting "self-awareness". The deductions are performed directly with the natural language text without previous translation into a formal representation. The "metagnostic" effect is accomplished by representing and processing the state of linguistic processing and reasoning of the system. The system is implemented in Prolog and uses a text grammar to parse sentences that contain the information being sought. The system uses reasoning rules, lexicon, ontology, prerequisite knowledge and the history of its state. The system may easily be adapted to completely different domains such as biomedical texts and texts of the proofs of theorems of Euclidean geometry. An evaluation performed with real sentences from these two completely different domains gave satisfactory results of accuracy and facility of domain adaptation.

Keywords: deductive question answering, question answering from texts, inference with texts, self awareness, explanation.

1 Introduction

In the present paper we propose the use of the term "metagnostic" instead of the term "metacognitive" used in [1] for computer systems that exhibit self-awareness. The reason for using this neologism is that we want to avoid any confusion that may arise from the psychological connotations of the adjectives "metacognitive" and "self-aware" that may be connected with human metacognition [2].

The nature of the behaviour of this kind of computer systems is manifested by the fact that the explanations they generate refer explicitly to the history of their linguistic processing and reasoning.

Our system AMYNTAS (Automatic Metagnostic YpologistikoN Trainable Answering System) generates explanations either directly or after applying deductive inference and using meta-information concerning the state history of the system. Deduction is accomplished without translation of the texts into any formal representation following the ARISTA method [4],[5].

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 72–80, 2011.

2 Related Work

The system AMYNTAS presented in the present paper answers questions by combining information from unstructured natural language texts and justifies answers to the questions by generating an explanation that exhibits behaviour inspired from human self-awareness. It is a system related to the system AROMA [5] that was listed in [3] as the sole inference-based biological question answering system in existence at the time of the review. The system AROMA used analysis of textual rhetoric relations for model-based question answering for explanation generation that AMYNTAS does not do at present. However AROMA lacked the main capability of AMYNTAS for generating explanations inspired from human self-awareness.

To the best of our knowledge the only modern systems in existence and current development that attempt to exhibit behaviour inspired from human self-awareness are the CASSIE system [6] and the EPILOG system [7]. However these two systems extract information solely from formal knowledge bases and they do not perform inferences and generate explanations directly from text as our system does.

3 The System Implementation and Operation

The question answering system AMYNTAS presented here was implemented in Prolog and consists of six modules implemented as separate programs totalling about 50 pages of code. These modules communicate through some temporary files that store intermediate results. The six modules are: the question processing module, the text pre-processing module, the ontology extraction module, the shallow parsing or text chunking module, the question answering module and the metagnostic processing module. The architecture of the system AMYNTAS is shown in Fig. 1.

The question processing module extracts information from the input question. The information extracted is a list consisting of the entities mentioned in the question and the relation that connects them. For example in the question "what influences p53" the entity is p53 and the "blank" entity standing for the unknown entity that is sought and the relation is "influence".

The text pre-processing module represents each word of a sentence as a fact with three arguments the first being the word itself, the second being the identifier of the sentence and the third being the position of the word in the sentence counting from left to right.

The ontology extraction module locates linguistic patterns in the input text corpus that may be used to extract automatically meronymic and taxonomic knowledge that may be used at question answering time.

The shallow parsing or text chunking module locates a verb related to the relation contained in the question and extracts the two substrings of the text sentence being analyzed that appear to the left and the right of the verb and end at some stop-word or punctuation mark.

The question answering module finds the answer to the question from the pre-processed text. The question answering module accepts questions that potentially require the combination of facts with the use of prerequisite knowledge for answering them. The prerequisite knowledge available to our system includes ontological

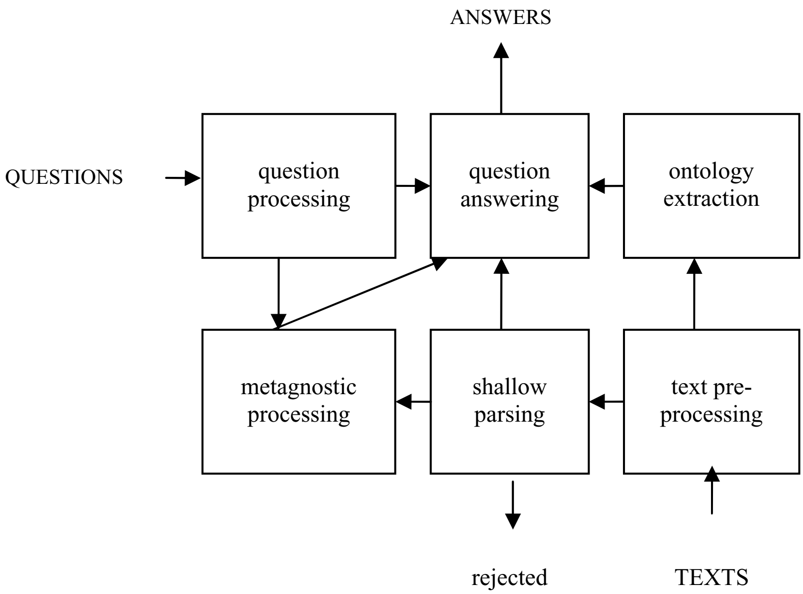


Fig. 1. The architecture of the AMYNTAS system

knowledge, inference rules and synonyms of the named entities involved of the domain which used in order to combine two or more facts mentioned in the text corpus.

At question answering time three looping operations are taking place. The basic loop concerns the search for an entity in a chunk related to the relation of the question. The second loop concerns the transformation of the list obtained from the question by following a particular strategy from the explicit list given to the system. The third loop searches for chains of facts using the matching of named entities occurring in the right part of one fact and the left part of another fact.

Table 1. The roles of the arguments of the predicate "strategy"

Variable Name	Role
Synonym	flag determining whether synonyms will be used or not
LSide	flag determining searching for a synonym at the left side of the verb
RSide	flag determining searching for a synonym at the right side of the verb
Inversion	flag determining whether the two sides of the text sentence i.e. to the left and to the right of the verb will be inverted or not
Relstrategy	specification of the relation denoted by the verb of the question
Pk	flag determining whether prerequisite knowledge will be used or not for answering the question
Strategy	the name of the strategy being applied

The metagnostic processing module collects information during the execution of all three loops and stores it either in an internal database or an external file. The metagnostic processing module also processes the history of state of the system in order to supply data for the generation of the explanations.

The strategies for processing the lists obtained from the transformation of the questions and the prerequisite knowledge provided is defined by the predicate: "strategy(Synonym, LSide, RSide, Inversion, Relstrategy, Pk, Strategy)".

The roles of the variables for each argument of the predicate "strategy" are given in Table 1. An example will be used to illustrate the operation of our system using text fragments from two abstracts found in PubMed that concern the very important proteins p53 and mdm2 for carcinogenesis. This operation includes the question answering process of the system and the manifestation of "self-awareness" or "metagnosis" in the explanation generated.

The first abstract consists of six sentences from which the following two sentence fragments were selected automatically using the entities p53 and/or mdm2 as keywords:

[1] "The p53 protein regulates the mdm2 gene"
[2] "regulates both the activity of the p53 protein"

The second abstract consists of seven sentences from which two were selected from which the following two fragments were selected automatically using again p53 and/or mdm2 as keywords:

[3] "The mdm2 gene enhances the tumorigenic potential of cells"
[4] "The mdm2 oncogene can inhibit p53_mediated transactivation"

The question answering module answers questions deductively by combining sentences with common entities. More formally given two sentences "A rel1 B" and "B rel2 C" it finds that "A rel3 C" where what rel3 is generated depends on some prerequisite knowledge about the result of combining the relations rel1 and rel2.

A form of the questions accepted by the question processing module is "What is <verb> by <entity> ?" where <verb> stands for the passive form of a member of the group of verbs known to our system as significant verbs and <entity> stands for a noun denoting a named entity.

Examples of named entities are proteins and genes such as p53 and mdm2. Examples of verbs are the verbs regulate, enhance, influence and inhibit.

Another form of question that is accepted by our system is "Why <entity1> <verbphrase><entity2>" where entity1 and entity2 are named entities and <verbphrase> is standing for a relation and may stand e.g. for "is equal to".

The example question "What influences p53" is input to the system and it is answered using the above sentence fragments. The relation "influences" is not symmetric and therefore none of the above mentioned 16 strategies is useful for answering this question. Another strategy (strategy17) is used which is appropriate for questions with asymmetric relations. The answer in this case after applying strategy17 is "p53 influences p53" i.e. the system detects a causal chain that forms a closed feedback loop based on the sentence fragments [1] and [4]. The verb "influence" is given to the system as common hypernym of the two verbs or relations "regulates" and "inhibits".

The entity mdm2 is detected in the two sentence fragments “The **p53** protein **regulates** the **mdm2** gene” and “The **mdm2** oncogene can **inhibit p53** mediated transactivation” giving the causal chain “regulates+inhibit”.

Given that a deduction rule is also known to the system stating that “if rel1=regulate(s) and rel2=inhibit(s) then rel3=influences” the chain may be conflated to rel3 i.e. “influences”. This deduction rule enables the system to compose the two facts and generate an answer based on the recognition of the causal chain formed by the two sentence fragments [1] and [4].

The analysis performed by the question processing module extracts the components of the question such as the verb denoting a relation namely “influenced” in the example question and the entity such as “p53” for the example question. These components are stored in the internal data base.

4 The Application to Two Unrelated Illustrative Example Domains

The detailed operation of the system and the explanations generated when answering questions by extracting information form texts belonging to two unrelated illustrative example domains is presented in this section. The two illustrative domains are a biomedical and a geometrical domain.

The biomedical example uses a subset of the example text briefly analysed above. A more detailed analysis for this example is given in this section.

In the biomedical domain the system is able to identify the sentences of the input text that contain causal information about the p53 protein. Two example sentences with their identification numbers 1411 and 1421 are:

<1411>: <the p53 protein regulates the mdm2 gene>
<1421>: <the mdm2 oncogene can inhibit p53 mediated transactivation >

The example question “What influences p53” is also input to the system.

When the system processes the first sentence the explanation as verbatim generated by our system runs as follows:

I searched the input for the entity <p53>.
in the chunk <the p53 protein> of the sentence <1411>.
Since its first token is not an entity, I tested the rest of the tokens.
Repeating the search I found that
the entity <p53> is one of the tokens of the chunk <the p53 protein>
which is the chunk to the left of the verb of the sentence <1411>.

I found that the chunk to the right of the verb of the sentence <1411>
is the chunk <the mdm2 gene > and
Since its first token is not an entity, I tested the rest of the tokens.
Repeating the search the entity <mdm2> is one of the tokens of the chunk <the mdm2 gene >
which is the chunk to the right of the verb of the sentence <1411>.
I found that the entity <p53> is included in the sentence
<1411>:<the p53 protein regulates the mdm2 gene >

When the system processes the second sentence namely the sentence
<1421>: <the mdm2 oncogene can inhibit p53 mediated transactivation >

the explanation as verbatim generated by our system runs as follows:
I found that the entity <mdm2> is one of the tokens of the chunk <the mdm2 oncogene can>
which is the chunk to the left of the verb of the sentence <1421>.

I found that the chunk to the right of the verb of the sentence <1421>
is the chunk <p53 mediated transactivation >
and <p53> is the first token of the chunk
the entity <p53> is one of the tokens of the chunk
<p53 mediated transactivation >
which is the chunk to the right of the verb of <1421>.

And now the system detects a chain formed by <1411> to <1421> and back to <1411> and explains as follows:
In my effort for answering the question "What influences p53"

I found from the following sentences:
<1411>: <the p53 protein regulates the mdm2 gene >
<1421>: <the mdm2 oncogene can inhibit p53 mediated transactivation >
That p53 influences p53
because influence is hypernym of regulate and inhibit.
From which it follows that there is a loop for <p53>.

In the geometrical domain the system is able to identify the sentences of the input text that logically justify a statement included in the English translation [8] of the text of the proof of the first Proposition from Euclid's Elements. For illustration of the answering of questions from the text of the proof of the first Proposition of the Elements is sketched below. The prerequisite knowledge we use involves various kinds such us Euclid's common notions, postulates and definitions. The first Proposition of the Euclid's Elements may be stated in English: "construct an equilateral triangle on a given finite straight line ".

The equilateral triangle abc is constructed by the points of intersection of two equal circles and their centres a and b where these centres lie on each others circumference and the points a and b are the two endpoints of the given finite straight line. The Euclid's Elements text includes a proof that the triangle constructed in this way is equilateral.

We use the text of this proof as a text base for answering deductively questions and generating explanations concerning the justification of statements present in the proof text. The meaning of the geometrical entities involved in this Proposition may be understood by using the diagram shown in Figure 1.

Three illustrative questions are answered by the system by analyzing automatically the corresponding Euclidean proof as follows:

Question 1: "why is side ab equal to side ac?"
Answer 1: "because they are radii of the same circle c1"
Question 2: "why is side bc equal to side ab?"

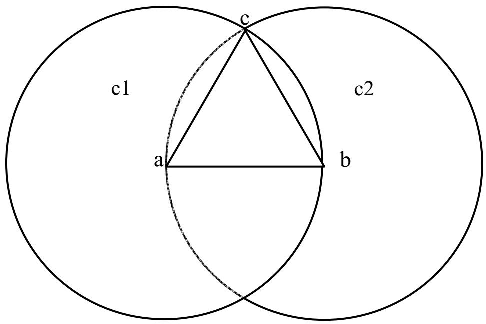


Fig. 2. The equilateral triangle abc constructed by the straight segment ab and by the point of intersection c of the two equal circles c1 and c2 with centres a an b

Answer 2: "because they are radii of the same circle c2"
Question 3: "why is side ac equal to side bc?"
Answer 3: "because each of ac and bc are equal to ab"

These answers are only partial. The full answers contain explanations too.

The operation of the system in the case of the Geometry domain will be first explained by showing how Question 1 namely "why is side ab equal to ac" is processed by the system. The answer to this question is found by following strategy1. This strategy uses the expression of the question as it stands without any transformation. It should be noted that no prerequisite knowledge is used when strategy1 is followed. The answering of other questions may follow other strategies that apply various transformations of the questions such as inversion of its terms and replacement of an entity by each synonym. More specifically the inversion of "ac is equal to ab" is "ab is equal to ac" and the synonym of "ac" is "ca".

The name of the current strategy used is stored in an internal data base by the metagnostic sub-module so that the system is aware of the strategy being followed at each step. The chains of the proof steps as mentioned in the proof text are also stored so that they may be used at explanation time.

The original or transformed question is compared with the sentences of the text of the proof. When a match is found an answer is generated giving an explanation that justifies the answer. The transformed parts of the question are retrieved from the internal data base. The program finds the position of the left part of a proof step that matches the left part of the transformed question and continues by matching the relation and the right part of the question with some step of the proof.

Following the satisfaction of the matching operation the system tries the first strategy in the list and a sentence of this form is not found in the proof text. For this reason the system will try all other strategies in order to be able to answer the question. These strategies involve the replacement of entities with their synonyms, the inversion of the entities since the relation equal is symmetric and the use or not of prerequisite knowledge. Concerning the present question the strategy that succeeds is strategy16. This strategy involves the replacement of the two entities ac and bc with their synonyms ca and cb respectively, the inversion of the two entities and the use of prerequisite knowledge for answering the question. The prerequisite knowledge used is:

«things which are equal to the same thing are also equal to one another» which is a statement called “common notion” in Euclid’s elements.

The answer generated by the system reports all the information it used to find the answer in order to justify it.

This answer runs verbatim as follows:

Your question is:
why <ac> is <equal> <to> <bc> ?
I analysed this question using strategy : <strategy16>
which involves : mesynonyms, left, right, meinversion, equality, pk
My answer using Euclid's proof text offers the justification:
because
<bc> is equal to <ab> and
<ac> is equal to <ba>
Furthermore because
<ab> and <ba> are synonyms, which means that
<ab> and <ac> are equal to each other
and because of the common notion
< things which are equal to the same thing are equal to one another>.

5 The System Evaluation

The performance of the system was evaluated using a set of 127 sentences obtained from the PubMed Data Base that were selected from the titles of papers. The criteria of selection were that they contain the name of the protein p53 and the influence verb “enhance”. These 127 sentences are all that were found from PubMed on October the 7th 2008. This set constituted an input text to the system and was checked by the biologist Dr. Ourania Kosti. The results of evaluation of the parsing module quantified in terms of “precision” and “recall” are: Precision= 80% and Recall= 94 %, where precision and recall were computed by the formulae:

Precision=

correct answers generated (True Positives) Divided by # total answers extracted (True Positives + False Positives)

Recall=

correct answers generated (True Positives) Divided by # total correct answers identified by the biologist (True Positives + False Negatives).

The flexibility of the system was proven by showing that it may very easily be adapted to a completely different domain such as the proofs of Euclidean geometry. The only changes necessary were the enrichment of the lexicon and the ontology so that it includes the entities, the concepts and the relations of the new domain.

6 Conclusion

The present paper presented a system for metagnostic question answering from texts. Two simpler versions of the system were presented in [1] and [9]. The present system can extract information from texts and answer questions and provide explanations that exhibit self-awareness. This was accomplished by representing and processing its state of automatic text analysis which includes deductive reasoning performed directly with the text without any previous formal representation of its sentences. The explanations generated are offered in order to convince the user of the correctness of its output. Explanations are also provided for the sentences rejected because they do not state a relation of interest of the entity of interest and another entity.

The system was implemented in Prolog it consists of the six modules namely the question processing module, the text pre-processing module, the ontology extraction module, the shallow parsing or text chunking module, the question answering module and the metagnostic processing module and uses a question grammar, a lexicon, an ontology and a history of the state of the system.

The system presented in the present paper constitutes a system appropriate for metagnostic question answering in different domains. Future extensions of the system may include a feedback loop that suggests corrections and/or expansions of its knowledge bases in order to be able to process a larger variety of text sentences. Such a feedback loop will be based on the self awareness capability already exhibited by the present system.

The explanations generated by the system may also be used for the support of handicapped persons when they fail to communicate with the system due to their own reduced capability of phrasing a question.

Acknowledgement. We thank the biologist Dr. Ourania Kosti for the invaluable help that she generously provided for the evaluation of our system.

References

1. Kontos, J., Armaos, J.: Metacognitive Question Answering from Euclid's Elements Text. In: Proceedings of HERCMA 2007, Athens, Hellas (2007)
2. Maniadakis, M., Tani, J.: Acquiring Rules for Rules: Neuro-Dynamical Systems Account for Meta-Cognition. Adaptive Behavior 17(1), 58–80 (2009)
3. Athenikos, S.J., Han, H.: Biomedical Question Answering: A survey. Comput. Methods Programs Biomed. (2009) (in press)
4. Kontos, J.: ARISTA: Knowledge Engineering with Scientific Texts. Information and Software Technology 34(9), 611–616 (1992)
5. Kontos, J., Malagardi, I., Peros, J.: Question answering and rhetoric analysis of biomedical texts in the AROMA system. In: Proceedings of the 7th HERCMA: Hellenic European Conference in Computer Mathematics and Its Applications, Athens, Greece (2005)
6. Shapiro, S.C., Rapaport, W.J., Kandefer, M., Johnson, F.L., Goldfain, A.: Metacognition in SnePS. AI Magazine 28(1), 17–31 (2007)
7. Morbini, F.: Computational Infrastructure for a Self-Aware Agent. Ph.D. thesis, University of Rochester, Rochester, New York (October 2009)
8. Heath, T.L.: The Thirteen books of Euclid's Elements. Dover, N.Y. (1956)
9. Kontos, J., Armaos, J., Malagardi, I.: Metagnostic Question Answering from Biomedical Texts. In: HCI, San Diego, pp. 332–336 (2009)

Collecting an American Sign Language Corpus through the Participation of Native Signers

Pengfei Lu[1] and Matt Huenerfauth[2]

[1] The City University of New York (CUNY)
Computer Science Doctoral Program, Graduate Center
365 Fifth Ave, New York, NY 10016, USA
[2] The City University of New York (CUNY)
Computer Science Department, Queens College
65-30 Kissena Blvd, Flushing, NY 11367, USA
pengfei.lu@qc.cuny.edu, matt@cs.qc.cuny.edu

Abstract. Animations of American Sign Language (ASL) can make more information, websites, and services accessible for the significant number of deaf people in the United States with lower levels of written language literacy – ultimately leading to fuller social inclusion for these users. We are collecting and analyzing an ASL motion-capture corpus of multi-sentential discourse to seek computational models of various aspects of ASL linguistics to enable us to produce more accurate and understandable ASL animations. In this paper, we will describe our motion-capture studio configuration, our data collection procedure, and the linguistic annotations being added by our research team of native ASL signers. This paper will identify the most effective prompts we have developed for collecting non-scripted ASL passages in which signers use particular linguistic constructions that we wish to study. This paper also describes the educational outreach and social inclusion aspects of our project – the participation of many deaf participants, researchers, and students.

Keywords: American Sign Language, animation, accessibility technology for people who are deaf, data collection, social inclusion, motion capture.

1 Introduction

American Sign Language (ASL) is the primary means of communication for one-half million deaf people in the U.S. [1], and ASL and English have distinct vocabulary and word-order. Due to educational and language exposure reasons, many deaf adults in the U.S. have relatively low levels of written English literacy [2], making it difficult to read English text on websites or other sources. Animations of American Sign Language (ASL) make information and services accessible for these individuals.

We are studying how to create high-quality ASL animations by analyzing recordings of people. We are in the middle of a 5-year study to collect, annotate, and analyze an ASL motion-capture corpus of multi-sentential discourse. Our goal is to create an annotated collection of ASL movement data using video camcorders and

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 81–90, 2011.

motion capture equipment, and we are seeking computational models of various aspects of ASL linguistics. These models will help us to synthesize more understandable ASL animations, enabling their use in applications for deaf individuals with low literacy. For our research, we want the stories we collect to contain certain linguistic phenomena but not others. This paper explores different ways we have prompted people to perform ASL stories in order to optimize this.

Section 2 describes the corpus collection goals of our study and other related corpus building projects. Section 3 summarizes our previous work on eliciting and collecting our corpus using alternative prompting strategies and the main results. Section 4 presents our corpora collection, annotation procedure, and the experiments in the second year. Section 5 contains the results, conclusions, and future research plans. Section 6 discusses the participation of people who are deaf in this research.

2 Our Corpus Collection Goals

For our corpus, we record native ASL signers performing short stories or passages in our laboratory while being videotaped and while wearing motion-capture gloves, an eye-tracker, a head-tracker, and a set of sensors on a special bodysuit. This equipment records the signer's handshape; hand location; palm orientation; eye-gaze vector; and joint angles for the wrists, elbows, shoulders, clavicle, neck, and waist. Three video cameras record front, side, and face-close-up views. Details of our studio configuration appear in [3]. Another native signer (called the "prompter") sits behind the front-view camera to converse with the performer being recorded. Signers tend to perform more natural ASL performances in this type of conversational setting [4]. Because we want to record natural ASL performances (to use as a basis for our research), we do not pre-script the passages to be performed; however, it is necessary to give the performer prompts to encourage them to sign about a particular topic for some period of time. As discussed below, we also want to encourage the performer to use particular linguistic constructions (and not others) in their ASL signing (without giving them a specific script to perform). Therefore, we have experimented with various forms of prompting to elicit ASL passages that are optimally suited to our research needs. Identifying a set of such prompts is the focus of this paper.

After we record a human signer performing a multi-sentence passage in ASL, then our team of ASL linguistics experts watch the recording and create a timeline of the signs performed. The experts also analyze the performance to note various linguistic constructions and other details about the performance, and this information is also added to the timeline for each story. Details are described in [4]. This linguistic annotation of each story facilitates our later analysis and use of the data in the corpus.

Several researchers have collected video-based corpora of sign language, e.g. [5][6][7], or short sign language recordings via motion-capture, e.g. [8][9]. However, our project is the first to record a large corpus of sign language passages while using motion-capture equipment. Previous researchers have also designed schemes for annotating the referential use of signing space [10] on a timeline, but our project is the first to analyze this linguistic use of signing space in a motion-capture corpus.

2.1 Spatial Reference Points in Our ASL Corpus

Signers associate people, things, and concepts under discussion with 3D locations around them in space [5][11][12]. The typical way in which a signer establishes a spatial reference point (SRP) is by pointing to a location around them in space immediately before or after mentioning someone or something for the first time. Signers may set up several of these SRPs around them in space during a conversation. After being established, these points are later used by the signer in the following ways:

- The signer will point at the location to refer to the person or thing.
- Some signs change how they are performed to aim at these SRPs (e.g., the motion path of some verb signs goes from their subject toward their object).
- Signers may aim their eyes or head at these SRPs for linguistic reasons.

SRPs are frequently used and are essential to *human* ASL performances, and they are also important for producing good quality ASL *animations*. Huenerfauth [13] found that native signers' comprehension of ASL animations improved when the animations included association of entities with locations in the space and the use of verbs whose motion paths were modified based on these locations. Our research focuses on adding these capabilities to ASL animation synthesis technologies; we believe that mathematical functions of verbs' motion paths can be induced from 3D motion data we are collecting from human signers performing those verbs. We are also analyzing the ASL passages we are collecting to learn when to associate entities with 3D locations, where to place them, and how these locations affect sign movements. The models of ASL spatial use we learn will be embedded into ASL animation software to produce more natural looking and understandable animations. Therefore, it is desirable that the stories and passages we collect in our corpus contain many examples of signers setting up SRPs and using them in a story or passage.

Fig. 1 shows a sample of some of the linguistic information for one story in our corpus; it is a timeline of an ASL passage discussing a girl using her computer. For our project, we add various linguistic information to a timeline corresponding to the video and motion-capture recording of each story collected. Fig. 1 shows only a subset of that information: the sequence of signs, the establishment of SRPs (indicated by a line on the "SRP#1 Establishment" row of the timeline), and the references made to those SRPs during the passage (indicated on the "SRP#1 Reference" row of the timeline). In this case, the first time that the signer points to a location in 3D space around his body, he establishes an SRP at that location to represent the girl being discussed; this SRP is referred to again later in the passage when the signer performs another "POINT" sign. A loose translation of the passage in Fig. 1 would be: "Wow. There was this 12-year-old girl, and she was on the Internet typing…"

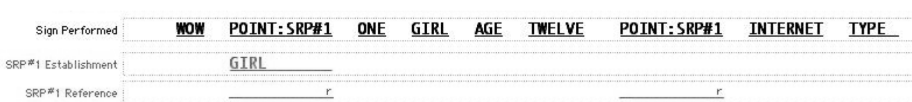


Fig. 1. Example of a timeline from a story from our corpus that contains an SRP

2.2 Classifier Predicates in Our ASL Corpus

Classifier predicates (CPs) are a linguistic construction in ASL that also uses the space around the signer's body – but in a different way than SRPs. CPs are complex signs in which the signer creates movement for the hands (or sometimes the body) to indicate the spatial arrangement, size, shape, or movement of people/objects in a 3D scene being described [14]. During CPs, entities under discussion are associated with locations in space around the signer, but unlike SRPs, during CPs, the arrangement reflects a real-world 3D configuration or arrangement of objects. CPs are not our current research focus, and because they lead signers to use space around their bodies in a different way than SRPs, we don't want to record stories that contain a lot of CPs in our corpus. Building this corpus (recording people performing ASL and then linguistically analyzing the recordings) is very time-consuming, we therefore want to optimize the stories that we collect so that they contain primarily SRPs and not CPs.

2.3 How We Evaluate Our Prompting Strategies

Research projects collecting video recordings of sign language for linguistic study have used scripts or various prompting strategies to encourage signers to perform stories or sentences that contain specific linguistic phenomena of interest to the researchers [5][6][7][15]. We have adopted some of these prompts for our project and have invented others; sections 3 and 4 discuss how we evaluate the success of our prompting strategies. For our research, an ideal ASL passage to be collected would:

- *Be long enough to allow for establishment of SRPs.* If a story is too short, then the signer might not set up many SRPs or refer to them in the story. So, we will count the length of the stories we collect – as measured in seconds of time or in the total number of signs performed. By measuring the length of the stories collected using each of our prompting strategies (details in sections 3 and 4), we will be able to determine which prompting strategy is most effective.
- *Contain several SRPs established by the signer.* Collecting stories in which signers establish large numbers of SRPs around them in space can sometimes be difficult; so, we will count the number of SRPs established during each story we collect to measure the effectiveness of our different prompting strategies.
- *Contain many pointing signs or verbs that refer to SRPs.* With many examples of these spatial references (SRs), we will be able to study diverse forms of spatial use and reference in ASL signing.[1] Some linguists regard a signer pointing to himself (to say "I" or "me") as a form of spatial reference; so, we'll also count the number of these "first person references" in our data as a separate total.
- *Contain as few CPs as possible.* So, we will count the number of CPs that occur during the stories we collect; unlike the other items we discussed above, we would prefer to see a small number of CPs in the stories collected in the corpus.

[1] When a signer points to a location in space the first time, this establishes an SRP for some entity under discussion; we refer to this as an *SRP establishment*. When the signer points to the SRP again later, then this is a *spatial reference* to a previously established SRP. We also count the first pointing gesture that established the SRP as a spatial reference.

3 Summary of Our Previous Study: Year 1 of Our Project

During the first year of our project, we recorded and annotated 58 ASL passages from 6 signers (approximately 40 minutes of data). For this data collection, the prompter behind the camera used 9 different prompting strategies to elicit ASL passages:

- Tell a story: Invent a story using this topic: “If I had a genie that could grant three wishes, I’d…”
- Children’s book: Read this short children’s book, and then explain the story as you remember it.
- Repeat Conversation: Watch this 3-minute video of an ASL conversation or of a captioned English conversation, and then explain what you saw.
- Wikipedia Article: Read this 300-word Wikipedia article on “The History of Racial Segregation in the United States,” and now explain/recount the article.
- Recount Movie/Book: Tell me about your favorite movie or your favorite book.
- Compare (not people): Compare two things: e.g. Mac vs. PC, Democrats vs. Republicans, high school vs. college, Gallaudet University vs. NTID, travelling by plane vs. travelling by car, etc.
- Personal Intro/Info: Introduce yourself, describe some of your background, hobbies, family, education, etc.
- Hypothetical Scenario: What would you do if: You were raising a deaf child? You could have dinner with any two famous or historical figures?
- Compare (people): Compare two people you know: your parents, some friends, family members, etc.

In a prior paper [4], we presented the characteristics of the stories that signers performed in response to each of these different prompts. Our results from analyzing the stories collected during year 1 are summarized here: The “tell a story,” “children’s book,” and “repeat conversation” prompts elicited ASL stories with high CP/SR ratios (undesirable). These prompts that related to spatially/visually descriptive topics led to many CPs performed by signers. The “wikipedia article” and “recount movie/book” prompts yielded long story lengths, high number of SRPs established, and modest CP/SR ratios (desirable). While they elicited shorter passages, the “compare” and “personal intro/info” prompts also yielded stories with low CP/SR ratios (desirable). The “personal intro/info,” “tell a story,” and “hypothetical scenario” prompts led to many first-person references; the signer often discussed himself in these passages.

4 Our New Study: Year 2 of Our Project

The analysis of the different prompting strategies in year 1 of our project (discussed above) guided our data collection procedure in year 2. This section presents a new study we conducted to analyze our revised set of prompts used during this second year of the project. Specifically, we stopped using the “tell a story,” “children’s book,” and “repeat conversation” prompts from year 1, and we continued to use the “wikipedia article,” “recount movie/book,” “compare,” and “personal intro/info.” Each of these is explained in Table 1. In an effort to encourage signers to tell even longer stories, use more SRPs, and use fewer CPs, we tried several new prompting strategies during year 2 (which are also listed in Table 1). As we gained additional experience at recording signers using motion-capture equipment and analyzing stories, we were able to collect a larger set of stories in this second year of the project. The number of stories collecting using each type of prompt is also listed in Table 1. One of the new prompting approaches ("photo page") involved showing a page of images to a participant to encourage him to tell a story; an example of a page of images similar to those used in our study is shown in Fig. 2. Typically, the photographs included popular celebrities, athletes, or politicians who were currently in the news.

Table 1. Types of prompts used in Year 2 and number of stories of each type collected (N)

Type of Prompt	N	Description of This Prompting Strategy	Year This Was Used
News Story	12	Please read this brief news article (about a funny or memorable occurrence) and recount the article.	Year 2
Opinion / Explain Topic	5	Please explain your opinion on this topic (given) or explain the concept as if you were teaching it to someone.	Year 2
Compare (not people)	19	Compare two things: e.g. Mac vs. PC, Democrats vs. Republicans, high school vs. college, Gallaudet University vs. NTID, travelling by plane vs. travelling by car, etc.	Year 1 & 2
Compare (people)	2	Compare two people you know: your parents, some friends, family members, etc.	Year 1 & 2
Personal Intro/Info	8	Introduce yourself, describe some of your background, hobbies, family, education, etc.	Year 1 & 2
Personal Narrative	3	Please tell a story about an experience that you had personally.	Year 2
Photo Page	5	Look at this page of photos (of people who are in the news recently) and then explain what is going on with them.	Year 2
Recount Movie Book	9	Recall a book you're read recently or a movie you saw, and then explain the story as you remember it.	Year 2
Wikipedia Article	3	Read a brief Wikipedia article on some topic and then explain/recount the information from the article.	Year 1 & 2

Fig. 2. Example of what the page of photos looked like for the "photo page" prompts

In the second year of our study, we recorded and annotated 66 ASL passages from 3 signers (approximately 75 minutes of data). As was done in year 1 of the project, we used a set of prompts to elicit unscripted multi-sentential single-signer passages. The prompts used in year 2 are listed in Table 1. As was done in year 1 of the project, our team of ASL linguistics experts analyzed the stores collected to produce a timeline of each performance that includes the sequence of signs, the establishment of SRPs, the references to SRPs, the use of CPs, and other linguistic phenomena of interest to our research. In order to evaluate the set of prompts used during year 2 of our project, we calculated the average passage length (measured in the number of signs performed or the number of seconds); the results are shown in Fig. 3. We would prefer longer stories in our corpus because this increases the opportunity for the signer to establish several SRPs and to refer to them again. Further, we have found it easy to record very short stories from signers during our recording sessions; finding prompts that encourage a signer to perform a longer multi-sentence passage are therefore valuable to identify. Note: Error bars in Figures 3, 4, 5, and 6 indicate the standard error of the mean for each value.

Fig. 4 displays the average number of classifier predicates per second and the average number of spatial references per second in each type of ASL story. As discussed in section 2, we would prefer stories in our corpus with a low number of CPs and a high number of SRPs. The "opinion explain topic," "news story," and "compare people" prompts all led to high CP/SR ratios and long story lengths.

Fig. 5 displays the average number of the spatial reference points established per story – for each prompting strategy. For example, if a signer sets up three different points in space around their body to represent three entities under discussion in a passage, then we would say that such a story would have 3 SRPs *established*. If the signer continues to refer to these entities multiple times throughout the story (i.e., pointing to these locations in space again and again during the story), then the number of *spatial references* in the story would be much higher.

Fig. 6 displays the average number of 1st-person references per second in the stories collected using each prompting strategy. As discussed in section 2, we present the results for first-person references separately because some ASL linguists would consider this a form of spatial reference and some would not. We would say that a first-person reference occurs whenever the signer is pointing to himself. The "personal narrative" and "personal" prompts led to a high number of 1st-person references; this is not surprising since these prompts led signers to discuss themselves.

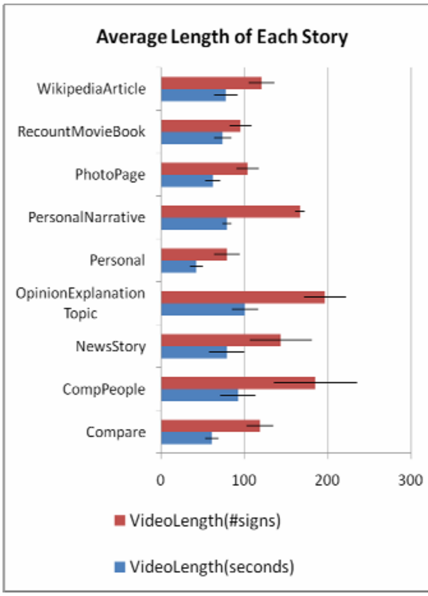


Fig. 3. Length of the ASL stories collected

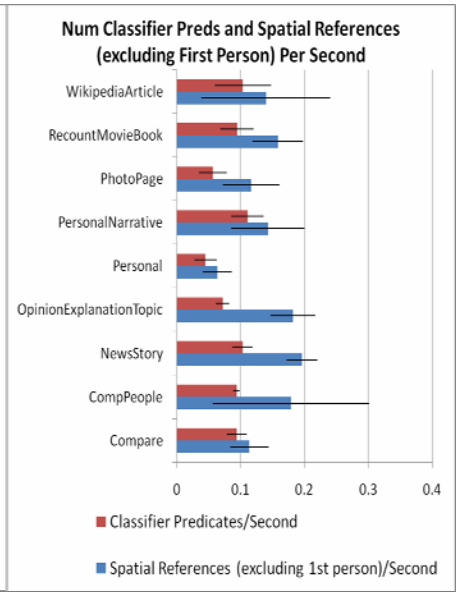


Fig. 4. Number of classifier predicates and spatial references per second in each story

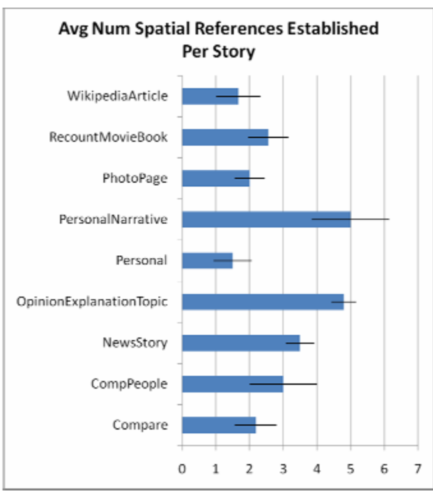


Fig. 5. Number of spatial reference points established in each story

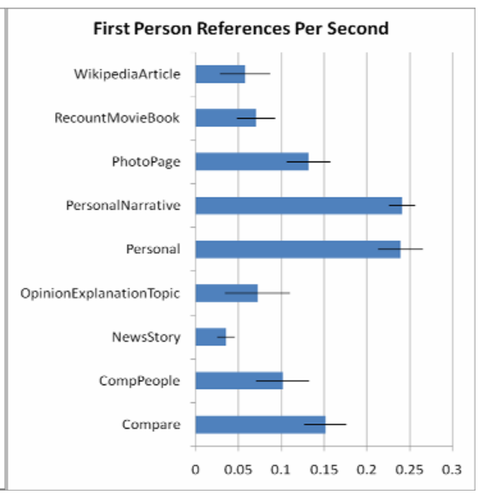


Fig. 6. Number of first-person references in each ASL story

5 Discussion and Future Work

We were pleased with the results of the "opinion/explanation topic," "compare people," "news story," and "recount movie/book" prompts, which led to good story lengths and high SR/CP ratios. The "opinion/explanation topic" and "news story" prompts led to large numbers of SRPs established in the signing space. In year 3 of the project, we intend to use a larger proportion of these prompts during our data collection. While the results of our study have immediate benefits for our own research project, other linguistics and computer science researchers who are conducting sign language data collection will benefit from our comparison of various forms of prompting. This research therefore has benefits for sign language animation synthesis and sign language recognition research, which has accessibility benefits for people who are deaf. We are not aware of other systematic analysis of the benefits of various forms of linguistic prompting used in a sign language data collection study.

As we gather and analyze our corpus of ASL performances, we are beginning to analyze this data to study how human ASL signers set up SRPs in space, how they perform ASL verbs whose motion paths change based on SRP arrangements, and other details of sign language performance (e.g., the timing and speed of signs). We intend to incorporate these findings into our ASL animation synthesis technology to produce more natural and understandable animations of ASL [3][4][13].

6 Inclusion of Deaf Participants, Students, and Researchers

As discussed in section 1, ASL animation technology has the potential to make more information accessible to people who are deaf that have lower levels of English literacy – ultimately leading to fuller social inclusion of these individuals. Our research project also has a more immediate impact in this regard due to the extensive participation of people who are deaf in the various stages of the research process. Based on New York City, we advertise through Deaf community websites and other local resources to identify participants to come to our lab to be recorded for the corpus and to participate in experiments evaluating the quality of ASL animations.

Several deaf and signing researchers also participate in our project (including deaf high school and undergraduate students). Each summer, two or three students from local deaf schools or mainstream programs in the greater New York City area participate in three-month research experiences at our lab. In addition, we have also hosted a deaf undergraduate student visiting from Gallaudet University and another undergraduate student majoring in linguistics with excellent fluency in ASL for summer research experiences at the lab. Further, a graduate-level deaf research assistant also coordinates the projects at the lab throughout the year. Further details of the participation of deaf students and researchers in our project are described in [16]. Our goal is for these students to gain knowledge and practical experience working in a research laboratory and insight into the process of applying for and succeeding in future study and careers in scientific research – ultimately leading to fuller inclusion of people with disabilities in the fields of computer science and accessibility.

Acknowledgments. This research was supported by the National Science Foundation (award #0746556), Siemens (Go PLM Academic Grant), and Visage Technologies AB (academic software license). W. Clarke, K. Gallagher, A. Krieger, J. Lamberton, A. Pagan, J. Penzellna, and M. Turtletaub assisted with data collection/annotation.

References

1. Mitchell, R., Young, T., Bachleda, B., Karchmer, M.: How Many People Use ASL in the United States? Sign Language Studies 6(3), 306–335 (2006)
2. Traxler, C.: The Stanford achievement test, ninth edition: national norming and performance standards for deaf and hard-of-hearing students. J. Deaf Studies & Deaf Education 5(4), 337–348 (2000)
3. Lu, P., Huenerfauth, M.: Collecting a Motion-Capture Corpus of American Sign Language for Data-Driven Generation Research. In: Workshop on Speech and Language Processing for Assistive Technologies (SLPAT), Human Language Technologies / North American Association for Computational Linguistics Conference (HLT-NAACL), pp. 89–97 (2010)
4. Huenerfauth, M., Lu, P.: Eliciting spatial reference for a motion-capture corpus of American Sign Language discourse. In: Workshop on the Representation and Processing of Signed Languages, LREC, pp. 121–124 (2010)
5. Neidle, C., Kegl, J., MacLaughlin, D., Bahan, B., Lee, R.: The syntax of ASL: functional categories and hierarchical structure. MIT Press, Cambridge (2000)
6. Bungeroth, J., Stein, D., Dreuw, P., Zahedi, M., Ney, H.: A German sign language corpus of the domain weather report. In: Vettori, C. (ed.) Workshop on the Representation and Processing of Sign Languages, LREC, pp. 2000–2003 (2006)
7. Efthimiou, E., Fotinea, S.E.: GSLC: Creation and Annotation of a Greek Sign Language Corpus for HCI. In: Stephanidis, C. (ed.) HCI 2007. LNCS, vol. 4554, pp. 657–666. Springer, Heidelberg (2007)
8. Brashear, H., Starner, T., Lukowicz, P., Junker, H.: Using multiple sensors for mobile sign language recognition. In: IEEE International Symposium on Wearable Computers, p. 45. IEEE Press, New York (2003)
9. Cox, S., Lincoln, M., Tryggvason, J., Nakisa, M., Wells, M., Tutt, M., Abbott, S.: Tessa, a system to aid communication with deaf people. In: ACM SIGACCESS Conference on Computers and Accessibility, pp. 205–212. ACM Press, New York (2002)
10. Lenseigne, B., Dalle, P.: A Tool for Sign Language Analysis through Signing Space Representation. In: Sign Language Linguistics and Application of Information Technology to Sign Languages Conference, Milan, Italy (2005)
11. Liddell, S.: Grammar Gesture and Meaning in American Sign Language. Cambridge University Press, Cambridge (2003)
12. Meier, R.: Person Deixis in American Sign Language. In: Fischer, S., Siple, P. (eds.) Theoretical Issues in Sign Language Research. Linguistics, vol. 1, pp. 175–190. University of Chicago Press, Chicago (1990)
13. Huenerfauth, M.: Improving spatial reference in American Sign Language animation through data collection from native ASL signers. In: Stephanidis, C. (ed.) UAHCI 2009. LNCS, vol. 5616, pp. 530–539. Springer, Heidelberg (2009)
14. Supalla, T.: Morphology of Verbs of Motion and Location. In: Caccamise, F., Hicks, D. (eds.) 2nd Nat'l Symposium on Sign Language Research and Teaching, pp. 27–45 (1978)
15. Nishio, R., Hong, S., Konig, S., Konrad, R., Langer, G., Hanke, T., Rathmann, C.: Elicitation methods in the DGS (German Sign Language) corpus project. In: Workshop on the Representation and Processing of Signed Languages, LREC, pp. 178–185 (2010)
16. Huenerfauth, M.: Participation of high school and undergraduate students who are deaf in research on American Sign Language animation. In: ACM SIGACCESS Accessibility and Computing, vol. (97), pp. 14–24. ACM Press, New York (2010)

Fuzzy-Logic Controller for Speaker-Independent Speech Recognition System in Computer Games

Moyen Mohammad Mustaquim

Department of Informatics and Media
Uppsala University
Uppsala, Sweden
moyen.mustaquim@im.uu.se

Abstract. Computer Games are now a part of modern culture. By using automatic speech recognition systems (ASRS), voice driven commands can be used to control the game, which can open up the possibility for people with disabilities and age related problems to be included in game communities and use the service offerered. Conventional speech recognition systems however, do not support emotions, attitudes, tones etc. This is a problem because such expressions can be vital for gaming, especially as the majority of the users to be considered for this operation are elderly adult. For instance, if a command is uttered in a low pitch the character in the game can act differently than when it is voiced in an elevated pitch. The traditional voice controller for such operation is usually based on logic or a probability supported algorithm that can serve the purpose but has certain limitations like long process time, cost and complexity in implementation. This paper proposes a fuzzy logic based controller to detect a user's emotion from their voice command for controlling the game in a different ways. This fuzzy logic controller can be programmed even on a microchip to drive certain voice driven actions as a built-in form of hardware. To use a fuzzy logic controller, instead of an approach based on traditional algorithms, gives more accuracy, control and efficiency when handling the user's voice input commands for manipulation.

Keywords: Universal design of computer games, fuzzy logic, fuzzy controller, speech controlled computer games design.

1 Introduction

1.1 Computer Games in Society Today

Games are the vehicle with which society is and will change the computer. But how will the games themselves are changed by society? We can expect two processes to affect games: one is the mass market and another is the flowering of heterogeneity. In some ways, these processes work against each other [1].

As participation in gaming increases around the world and across a larger part of the population, it is obvious that games are not only an entertainment media for children or young people anymore but games can be used to serve several purposes

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 91–100, 2011.

for all kinds of people in the society. Today's technologically inclined people are going to be older tomorrow, and they would not like to be excluded from future services for instance games, just because of design issues associated with the game. Also, anyone can suffer temporary or permanent disabilities regardless of age and it would be unfortunate to exclude them from using games. Moreover, both the human death rates and birth rates are generally decreasing, which results in an age distribution with a large number of older people (disabled or not) in the society in the near future. So the game industries should be concerned with such issues, so that they can include these groups of people and offer them properly designed services.

1.2 Using Human Modalities in Game Design

The advent of the information society, along with the emergence of novel technological paradigms such as mobile or ubiquitous computing, ambient intelligence, and the "disappearing computer", raises new and stimulating scientific challenges in research on multimodal human computer interaction (HCI) [2]. New input/output modalities and forms of multimodality are hence needed to provide easy and effective access for all users, including the very young and the elderly, as well as people with various types of disabilities; in all contexts of use, especially in mobile interaction with embedded systems, wearable computers, and augmented everyday objects integrating aspects of intelligent interactive behavior [2]. It is thus a wise approach to use different human modalities for controlling games.

2 Background

2.1 Game Controller

A game controller is a device used for controlling the playable character or objects or otherwise providing input in a computer game. The controller is usually connected to a game console or a computer by means of a wire, chord and nowadays also by means of wireless communication. Typical game controllers are keyboards, mice, joysticks, game pads etc. There are also special purpose devices such as steering wheels or foot pedals. The principal function of a game controller is to control the movement or actions of a playable body-object or otherwise influence the events in a video- or computer game. There is health concerns related to game controllers such as risk for injuries like repetitive strain injuries or carpal tunnel syndrome hence they need to be designed ergonomically to give best possible relaxation of hands and mind. However, certain groups of people have limited accessibility or no accessibility at all to such game controllers because of their physical limitations. Alternative ideas for game control are therefore important to consider for such groups of people.

2.2 Speech Technology

Speech technology seems to provide new opportunities to improve the accessibility of electronic services and software applications including games, by offering compensation for the limitations of specific user groups. These limitations can be quite diverse and originate from specific sensory, physical or cognitive disabilities—such as difficulties to see icons, to control a mouse or a keyboard [3]. Such limitations have both functional and emotional aspects that should be addressed in the design of user interfaces [4]. Speech technology can be an 'enabler' for understanding both the content and 'tone' in user expressions, and for producing the right information with the right tone [3].

2.3 Automatic Speech Recognition System

Automatic speech recognition (ASR) also known as speaker independent speech recognition is the automatic conversion of human speech to a sequence of words. The aim of ASR is to recognize automatically what has been said. Nowadays, ASR systems are generally based on the hidden Markov models (HMM) [12] for modeling the acoustics of speech and use either statistic language models (n-grams) or rule-based grammars to model the language component. Furthermore, there are many techniques that normalize the acoustic features of the signal and adjust the acoustic models to a specific speaker or different recording conditions[3].

If a speech recognizer is part of an application, usually it is used as an input technology or technique [5]. Various other performance measures can reduce the word error rate, depending on the application. For instance, van Leeuwen et al.[6] mentions the following issues that are important to consider when assessing the performance of an ASR application: the type and quality of feedback, error correction strategy, dealing with out-of-domain words, speed and response time, the user's situational awareness in the dialog structure, dialog/task success rate, and subjective impression of the overall performance of the system [7].

Nowadays, speech recognition technology is used in many commercial applications. In these applications, shortcomings of the technology are reduced by tailoring the interaction to the individual user and context, for instance, by making the system speaker-dependent (e.g., a dictation system), by limiting the vocabulary (e.g., voice dialing) or by using a simple and predictable syntax (e.g., command and control, data-entry)[3]. In some applications, recognition accuracy does not need to be perfect, e.g., for spoken document retrieval a word recognition accuracy of 70% produces similar retrieval results as manually generated speech transcripts [8].

2.4 Requirements of a Voice Controlled System for Games

While the traditional approach of game control is using input devices, people with certain limitation of doing so would like to use other alternatives for control. People with motor system problems, limitations of moving the upper limbs, problems with hand muscles, hand joints, muscular stress, and many other neurological disorders, can be highly benefited by using voice driven command systems in gaming. Age related issues can make the hand of a person weak which is quite common and hence use of speech is reasonable for such users for game control. Also the use of speech does not require any special hardware. Ordinary computers of today are able to carry out speech transforming operations from the user to the process handler of the game, hence there is no need for any extra hardware and disabled people do not need to wear or use any special device to use the speech system for controlling games.

2.5 Fuzzy Logic

Fuzzy logic is a problem-solving control system methodology that lends itself to implementation in systems ranging from simple, small, embedded micro-controllers to large, networked, multi-channel PCs or workstation-based data acquisition and control systems. They can be implemented in hardware, software, or a combination of both [13]. Fuzzy logic provides a simple way to arrive at a definite conclusion based upon vague, ambiguous, imprecise, noisy, or missing input information and fuzzy logic's approach to control problems mimics how a person would make decisions, only much faster. Fuzzy logic was first conceived as a better method for sorting and handling data but has later proven to be an excellent choice for many control system applications since it mimics the logics of human control. It can be built into anything from small, hand-held products to large computerized process control systems. It uses an imprecise but very descriptive language to deal with input data more like a human operator. It is very robust and forgiving of non-perfect operator performance and data input, and often works when first implemented with only little or no tuning.

2.6 Fuzzy Logic Controllers

Fuzzy logic control may be viewed as a branch of intelligent control. It can be seen, as an emulator of the human decision making process in that it is approximate rather than exact. The conventional control systems work either in a linear or a non-linear manner. Linear systems are simple and well explored but not too flexible. The non-linear systems on the other hand can be powerful but are often sensitive to modeling errors. A fuzzy controller is effective in situations where the control process is too complex to analyze by conventional quantitative techniques and also when the available sources of information are interpreted qualitatively, inexactly or uncertainly. A fuzzy logic controller has several advantages over a traditional controller. It is more flexible, and easy to understand. A fuzzy controller uses human linguistic terms for control. To design a fuzzy controller for parallel or distributed control, multiple fuzzy rules, or complex non-linear systems can be used. However, there are more parameters to tune in the fuzzy controller and the stability of the fuzzy controller is sometimes difficult to analyze mathematically. Regardless of its complexity a fuzzy logic controller can be used where it is difficult or impossible to model mathematically and where traditional strategies are too complex or non-linear to be controlled. Figure 1 shows the basic structure of a fuzzy logic controller.

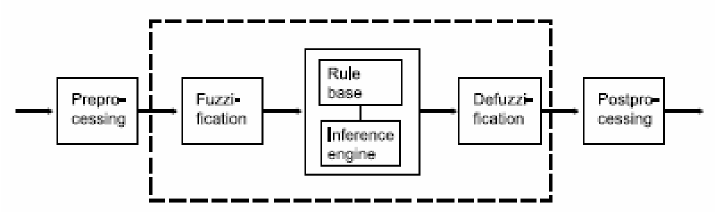


Fig. 1. Structure of a fuzzy controller

2.7 Traditional Approach of Detecting Emotion from Voice

The traditional control system to detect emotions in voice control systems for games consists of a standard automatic speech recognition system where the vocabulary array is built by fetching the commands from the most commonly used game platform. The user input is processed through the soundcard which converts the analog data to digital format and then checks with the user word model from the vocabulary register. Once the word uttered by the user matches with the word from the vocabulary list, it then sends the binary signal corresponding to that word to the process handler of the game. Users see that the action has been taken according to their given voice command, through the GUI of the game. Figure 2 shows the block diagram of such conventional system.

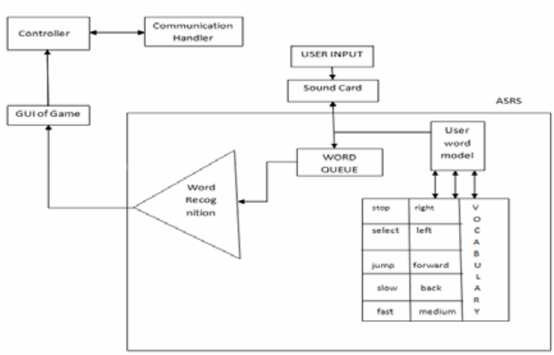


Fig. 2. Common speech recognition system for game control

3 Proposed System- The Fuzzy Controller

The proposed system replaces the traditional system's control part as shown in Figure 2, by a fuzzy-logic controller. There exists linguistic terms which are the content of the fuzzy set describing the command actions. The rule input commands are processed in to output. The fuzzy controller is described in the following subsection. The proposed fuzzy controller's block diagram is shown in Figure 3. The fuzzification process involves handling voice command variables (d1,d2….dn), output variables (ω and v) and transforming numerical variables into fuzzy sets. Two output variables ω and v are used respectively for determining the angular velocity and the linear velocity of an object. Hence state variables (d1,d2….dn), are decided with the error and change of error by the user's voice input.

The proposed system replaces the traditional system's control part as shown in Figure 2, by a fuzzy-logic controller. There exists linguistic terms which are the content of the fuzzy set describing the command actions. The rule input commands are processed in to output. The fuzzy controller is described in the following subsection. The proposed fuzzy controller's block diagram is shown in Figure 3. The fuzzification process involves handling voice command variables (d1,d2….dn), output variables (ω and v) and transforming numerical variables into fuzzy sets. Two output variables ω and v are used respectively for determining the angular velocity and the linear velocity of an object. Hence state variables (d1,d2….dn), are decided with the error and change of error by the user's voice input.

Here, error is determined by the error from the processed output minus the error from the last processed output. The controlled variables ω and v are decided, as said before to determine the angular and linear velocity of an object which is important to consider for the different movement based on the user-given voice command. The fuzzification process converts the crisp input values to fuzzy set values.

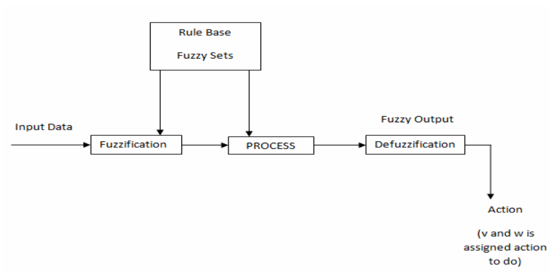


Fig. 3. Proposed fuzzy logic controller system

An inference method was chosen to design this fuzzy controller which is the Mamdani inference method[11]. Also the fuzzy singleton [14] fuzzification method is selected as it measures the state variables without uncertainty. The inference process defines the connective implication and rule combination operations. The controller uses minimum connectives and a singleton sum product inference mechanism. This choice relies on two reasons. First, the product preserves the shape of the output fuzzy set and second, with the sum, the result is influenced by difficult rules, reaching the same conclusion.

Output variables (ω and v) are also modeled as discrete fuzzy set as it increases speed in inference and defuzzification process. Figure 4 shows the shape of fuzzy input sets where d1, d2, d3……dn are considered to be the voice sensor data received from the user. The angle θ is used as a guidance operator.

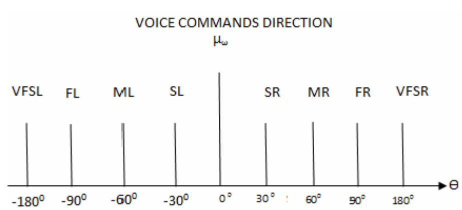


Fig. 4. Shape of fuzzy input sets deciding direction of object

That is, the way d is sensed from the user input, triggers the value of   and puts a decision value on the output variable describing at which angle the object should move. The value of   is discritized and normalized within the range of [-180,180].

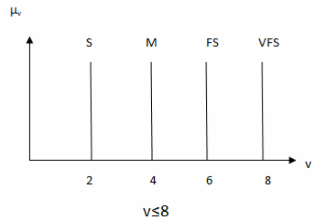


Fig. 5. Shape of fuzzy output sets for linear velocity

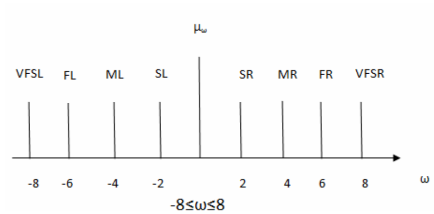


Fig. 6. Shape of fuzzy output sets for angular velocity

Figure 5 and figure 6 shows the shape of fuzzy output sets where limit of v and ω are discretized and normalized within the range of [0, 8] for v and [-8, 8] for ω.

The numerical values obtained from the two output sets are sent for a combination of two sets, and then sent to the communication handler of the game to execute the desired action. The content of the fuzzy set are linguistic terms which is the rule based fuzzy set used to measure the pitch of the voiced command to act differently from the character of interest.

The center of gravity (COG) [13] method is used for defuzzification strategy, as in this problem it results less processing time with no degradation of defuzzified value. Also COG method simplifies the computation of inference mechanism.

4 Proposed System - The Fuzzy Controller

4.1 Linguistic Terms

Sample linguistic terms for the rule base fuzzy set for few game commands are shown below. The knowledge of these commands is picked up from most widely used commands of controlling game using voice.

R=Right
SR= Right turn slow
MR=Right turn medium
FR= Right turn fast
VFSR= Right turn very fast
L=Left
SCT= Select
NS=Not select

ST= Stop
U=Up
MV=Move
S= Slow
M= Medium
FS= Fast
VFS= Very fast

4.2 Sum-Product Interference

Degree of each rule is determined by the following rule:

βl=m in(μAi(di)…… μAi (dk)). Ai are the input fuzzy sets that is the defined linguistic terms and dj (j=1….k) are the input variables.

Output fuzzy set Bl´ is constructed using following rule:

For each rule l, we use product operation μ βl´(v)= βl μBl(v). Here Bl are the output fuzzy sets and v is the output variable.

Combination of output fuzzy sets to single fuzzy set B´:

Sum operation: μB´l(v)= μB´l(v)+….+ μB´N(v)

4.3 Construction of Fuzzy Rules

The following example fuzzy rules are the representation of the gathered knowledge of user's voice command for game control.

Turn Left:
If θ is SCT and d0 is L and d1 is SCT Then v is ST and ω is L

Stop:
If θ is MV and d0 is SCT and d1 is ST Then v is ST and ω is ST

Right Turn Fast:
If θ is SL and d0 is R and d1 is FR and d2 is MV Then v is ST and ω is FR

Select and Move Forward
If θ is NS and d1 is SCT and d2 is FW and d3 is MV Then v is MV and ω is ST

5 Discussion

Although vocal interaction has technological limitations, these might be overcome when used as a component of a multimodal user interface. Such interfaces seem to have higher levels of user preference, among other things, because people experience a greater degree of flexibility [9]. The performance of the proposed fuzzy controller depends on how accurately it is implemented and fine tuned. The proposed fuzzy controller can be made more sophisticated while issues like word error rate, single word error rate and command success rate will be verified and fine tuned while combined with any algorithm such as HMM [10]. Use of Viterbi algorithm [12] is reasonable as long as it does not increase the cost of the overall controller. As the system will mostly be dealing with old or disabled people, decoding of the speech might be difficult sometimes, especially when the system is presented with a new utterance and it must then compute the most likely source word. Use of the Viterbi algorithm is thus prudent to find the best path which can be an added advantage in this system.

However in the proposed fuzzy controller, some issues have been considered in terms of performance. The measurement of uncertainty is the first issue where the classical rules to find the max or min value has been use. Second, the importance measurement was considered, that is the uncertainty measurement rule might be altered by the value of importance of the members in fuzzy set, which may not just give the priority to the maximum or minimum value in this case. The quality measurement rule was also considered. The quality of a rule is measured by multiplying the value obtained by a rule with the quality value assigned to a rule. The quality value assigned to a rule can be 0 or 1 where 0 describes a rule to be lower quality and 1 describes the rule to be of a good quality.

6 Conclusion

The use of speech to control games is only one step towards inclusiveness and thereby universal design and accessibility in games. The result from this paper, to use a fuzzy controller for designing voice controlled game has multiple possibilities in game design and PC industry. While any existing algorithm can be altered and new algorithm can be created for detecting user's emotions for executing one command in different way in game playing; a fuzzy-logic based controller programmed in chip and integrated with sound card, substantiate the lasting universal design of certain standard of voice command for playing game. This enables and opens enormous opportunity both for hardware manufacturers and game industries. Voice support for game, based on this paper's idea can show the way towards finding an industry standard of universal design in computer gaming in an optimal way.

References

1. Crawford, C.: The Art of Computer Game Design (1982)
2. Carbonell, N.: Multimodality: a step towards universal access. Universal Access in the Information Society 2(2), 89–90 (2003)
3. Neerincx, M., Cremers, A., Kessens, J., van Leeuwen, D., Truong, K.: Attuning speech-enabled interfaces to user and context for inclusive design: technology, methodology and practice. Universal Access in the Information Society 8(2), 109–122 (2009)
4. Picard, R.W.: Affective Computing. MIT Press, Cambridge (1997)
5. Hinckley, K.: Input technologies and techniques. Lawrence Erlbaum, Mahwah (2003)
6. Van Leeuwen, D.A., Martin, A.F., Przybocki, M.A., Bouten, J.S.: NIST and TNO-NFI evaluations of automatic speaker recognition. Computer Speech Language 20, 128–158 (2006)
7. Van Wijngaarden, S.J., Smeele, P.M.T., Steeneken, H.J.M.: A new method for testing communication efficiency and user acceptability of speech communication channels. In: Proceedings of Eurospeech 2001, Aalborg, pp. 1675–1678 (September 2001)

8. Ordelman, R.J.F., de Jong, F., Huijbregts, M.A.H., van Leeuwen, D.A.: Robust audio indexing for Dutch spoken-word collections. In: Proceedings of the XVIth International Conference of the Association for History and Computing (AHC 2005), Amsterdam, The Netherlands, September 14-17 (2005)
9. Oviatt, S.: User-centered modeling and evaluation of multimodal interfaces. IEEE 91(9), 1457–1468 (2003)
10. Eddy, S.R.: Hidden Markov models. Current Opinion in Structural Biology 6(3), 361–365 (1996)
11. Ying, H.: The simplest fuzzy controllers using different inference methods are different nonlinear proportional-integral controllers with variable gains. Automatica 29(6), 1579–1589 (1993)
12. http://www.iem.uni-due.de/~vinck/data-com/The%20viterbi%20algorithm.ppt
13. http://farfromneutral.com/kaizen/center-of-gravity-method/comment-page-1/
14. http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol2/jp6/article2.html

Building of Turn-Taking Avatars that Express Utterance Attitudes

A Social Scientific Approach to Behavioral Design of Conversational Agents

Masahide Yuasa and Naoki Mukawa

School of Information Environment, Tokyo Denki University,
2-1200 Muzai Gakuendai, Inzai,
Chiba 270-1382, Japan
{yuasa,mukawa}@sie.dendai.ac.jp

Abstract. In everyday communication, humans comprehend the attitudes of others conveyed via nonverbal behavior, such as facial expression, body posture and gaze behavior. In this paper, we describe a model for comprehending participants' desire to start to speak or to listen based on nonverbal behavior during conversation. We use a social scientific approach that is based on both an analysis of a video observation and an experiment using avatars. We explain the building of the model. We discuss detecting participant's attitudes using computer vision and the expression of their attitudes using their avatars' facial expressions and body postures.

Keywords: animated agent, avatar, turn-taking, nonverbal behavior, conversation.

1 Introduction

In everyday life, we communicate with our family, friends, co-workers, and others. We enjoy talking and eating together and sometimes have confrontational discussions. Such shared emotions are important for human beings as they contribute to the development of better relationships and create a sense of unity. In order to share emotions, we use both verbal and nonverbal behavior such as gaze direction, head orientation, facial expressions, and body posture. Nonverbal behavior plays an important role in comprehending subtle emotion. By using nonverbal communication, we express attitudes and can comprehend the degrees of others' feelings.

In our previous research about turn-taking in communication, we proposed a conversational utterance attitude model [14]. The model has two demands: "I want to start to speak" and "I want someone to start to speak." We use abstract animated agents that mimic human turn-taking in conversations, to confirm the validity of our model. However, the agents' shape was a simple sphere; which was simply not realistic enough. In order to improve upon our previous model, we needed a system that has animated avatars, agents, or humanoid robots.

Therefore, in this paper, we describe turn-taking avatars that express utterance attitudes. The avatar system detects participants' utterance attitudes using a camera, and

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 101–107, 2011.

the attitudes are then expressed by their avatars. Participants can use several utterance attitudes during conversation and can communicate smoothly without revealing their actual faces.

We used a social scientific approach in developing the model. We recorded an actual three-party conversation and observed the video. On the basis of the observation, we developed an utterance attitude model for demand such as “I want to speak”/ “I want you to speak” and built animated avatars using the model. Here, we discuss the value of the avatar system and our proposed model for conversational robots or agents.

2 Previous Research

Avatars are widely used in order to make communication more natural and fun. An avatar is the embodiment of a person. Avatars are used in online games, online communities, and 3D virtual worlds. Users can chat and interact virtually with other users. For example, BodyChat [12] uses embodied avatars to mimic face-to-face communication. There are also several virtual worlds; such as Second Life[8], Worlds.com[13] and There[9]. The avatars in these virtual worlds are capable of realistic nonverbal behavior (body posture, eye movements, head orientation, etc.) that closely resemble that in human-to-human interactions. Users can select the type of interaction for their avatar using a keyboard or a mouse. In addition, facial expressions created via facial motion-capture data are used in order to enrich communication between avatars. Yun et al. investigated the effectiveness of a local 3D facial avatar for a global audience [15].

However, the behavior of avatars is limited using these tools. Users can only choose from predefined basic behaviors. Designing ways for avatars to express their utterance attitudes using eye-gaze behavior or facial expressions during turn-taking is very important to make communication more natural and fun for users.

The psychological literature has analyzed the relationship between turn-taking and eye-gaze behavior [1, 3]. Sacks states that a listener being watched by a speaker tends to start speaking [7]. Thus, there are several existing agents and robots with a turn-taking function that detect a speaker’s gaze using eye-tracking equipment [5]. Peters et al. developed animated agents in a virtual world with a mathematical model of gaze behavior. In their research, when one agent looks at another hearer, the gaze behavior of the first attracts the attention of another, who then takes the next turn to interact. Traum et al. [10] developed a conversational agent that takes turns in a multiparty conversation. The agent gives a simple rule-based response to the user; that is, when a user looks at an animated agent, the agent responds to the user. Poggi et al. talk about specific communication signals called “Mind Markers,” including facial expressions and gazes that indicate a speaker’s beliefs, goals, and emotions [6]. However, their research only used a limited number of basic behaviors and their implied meanings.

Duncan et al. described several turn-taking categories (e.g., turn-maintaining, turn-yielding) and the complexity of turn-taking itself. We need to take into account more complicated rules of turn-taking.

In our research, we developed a conversational avatar system with effective and efficient turn-taking based on the utterance attitude model. Using the model, we developed conversational avatars that can communicate with humans in a lively and emotional manner. In our previous research, we proposed an utterance attitude model that presents a participant's utterance attitudes that "he/she wants to start to speak" or "he/she wants someone to start to speak" based on an analysis of the observed conversations. On the basis of this model, we built an animated avatar system that can express participants' attitudes detected by a camera. The participants can converse on the basis of the attitudes of each other's animated characters; they do not need to reveal their real faces and they can express their own attitudes. This system allows smooth communication and promotes good relationships among participants.

3 A Social Scientific Approach to Behavioral Design of Conversational Agents

In this section, we describe our social scientific approach to building our conversational avatars. Our approach consists of four processes that involve analyzing actual human behavior, developing a model, building avatars as a prototype system, and evaluating the system.

3.1 Analysis of Actual Chatting

We observed and analyzed a video-recorded 20-min conversational scene of three female university students using ethnographic conversation analysis. We observed and carefully transcribed every action of these participants, including words (what the participants were saying) and nonverbal actions, such as gaze, head orientations, and facial expressions. After the analysis, we proposed an utterance attitude expression model.

Fig. 1. Conversation scene with three female university students

3.2 Utterance Attitude Model [14]

Figure 2 shows our proposed utterance attitude model [14]. In this model, utterance attitudes can be categorized into nine classes on a two-dimensional plane: the horizontal axis represents the expressions of a person who "wants to speak/not to speak," and the vertical axis represents a person who "wants someone to speak/not to speak." The plane also shows that expressions can be classified into two types: (1) subtle, implicit, expressive behaviors that are expected to be noticed by others and (2) direct and explicit behaviors that intentionally control the utterance behaviors of others. The inner ring of the model shows implicit attitudes displayed by participants, for example, (1) I want to speak. The outer ring indicates explicit attitudes that can control the other participants, for example, the (7) I want him/her to speak.

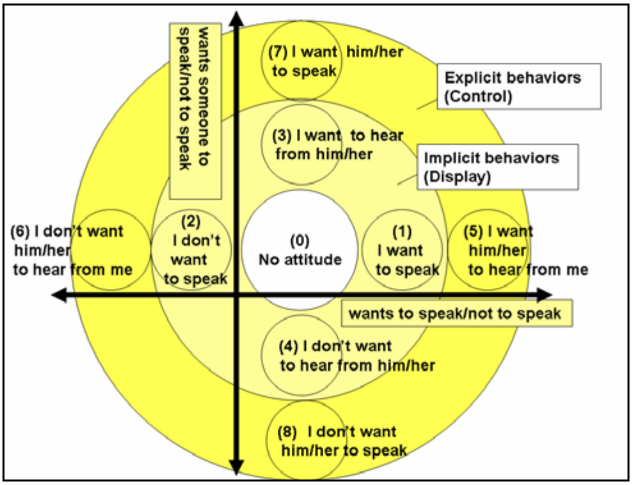


Fig. 2. Utterance Attitude Model

3.3 Building Turn-Taking Avatars

Utterance Attitudes for Avatars. We developed a system that enables three participants to converse using avatars. We implemented four expressions of utterance attitudes in our avatar system. Figure 3 shows avatar expressions that represents the utterance attitudes of participants.

As shown in Fig. 3 (a), the avatar's body expanding in height represents a positive utterance attitude. The avatar's contracting body shown in Fig. 3 (b) represents a

Fig. 3. Avatar's expressions and the representation of a participant's utterance attitudes

negative utterance attitude. The effects of expanding/contracting behaviors are well known in the literature about animation movies. In should be noted that the movements of the body reveal the avatar's implicit attitudes that "he/she wants to start to speak" or "he/she does not want to start to speak."

In addition, the avatar has hands for purposively controlling partners' utterances as shown in Fig.3, (c) promoting and (d) blocking. These movements represent explicit attitudes in order to control the other's turn. The avatar's promoting hand means it "wants you to speak." and the avatar's blocking hand means "please stop speaking."

The system can easily detect these four participants' attitudes using computer vision technology and can express the avatars' attitudes.

Fig. 4. User Interface

User Interface. This system has six cameras to detect a participant's behaviors, which are transmitted to the other participants. An audio link was established and the participants wore headsets with a microphone to converse.

One participant can see the behaviors of the other two participants on the display. The system was arranged so that when a participant looks at the avatar to the left, the participant represented by that avatar can see the front face of the avatar. The other participant represented by the avatar on the right can see the profile of that avatar's face (Fig. 4).

Detecting a Participant's Face and Hands. Our system detects head orientation (right/left), the height of the head (high/middle/low), and hand gestures (promoting/blocking) using cameras. We use six cameras set around two displays in front of a participant. The system determines whether a participant gazes at the right or the left display and estimates the height of the body by detecting the height of the participant's head. The avatar's body expanding or contracting in height is based on the participant's head movement in term of height. In addition, the system can detect hand gestures using a colored marker, which is attached to the participant's finger.

3.4 Preliminary Test of the Avatar

We tested the effectiveness of our avatar system using three participants. We recorded the participants' conversation using the avatars and observed the recorded video. We found that participants used expressions of utterance attitudes efficiently and an avatar's expressions were recognized by the others as utterance attitudes; therefore, our expressions of turn-taking were useful in avatar communication. For example, when a participant had an opinion and wanted a turn to speak, his/her head or body leaned forward, and his/her avatar's head and body were moving up. The participants recognized the movement as meaning "he/she wants to start to speak." When a participant did not want to start to speak, he/she stretched out his/her hand to promote conversation and his/her avatar did the same. This movement was interpreted as "he/she wants me to start to speak." Thus, the utterance attitudes we proposed were used effectively to communicate in the avatar system. In an interview with the participants, they stated that they felt there was less collision (when more than one person starts to speak at once) in turn-taking than during the case of voice chatting. One participant commented that he could prepare to start to speak and it was easy to speak at the turn-taking point of the conversation.

4 Discussion

We used only limited expressions (four attitudes) for avatars in the preliminary test. Even though participants could use only four expressions of utterance attitudes, they could still converse. We should use other expressions of attitudes, including subtle nonverbal behaviors during conversation. However, it is difficult to extract these behaviors because nonverbal behaviors can be very subtle. Even though we selected only four attitudes based on psychological findings, it was very difficult to select appropriate expressions for the other attitudes. In the preliminary test, our expressions of turn-taking were useful in avatar communication. Therefore, four expressions are enough, because it is the first step using both implicit and explicit attitudes in an utterance attitude model. Additional experiments are required to design more understandable expressions based on research about nonverbal behavior.

We did not take into account combinations of attitudes; such as "expansion" (the wants-to-speak attitude) and "blocking" (the does-not-want-a-partner-to-speak attitude). Further experiments on the combinatorial effects of attitude are required.

On the basis of these findings and future research, we will improve the avatar system to make communication and turn-taking smoother.

5 Conclusions

We proposed a new avatar system that detects participants' utterance attitudes and generates their avatars' utterance attitudes. We selected appropriate expressions of utterance attitudes for the avatars on the basis of psychological findings. In our test, participants could take turns in conversing using the avatar system and the expressions were used effectively.

References

1. Argyle, M., Cook, M.: Gaze and Mutual Gaze. Cambridge University Press, Cambridge (1976)
2. Duncan Jr., S.: Some Signals and Rules for Taking Speaking Turns in Conversations. Journal of Personality and Social Psychology 23, 283–292 (1972)
3. Kendon, A.: Some Functions of Gaze Direction in Social Interaction. Acta Psychologica 32, 1–25 (1967)

4. Ma, C., Osherenko, A., Prendinger, H., Ishiizuka, M.: Chat system based on emotion estimation from text and embodied conversational messengers. In: Proc. of the 2005 International Conference on Active Media Technology (ATM 2005), pp. 546–548 (2005)
5. Matsusaka, Y., Fujie, S., Kobayashi, T.: Modeling of Conversational Strategy for the Robot Participating in the Group Conversation. In: Proc. of Eurospeech 2001, pp. 2173–2176 (2001)
6. Poggi, I., Pelachaud, C., Caldognetto, E.M.: Gestural Mind Markers in ECAs. In: Proc. AAMAS 2003, pp. 1098–1099 (2003)
7. Sacks, H., Schegloff, E., Jefferson, G.A.: A Simplest Systematics for the Organization of Turn-taking for Conversation. Language 50(4), 696–735 (1974)
8. Second Life, `http://www.secondlife.com/`
9. There.com., `http://www.there.com/`
10. Traum, D., Rickel, J.: Embodied agents for multiparty dialogue in immersive virtual worlds. In: Proc. of AAMAS 2002, pp. 766–773 (2002)
11. TVML, `http://www.nhk.or.jp/strl/tvml/`
12. Vilhjalmsson, H., Cassell, J.: Bodychat: autonomous communicative behaviors in avatars. In: Proc. of the Second International Conference on Autonomous Agents, pp. 269–276 (1998)
13. World.com., `http://www.world.com/`
14. Yuasa, M., Mukawa, N., Kimura, K., Tokunaga, H., Terai, H.: An Utterance Attitude Model in Human-Agent Communication: From Good Turn-taking to Better Human-Agent Understanding. In: CHI Extended Abstracts 2010, pp. 3919–3924 (2010)
15. Yun, C., Deng, Z., Hiscock, M.: Can local avatars satisfy a global audience? A case study of high-fidelity 3D facial avatar animation in subject identification and emotion perception by US and international groups. Computers in Entertainment 7(2) (2009)

Part II

Interacting with Documents and Images

Visual Design Elements of Photos on Taiwanese Female Clothing On-Line Sales

Ro-Han Chang, Tsai-Rong Yu, and Yao-Cheng Tsai

Department of Commercial Design, Chung Yuan Christian University,
No.200, Zhongbei Rd., Zhongli City, Taoyuan County 320, Taiwan
rohan@cycu.edu.tw, phices0316@hotmail.com,
tyc_production@hotmail.com.tw

Abstract. The number of on-line shoppers has gradually increased over time. Due to the change of the market, a virtual store has to think highly about the way its brand is represented on its website. In order to be competitive in the market, building a brand image is vital to future success. Recorded documents state that besides the quality of the products, the reputation of the store, and the price of the products, the way in which a store represents itself is of the upmost importance. From an academic perspective, the style and the presentation of the photos from the on-line store is worthy of discussion and systematization. This research will be separated into three parts: (a.)Use a photography perspective to study the difference between traditional commercial photography and photos from on-line stores. (b.)Use a morphology perspective and organize the visual design elements (lighting, framing, composition, angle of shooting, shooting scene and the presentation of products) of the photos of products from on-line clothing store. (c.)Establish how visual design is composed through the use of multiple photography styles with varying emphasis on focal points.

Keywords: visual design element, on-line sale, photo, female.

1 Introduction

Due to the change of the market, a virtual store has to think highly about the way its brand is represented on its website. Recorded documents state that besides the quality of the products, the reputation of the store, and the price of the products, the way in which a store represents itself is of the upmost importance [1].

Due to the concept of an online store, browsers do not have physical contact with the product. The goal of every on-line seller in this particular industry is to present unique and stylish photographs of clothing items. Achieving this will increase trustworthiness from the female consumer toward shopping websites in regard to the professionalism of the photos. Furthermore, unique presentation can make a tactile impression on the browser [2]. Hui-Chen Chung stated that females pay more attention on the impression of the clothing website than males when it comes to on-line shopping and decision

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 111–120, 2011.

making [3]. Ya-Wen Yu also indicated that the focal point of a photograph can influence and affect the feelings of consumers [4]. Due to the shopping behavior of females, they value the elements of web design more than males. Therefore, it is critical to create a shopping environment within the on-line market especially for females. The presentation of photos from on-line stores might affect the feelings of, and decisions made by a shopper. From an academic perspective, the style and the presentation of the photos from the on-line store is worthy of discussion and systematization. For the purpose of this research, the following will occur: (1) Study the difference between traditional commercial photography and photos from on-line stores. (2) Establish how visual design is composed through the use of multiple photography styles with varying emphasis on focal points. (3) Discuss the composition of visual design elements by analyzing different photos. On conclusion of this research, it is anticipated a discussion can take place. The focus of this discussion will be whether the combination of different visual design elements can create a selling environment and deliver style and brand image to the market.

2 Literature Review

2.1 Presentation of the Photos from On-Line Clothing Stores

At present, the main presentation on a store's website is as follow: (1) Floor photography: It means photograph an object which lies on the floor. Notomi Yasukuni indicated that it was used by the web seller in the early stage [5]. (2) Hang clothes on a hanger. This method is usually used by the seller of a small company or second hand seller in the event of a clearance sale when there is zero cost of human resource [6]. (3) Mannequin: Misawa Atsushi once addressed that using a mannequin could be more effective and three-dimensional than (1) and (2) above [7]. (4) Self-portraits by seller. Recently, sellers of much smaller companies have taken on the role of model and photographer in presentation due to budget considerations. They usually dress themselves in the company's products and take pictures from the reflection in mirror. The angle of the photograph is well positioned so their facial features are not exposed in order to create a mystical feeling. (5) Model. Pictures can either be shot indoors or outdoors. When photography shoots take place indoor; lighting, composition, and the posture of model need to be considered. When photography shoots take place outdoor; climate, framing, and scenery need to work together with the composition and the posture of model [8].

2.2 The Difference between Traditional Frame Photography and Photography Styles from On-Line Stores

The following tables summarize information from recorded documents and statements from experts. The following will define the difference between traditional commercial photography and photos from on-line stores.

The Difference between Traditional Commercial Photography and Photos from On-line Store for Framing

Table 1 shows the difference between traditional commercial photography and photos from on-line store for framing.

Table 1. The difference between traditional commercial photography and photos from on-line store

Method of Framing	Traditional commercial photography	Photos from on-line store
Full Shot	It's better on performing an associative perception [9]. Not gloomy if enough space is provided [10].	Taiwanese Famous photographer Hei-Mian indicated that it can work with model to bring out the associative perception [11].
Medium Full Shot	The range is between a model's face to knees, placing emphasis on the upper body [12]. Chiung-Chieh Chin &Well Planning stated that it's better to crop the parts of a photo below the knees to avoid being unnatural [8].	Medium full shot can be used to present the clothing on the upper body [13].
Medium Shot	David Image Studio said used to take photos above waist [14]. It is best to show the model's expression and have it harmonized with the background [15].	Medium shot can be used to present dresses or long sleeve tops. It can show the expression of models and their posture [15].
Bust Shot	Bust shot is the most common framing used by traditional commercial photographer. It completely reveals the expression of a model [9].	Bust shot cannot present the full feature of clothing products, so it is less applied in the clothing industry. [13].
Close-up Shot	Close-up shots emphasize the emotions of the model.	Close-up shots do not allow the features on clothing items to stand out [11].

The Difference between Traditional Commercial Photography and Photos from On-line Store for Composition

Composition is the placement or arrangement of visual elements or ingredients in a work of art in photography. Framing, shooting distance, and shooting angle are all part of the process (see the Table 2).

Table 2. The difference between traditional commercial photography and photos from on-line store for composition

Composition	Traditional commercial photography	Photos from on-line store
Dot Composition	Shrink character into a dot and centralize it in a clean picture [16].	Using dot composition can make the character stand out [11].
S Composition	S composition: a curve style which works perfectly for the presentation of the female body [9].	S composition presents maturity and sexiness from the photos on the internet [16].
Three aspects Composition	It is also called the golden cropping composition. It shows a sense of proportion [9].	Three aspects composition can fully present clothing products and combine the expression and posture of the model [17].
Two aspects Composition	It separates the image into two parts. It is usually used for outdoor shots and focuses on expression [12].	Two aspects composition will focus on the expression of the model, but combined with the framing can emphasize the features on the clothes [9].
Circle Composition	This composition is used to focus on a main feature and grab attention [12].	This composition will emphasize the facial expression rather than features on clothing items [14].
Triangle Composition	It gives a sense of stability above the knees [16].	Model can use appropriate posture in order to present the triangle composition [10].

The Difference between Traditional Commercial Photography and Photos from On-line Store for the Angle of Shooting

Shooting angle is the placement of the angle from a camera. Different angles affect the photo result (see the Table 3).

Table 3. The difference between traditional commercial photography and photos from on-line store for the angle of shooting

Angle of shooting	Traditional commercial photography	Photos from on-line store
Eye Angle	Misawa Atsushi indicated that it is important to add a gaze from the model to avoid being unnatural [7].	The most traditional way of shooting but communicating as though a story is being told [9].
Low Angle	Low angle can make figures slimmer and taller [17].	It can make the model look slimmer and taller [11].
High Angle	High angle will make the head big and body smaller [4]. It can produce cute expressions [18].	It can work on cute clothing and youth clothing in order to show a lively feeling [9].

The Difference between Traditional Commercial Photography and Photos from On-line Store for Location

When taking photographs, the location and model must correspond with each other in order to create an environment that appeals to consumers. Table 4 shows the difference between indoor and outdoor.

Table 4. The difference between traditional commercial photography and photos from on-line store for location

Scene	Traditional commercial photography	Photos from on-line store
indoor	Roger Hicks indicated that round 10am in the morning is the best time to get a clear tactile impression or detailed subject for the photo [19]. In summer, the best time to take a photo is sunrise or sunset. In winter, photographs facing the southern direction of a room can be taken any time of day [7].	The advantage to shoot indoors is convenience. A model has to provide a lively posture due to the simple background [16].The simplest way to shoot indoor is to use white background. The result will be clean and comfortable [11].
outdoor	The best time to shoot outdoor is between 7-10am or 3-6pm [11]. During this time the sunshine is gentle but sufficient and will not leave a shadow on the face [16].	David Image Studio said the goal of shooting the photo outdoor is to make the consumer feel comfortable and the background should not be messy [14].

3 Methods

This research is separated into two parts: (a) Analyze the visual design element of the present female clothing photos found on on-line stores. (b) Discuss how visual design is composed through the use of multiple photography styles with varying emphasis on focal points.

3.1 Analyze the Morphology of Visual Element of Photography

- According to ARO(Access Rating Online), MIC(Market Intelligence & Consulting Institute), and Nielsen Net Watch, Yahoo Auction was Taiwan's most popular website for on-line selling [20]. Therefore photos found on Yahoo Auction from the past three years were collected and organized in order to analyze popular female brands.

- Scope of reference : To avoid multiple styles appearing among the research; female fashion for the age category of 18 to 28 was selected for this particular research. After collecting and filtering various fashion photos found within the Asian market, 40 product photos were chosen as samples.

- Method of analyzing： Using a sound level of photography knowledge, the visual design elements of the 40 sample photos in morphology perspective were analyzed.

3.2 Styles of Visual Design Elements on Photo

A focus group was formed to discuss the styles of visual design elements on photo.

Participators：

(1) Female consumer who has more than 3 years on-line shopping experience.
(2) Expert who has studied visual design for at least 2 years, has passed through a cognitive process and can perceive and share concepts with others.

Content of discussion：

(1) The 40 samples into groups and discuss the style and presentation of each group.
(2) Discuss the composition of the visual design element of each group.

4 Conclusions

4.1 Synthesize the Visual Design Element of the Present Female Clothing Photos on On-Line Store as Table 5

Table 5. The visual design element of the present female clothing photos on on-line store

	A Lighting	B Framing	C Composition	D Angle of Shooting	E Location	F Presentation of products
a	Front Light	Full Shot	Dot Composition	Eye Angle	Indoor	Floor photography
b	Plain Light	Medium Full Shot	S Composition	Low Angle	Outdoor	Hang Clothes on a hanger
c	Side Light	Medium Shot	Three aspects Composition	High Angle		Self-portraits by seller
d	Rembrandt Light		Two aspects Composition			Model
e			Circle Composition			
f			Triangle Composition			

4.2 After Synthesizing, the Photos Were Categorized into 4 Groups: (1) Modern Fashion, (2) Mellifluent Fashion, (3) Simple Fashion, and (4) Clothing Focus Fashion

1. Modern fashion: Photography shoots in nature scenes, street scenes or in front of a featured store all belong to this category. These types of scenes are usually seen in female fashion magazines. A sense of fashion can be felt on the streets of the city. This type of photography mostly uses side lighting and eye angle. In order to get the associative perception, full shot and three aspects composition are involved. Table 6 shows the composition of visual design element from the photo category of modern fashion.

Table 6. The visual design element of modern fashion

Sample	A lighting				B Framing			C Composition						D Angle of Shooting			E Location		F Presentation of products			
	a	b	c	d	a	b	c	a	b	c	d	e	f	a	b	c	a	b	a	b	c	d
1		✓			✓					✓				✓				✓				✓
2				✓	✓					✓					✓			✓				✓
3		✓			✓						✓					✓		✓				✓
4		✓			✓					✓				✓				✓				✓
5		✓				✓				✓				✓				✓				✓
6		✓			✓					✓					✓			✓				✓
7			✓				✓				✓			✓				✓				✓
8		✓				✓					✓				✓			✓				✓
9		✓			✓			✓						✓				✓				✓
10	✓				✓			✓						✓				✓				✓
11		✓			✓					✓					✓			✓				✓
12				✓	✓					✓				✓				✓				✓
13			✓		✓			✓						✓				✓				✓
14		✓				✓					✓			✓				✓				✓
15		✓			✓					✓				✓				✓				✓
16			✓		✓			✓						✓				✓				✓

2. Mellifluent fashion: white background is mostly applied in this category. It delivers a gentle, clean and calm feeling. This type of photo is usually shot indoors against white or wooden background. The photographer uses the reflections of natural lighting from outside due to space limitations. The photographer uses full shots with side lighting and high angles. Two aspects or three aspects composition is used to accurately place the model in the photograph. Table 7 shows the composition of visual design element from the mellifluent fashion category.

Table 7. Tthe visual design element of mellifluent fashion

Sample	A lighting				B Framing			C Composition						D Angle of Shooting			E Location		F Presentation of products			
	a	b	c	d	a	b	c	a	b	c	d	e	f	a	b	c	a	b	a	b	c	d
17		✓				✓					✓			✓			✓					✓
18				✓	✓			✓						✓			✓					✓
19	✓						✓					✓				✓	✓					✓
20				✓		✓				✓				✓			✓					✓
21			✓			✓				✓						✓	✓					✓
22				✓	✓			✓						✓			✓					✓
23			✓			✓					✓					✓	✓					✓
24	✓				✓						✓			✓			✓					✓
25			✓		✓			✓								✓	✓					✓
26			✓		✓			✓							✓		✓					✓

3. Simple fashion: photography in this category does not use models. Products are photographed on the floor against a white background. Even though the focal point is clear, the consumer will feel the product is too simple and the shopping desire is easily reduced. This type of photography is simple: products are located in the central of the photo. If the clothes are photographed on a hanger, the angle of shooting has to be parallel with the camera. If items are photographed on the floor, the angle of shooting has to be vertical with the camera. Table 8 shows the composition of visual design element from the simple fashion category.

Table 8. The visual design element of simple fashion

Sample	A Lighting				B Framing			C Composition						D Angle of Shooting			E Location		F Presentation of products			
	a	b	c	d	a	b	c	a	b	c	d	e	f	a	b	c	a	b	a	b	c	d
27	✓				✓			✓								✓	✓		✓			
28	✓				✓						✓			✓			✓			✓		
29	✓				✓					✓						✓	✓		✓			
30		✓			✓					✓				✓			✓			✓		
31		✓			✓							✓		✓			✓			✓		
32		✓			✓			✓						✓			✓			✓		

4. Clothing focus fashion: this category makes use of models for photo shoots. The angle of shooting attempts to keep the model's face from being revealed. Therefore this style creates a polarized feeling. Some may think it will emphasize the features on the clothes. Others, however, may be distracted by the fact that the model's face cannot be seen and thus believe it spoils the entire photograph. The photos in this category all use full shots. Three aspects composition is applied a lot when shooting takes place outside. When shooting

takes place indoors, two aspects composition is more likely to be applied with high angle shooting techniques. This will accurately place the model in the photograph. In comparison, shooting outdoors provides the photographer with greater available space. Table 9 shows the composition of visual design element from clothing focus fashion category.

Table 9. The visual design element of clothing focus fashion

Sample	A Lighting				B Framing			C Composition						D Angle of Shooting			E Location		F Presentation of products			
	a	b	c	d	a	b	c	a	b	c	d	e	f	a	b	c	a	b	a	b	c	d
33	✓				✓			✓						✓				✓				✓
34	✓				✓						✓				✓			✓				✓
35	✓				✓								✓			✓	✓				✓	
36		✓			✓				✓		✓			✓			✓					✓
37		✓			✓						✓			✓			✓					✓
38		✓			✓								✓	✓			✓					✓
39					✓								✓			✓	✓				✓	
40					✓				✓						✓		✓					✓

5 Discussion

In the past, recorded documents stated that low angle shots represented pride and distant feelings. High angle shots created quirky images. Nevertheless it was discovered that there isn't great difference among varying angle shots. After analyzing the photos, our assumption was in regard to tones and the way in which they can affect the impression made by a photo.

According to the discovery from the 40 product photos, it is found that the composition been applied the most were (1) three aspects composition, (2) two aspects composition and (3) dot composition. Three aspects composition demonstrates proportion, and it works on both vertical and horizontal shooting. Two aspects composition corresponds with full scene shooting and allows clothing items to stand out. Two aspects composition however lacks the associative perception. The purpose of using dot composition is to emphasize the creation of a shopping environment. In comparison to three aspects and two aspects composition, it cannot completely present the features and details of products. The most common framing method is full shot along with three aspects composition; the second most common method is medium full shot with side lighting. It can produce an outstanding result. Medium full shot has already usually applied in fashion magazines in the past because of the gentle natural lighting.

In conclusion of the above research, it is strongly believed that different composition can affect the feelings of female consumers. Although various angles of shootings did not appear to have much difference, the photos of products shot outside corresponded to lighting, framing, and composition outlines in recorded documents. More importantly, the critical element for presenting products on an on-line store is the location. Due to the increasing trend of on-line shopping, it is hoped this information can be of value to virtual stores in regard to visual design of female clothes.

References

1. Lin, H.C.: Taiwan's Online Auction Market of Women's Clothing and Accessories. National Digital Library of Theses and Dissertations in Taiwan, Taipei (2005)
2. Kim, J., Moon, J.Y.: Designing towards Emotional Usability in Customer Interfaces-trustworthiness of Cyber-Banking System Interfaces. Interacting with Computers 29(1), 1–29 (1998)
3. Chung, H.C.: Factors Contributing the Clothing Purchase on the Internet and Its Gender Effects. National Digital Library of Theses and Dissertations in Taiwan, Taipei (1999)
4. Yu, Y.W.: Applied Analysis on Experiential Design Elements of Online Auction Website - Take Super sellers of Women's Clothing on Yahoo! Kimo Auction Site for Example. National Digital Library of Theses and Dissertations in Taiwan, Taipei (2004)
5. DrMaster Press (Translator). Ultra-Efficient Products Online Photography (Notomi Yasukuni). Drmaster, Taipei (2007)
6. Wu, C.L. (Translator): Commercial Photography and Editing Skills (Kumon Yasush). GOTOP, Taipei (2010)
7. Misawa, A.: Camera Lesson Pro. JenJen, Taipei (2010)
8. Lee, K.H., Pu, T.Y., Kao, Y.H.: Win with Atmosphere! DSLR Photography Style Fully Resolve. Delight, Taipei (2009)
9. Chin, C.C.: Well Planning: Skill of Portrait Photography. DrMaster, Taipei (2010)
10. Chia-Yin Online: The World of Square: Successful Composition. GrandTech, Taipei (2010)
11. Hei, M.: Portrait Photography. Delight, Taipei (2008)
12. DrMaster Press (Translator): Product Photography Lighting, Composition and Post-Production Editing Techniques (PLTHINK). DrMaster, Taipei (2009)
13. Chia-Yin Online: The World of Square: Light of Photography. GrandTech, Taipei (2010)
14. David Image Studio: Digital Photography. Flag, Taipei (2007)
15. DrMaster Press (Translator). Graphic Portrait Photography: People Photography Composition. Metering and Exposure (Pearl Studio, BAEK Cheul-Min). DrMaster, Taipei (2010)
16. Lei, Y.L., Liang, T.: Portrait Photography. UNALIS, Taipei (2010)
17. Chen, H.J.: Seminar of Online Auction, Popularity Product Photo. Flag, Taipei (2006)
18. Shih, K.Y.: D-SLR Portrait Photography. To Lin, Taipei (2009)
19. Hicks, R.: Secrets of Light. Continuum International, GBR (1994)
20. Yahoo! Business Blog, `http://www.wretch.cc/blog/ycorpblog/12177400`

Web Application for Analysis, Manipulation and Generation of Accessible PDF Documents

Alireza Darvishy, Hans-Peter Hutter, and Oliver Mannhart

ZHAW Zurich University of Applied Sciences
InIT Institute of Applied Information Technology
P.O. Box 805
CH-8401 Winterthur, Switzerland
{alireza.darvishy,mano}@zhaw.ch

Abstract. This paper presents a software architecture for a web-based service that checks the accessibility of PDF documents and is capable of rendering them accessible. Users will be able to detect accessibility issues related to a PDF document and use a web browser to fix them. The implementation includes a user interface component as well as a PDF analysis and tagging engine (work in progress). The user interface component should be intuitive, in order to support users who do not have extensive knowledge regarding the accessibility of PDF documents. No additional software and installation is needed to use the system, the software should be available for use in educational and e-government applications.

Keywords: Accessibility, Document accessibility, Accessibility Analysis and manipulation of PDFs, visual impairment, tagged PDF, software architecture, user interface design, web content accessibility guidelines.

1 Introduction

The PDF standard is used in a variety of fields and preferred by many authors and authorities. Many PDF documents are available on internet platforms, unfortunately most of them are not accessible for users with disabilities. All major word processors (such as Microsoft Word [1], OpenOffice [2], LaTeX [3], etc.) support a way to export a document to a PDF file. Yet most of these methods do not produce tagged PDFs (which are required for accessible PDFs), or the resulting PDFs are tagged incompletely (not compliant with WCAG 2.0 [10]) and have to be corrected manually. Obviously this is a tedious work and requires specialist know-how. Consequently, the majority of the publishing community is unable to cope with this task. Two key-problems must therefore be addressed:

1. The lack of accessibility knowledge on the part of authors.
2. Different word processors are used for the generation of PDFs.

This project aims to develop a tool, which can be used with very limited knowledge of accessibility and does not require any installation on the local computer.

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 121–128, 2011.

2 Our Solution

One approach to solve these problems is to extend the functionality of the word processors [9]. However, this is a complex and time-consuming task since every tool has to be adapted.

Our solution therefore works directly on the PDFs themselves, since all major word processors are able to generate PDF documents. Our system is capable of checking PDFs as well as the correctness and completeness of false or missing accessibility tags. The following diagram shows a simplified outline of the architecture used for the web application.

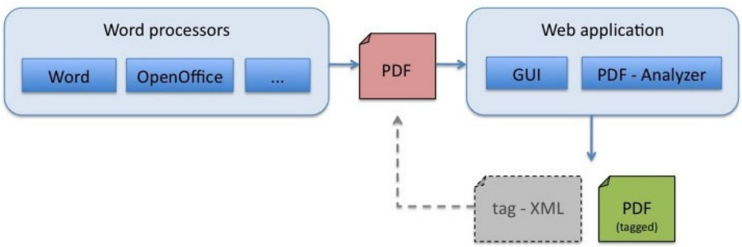


Fig. 1. Architecture of web application

As shown in Fig. 1, the application's output contains a fully accessible PDF and an optional file that contains tag information in XML format. XML is useful in several situations. For example, a student might use the web application to create an accessible PDF, and discover shortly thereafter spelling mistakes in the document. Thanks to XML, he or she does not have to reenter all of the accessibility information again. Instead, he or she can directly generate a new, fully accessible PDF. The XML file can also be used with a document that has a similar structure, a different number of images, or even new tables compared to the original PDF. The application will detect such discrepancies and ask only for the additional information it needs (e.g. the description for a new image).

Based on its concept and architecture, the web application liberates authors from secondary accessibility tools (e.g. accessibility checkers) and from the need for specific accessibility and tagging knowledge. Authors neither need to know how to tag PDFs nor what exactly has to be tagged. All this know-how is held by the application and is presented through a dynamic web page (GUI).

3 Problem Domain

Because a PDF does not necessarily have to be enriched with structural information, it can be very difficult to assemble this information retrospectively (a table could comprise only positioned text). It is therefore crucial to integrate the user, as only the user can reliably provide information on the logical structure of the PDF.

The system reads the PDF in two different ways. Firstly, it extracts existing structural information (tags), and secondly it proposes potential tags according to given rules. It can thus identify text blocks as paragraphs, for example. Such potential tags are presented to the user as errors because the system deems these tags to be missing. Of course, original tags with missing attributes are also displayed to the user as errors. In addition to marking and defining structural information, the reading order can also be derived from the tag tree of the PDF. A logical representation of a document (in this context, a PDF) has therefore been defined for the web application. It should be noted that although this representation does have similarities with the structure of a tagged PDF from ISO 32000-1 [12], it is not the same but only an abstraction of it. When a PDF is processed by the web application, however, all "standard structure types" defined in ISO 32000-1 must be mapped accordingly. Fig. 2 shows this logical representation as an example. It should be noted that neither types nor their attributes are shown in full detail. The graphic serves simply to illustrate the concept. The reason behind defining a logical representation of a document is that we want to separate the web application from the PDF standard. As a result, changes or extensions to the PDF standard lead only to reimplementations in the PDF parser.

3.1 Representation of a Document

A document comprises various elements. Derived from these basic elements are various types (section, heading, paragraph, etc.). Certain elements (e.g. section) can contain further elements. This results in a structure tree of the document. If it is processed depth-first, the result is the order in which the elements appear in the document.

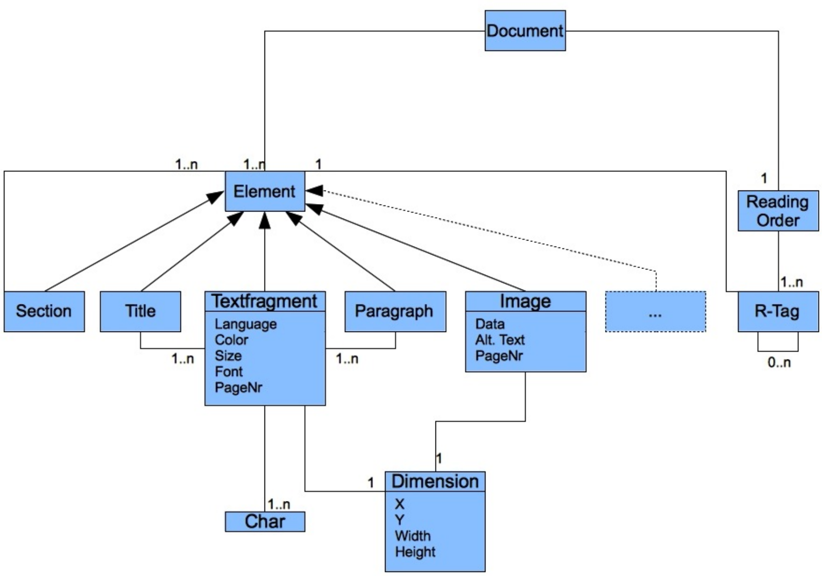


Fig. 2. Logical Structure

The important thing is that the page and position of each element can be determined. For a heading it can be determined via its text fragments. In the case of an image, on the other hand, the information pertaining to page and position can be determined directly (by means of dimension). These attributes are used primarily for the graphic presentation in the GUI.

In addition to the elements, a document also has a reading order which can differ from the absolute positioning of an element in the document. The reading order is also structured as a tree, whereby each R-tag of this reading order (see Fig. 2) references one element. Consequently, the reading order can be arranged freely without having to make any changes to the elements. If a tag does not reference an element, nothing is marked during synchronization with the PDF. The same occurs if an element without a counterpart is created in the reading order.

In order for the system to have this functionality, it has to be able to export the tags and errors from the PDF. To do so, it outputs the Reading Order tag tree, if available. For every tag, there must be a corresponding element in the PDF. So, an element tree can also be constructed (users can view incomplete elements displayed as errors in the GUI). Additionally, the system tries to detect elements not yet included in the tag tree. These are then added to the element tree (these tags are displayed as errors in the GUI because the system considers the elements still missing). Users can now modify or expand the tag or element tree via the GUI (e.g. correcting incomplete elements). Finally, the tag tree as well as the element tree is written back to the PDF.

4 User Interface Concept

This section explains the GUI concept in detail. Only the main components of the GUI will be examined in detail here and underlying ideas will be highlighted.

The user's main task is to correct errors (e.g. inserting an alternative text), ensure a logical reading order, and adjust, insert or delete existing tags. In the GUI presented here, these functions are shown within one view. Thus, the work to be done is immediately apparent. Additionally, discrepancies (incorrect reading orders, for example) are easy to recognize, as the GUI intuitively highlights whether or not the PDF is coherent.

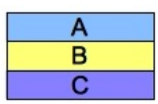


Fig. 3. Main Layout

The main GUI is divided into three sections (A, B, C), which will be discussed here in detail. These three sections are arranged on top of each other, as shown in Fig. 3.

4.1 A – Status Section

This section of the main GUI provides information on the status of the PDF. Thus, it is evident here whether or not it was possible to check the PDF and how many errors were detected. The user gets an overview of the PDF's status (from the accessibility point of view).

4.2 B – Unvisualized Events

A PDF not only displays errors or tags that concern a certain section of the PDF. The main language of the document cannot, for example, be assigned to a specific part of the PDF. Such errors therefore cannot be visually marked in the PDF and are thus itemized in a list (see Fig. 4).

Fig. 4. Unvisualized events

It should be noted that such errors or tags will not be translated to the PDF as tags. Most of them are PDF meta information. However, this is irrelevant for the user and introducing further terms (such as meta information) would only confuse the user.

Each line in the table shown in Fig. 4 comprises four cells. In addition to the type of error or tag (e.g. main language), the description, and the status, the user is given a number of options for dealing with tags or errors in the PDF. An error (see lines B1) can be corrected (option C in Fig. 4) or deleted (option D in Fig. 4). If an error is corrected, it does not disappear from the table, rather its status is changed and it becomes a full tag (see B2). It is important to leave corrected errors in the table so that the users can clearly see the results of their actions (e.g. correct an error).

If this logic is followed, it makes complete sense to also show tags in the same table. Unlike errors, however, tags can be edited not corrected.

The delete option, however, applies to both errors and tags. At first glance, it might appear counterproductive to delete an error. However, it must be assumed that the web application falsely identified the PDF structures (e.g. a table instead of simple text). In these cases, the user must be able to delete the incorrect error message.

4.3 C – PDF Visualization

Section C visualizes the PDF and the elements important in relation to accessibility. Section C primarily comprises three columns (see Fig. 5): The reading order on the left, the page thumbnails, and the actual tags and errors in the PDF on the right. Across these three columns, the correspondences of the "tagged PDF" are evident and can be intuitively manipulated. In the case of PDF documents with many pages, the PDF visualization is split over several pages. The user can then toggles between the individual pages or skip straight to the next error via the corresponding links.

Visualization: Reading Order. All tags/errors are listed in the reading order. Additionally, each element in the reading order displays a reference number that appears again both in the thumbnails and also in the tag/error visualization. This enables the user to quickly find the origin of a tag.

Certain tags (the section tag in Fig. 5, for example) can contain further tags. Such tags are displayed in the reading order with two tags. One of the tags marks the start and the other marks the end of the tagged section. These special tags, however, are used only for approximate structural markings (part, section, article, etc.). A table, by contrast, is displayed as a tag although strictly speaking it comprises several tags. The reading order within a table is then processed by means of sub-dialogs that have a structure identical to the GUI in Fig. 5.

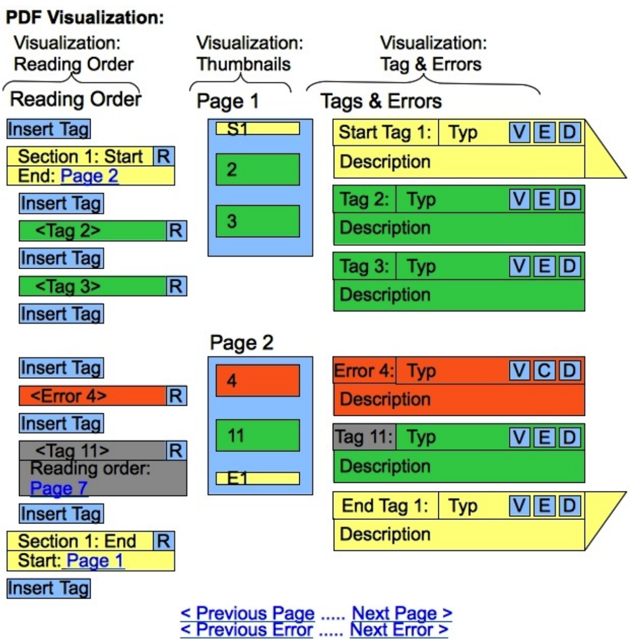


Fig. 5. PDF Visualization

The user is also to be given the option of inserting new tags into the reading order (marked in Fig. 5 with "Insert Tag") or changing the reading order (marked in Fig. 5 with "R").

However, the reading order can differ from the actual sequence of document parts. This aspect is also considered in the GUI calculation. Tag 11 in Fig. 5 depicts such a case. This tag is not supposed to appear in the reading order until much later (e.g. on page 7). Such a tag is shown twice in the GUI . Once to denote the absolute positioning of the tag with a reference to the page in the reading order (page 7 in this example). It has a distinctly different color to normal tags. The tag would also appear in the reading order on page 7 (see Fig. 6). Of course, the tag can now no longer be visualized in the thumbnails because it is actually located on page 2. Therefore, the user is again given a reference to the actual tag page (page 2 in our example).

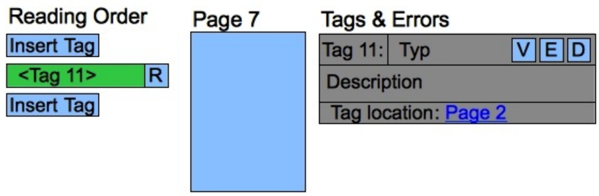


Fig. 6. Reading Order

Visualization: Thumbnails. The thumbnails are the key element of section C. They allow the user to quickly find things in the document and adjust the reading order and tags accordingly.

A thumbnail is a miniature view of a PDF page. One thumbnail is therefore created for each page. In the thumbnails (see Fig. 5), the individual tags or errors are highlighted in a different color. An error (e.g. image without alternative text) appears in red. A tag (e.g. a heading), on the other hand, is shown in green. In addition, each color marking is assigned a unique tag number that also appears in the reading order.

Visualization: Tag & Errors. The third and last part of section C contains the tags and errors that were found in the PDF. It should be noted that the errors are displayed only when the PDF has been checked. The GUI described in this section, however, can also be used to show tags only. As clearly shown in Fig. 5, the visualization of tags & errors corresponds to a detailed view of the PDF sections marked in the thumbnails. Besides various information on the tags or errors, it is particularly important to highlight the functions accessible to the user via this view. In Fig. 5, these are symbolized by the letters V, E, D, and C. V (View) is used to select a sub-dialog that displays the full view of the corresponding page and the tag (without thumbnail). E (Edit) can be used to edit a tag (e.g. change the type or an attribute). Other dialogs are of course selected, depending on the tag. However, listing them all would exceed the scope of this work. D (Delete) deletes the corresponding tag or error. C (Correct) allows an error to be corrected and then converted into a correct tag. The Tag & Error view also contains the nested structural tags (e.g. section tag 1).

5 Conclusion

The software architecture presented in this paper is suited for a broad audience (with no specific know-how required) and could therefore make a difference in the quality and quantity of fully accessible documents. Based on the proposed architecture, a prototype application is currently being implemented. The main advantage of the proposed concept is that it is independent from the authoring software (such as e.g. MS Office [8]), which results in lower maintenance and support costs.

References

1. Microsoft Office Word, `http://office.microsoft.com/word/`
2. OpenOffice, `http://www.openoffice.org/`
3. LaTeX, `http://www.latex-project.org/`
4. PDF Accesibility Checker, `http://www.access-for-all.ch`
5. Freedom Scientific JAWS for Windows Screen Reading Software, `http://www.freedomscientific.com/products/fs/jaws-product-page.asp`
6. NetCentric PDF Accessibility Wizard for MS Office, `http://www.net-centric.com/products/PAW.aspx`
7. Virtual508.com Accessible Wizard for Microsoft Office, `http://www.virtual508.com`
8. Microsoft Office (2010), `http://us1.office2010beta.microsoft.com`
9. Darvishy, A., Hutter, H.-P., Horvath, A., Dorigo, M.: Dorigo: A Flexible Software Architecture Concept for the Creation of Accessible PDF Documents. In: Miesenberger, K., Klaus, J., Zagler, W., Karshmer, A. (eds.) ICCHP 2010. LNCS, vol. 6179, pp. 47–52. Springer, Heidelberg (2010)
10. WCAG 2.0, `http://www.w3.org/TR/WCAG20/`
11. eGov PDF Checker, `http://accessibility.egovmon.no/en/pdfcheck/`
12. ISO 32000-1:2008. First Edition 2008-7-1: Document management - Portable document format - Part 1: PDF 1.7, pp. 573-610, `http://www.iso.org/`

Survey: Improving Document Accessibility from the Blind and Visually Impaired User's Point of View

Martin Dorigo[1], Bettina Harriehausen-Mühlbauer[2], Ingo Stengel[2], and Paul S. Dowland[1]

[1] Centre for Security, Communications and Network Research, University of Plymouth, Plymouth, United Kingdom
[2] University of Applied Sciences Darmstadt, Darmstadt, Germany
martin.dorigo@plymouth.ac.uk

Abstract. There are thousands of digital documents available on the internet, but many of them are not accessible for blind and visually impaired people. To find out what is of importance as to the reading of text and the navigation within documents from the user's point of view, a survey has been conducted among people concerned. They were asked how they handle text-documents, which problems they have and which features they would wish to have for a novel concept of assistive technology. The analysis of feedbacks from 205 participants leads to four key issues: The people concerned would like to have better access to digital documents as well as a better overview over their structure and content. In addition, the assistive technologies should be more easy, intuitive and standardised to use. A discussion is provided where the outline of preliminary solutions for the issues mentioned is proposed.

Keywords: Document accessibility, Visual impairment, Blind, Assistive Technology.

1 Introduction

1.1 Essential Components

On its way from the author to the reader with a disability, a digital document passes several components. To provide full access for the user, these components have to work together. If an accessibility feature is not implemented in one component, it does not result in an accessible user experience [1].

The author uses an authoring tool to create a document. The Authoring Tool Accessibility Guidelines (ATAG) [2] defines how authoring tools should help authors produce content that is accessible and conforms to the Web Content Accessibility Guidelines (WCAG) [3]. The content is typically read by a user agent. The User Agent Accessibility Guidelines (UAAG) [4] explains how to make user agents accessible to people with disabilities by passing the relevant information to the Accessibility API [5] of the operation system. This API helps assistive technologies like Screen Reader [6] to provide the information to the user with a disability.

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 129–135, 2011.

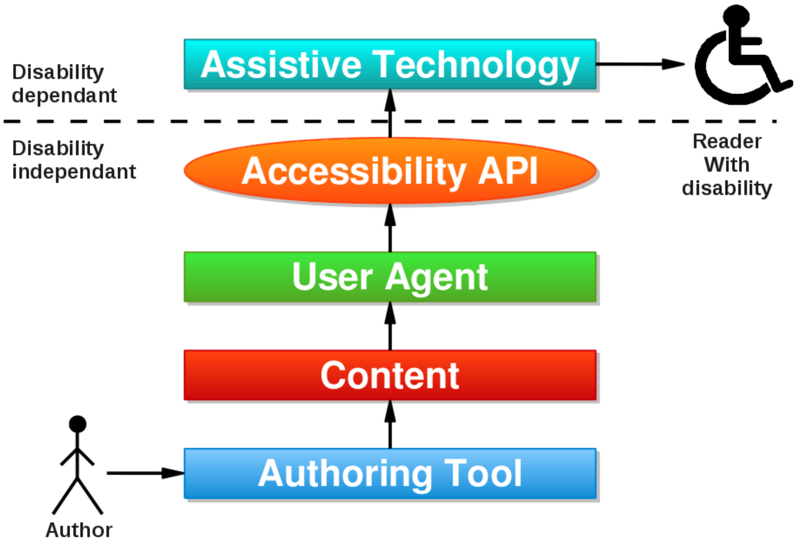


Fig. 1. Essential Components Overview

1.2 Document Abstraction Layers

During the processing by the different components, digital documents take on different forms. A static electronic document can be broken down into four abstraction layers [7].

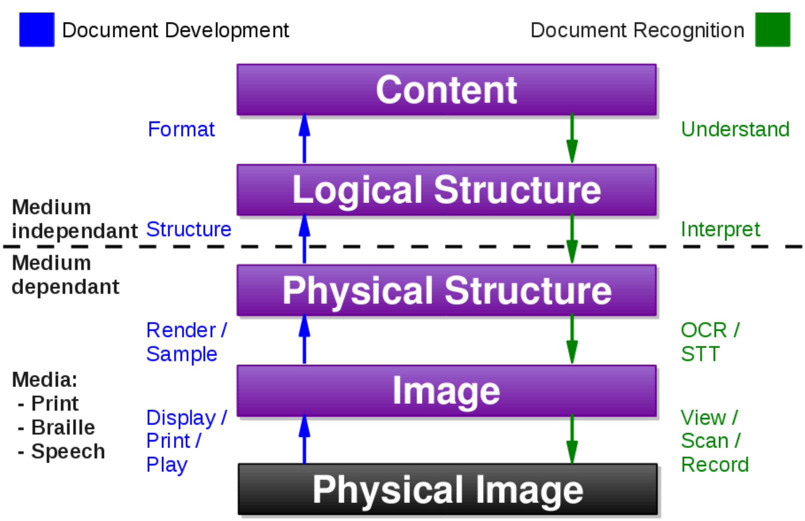


Fig. 2. Document Abstraction Layers

After the documents content had been organized into a logical structure, it is formatted into the physical structure of a specific medium. In addition to print, other media relevant to blind and visually impaired are speech and braille. The physical structure can be rendered to an image, which is than displayed to the user by a specific physical device.

While the development can be easily done automatically by a machine, the document recognition on the other hand is very computationally intensive and defective. In order to make a document accessible to people with a disability, using different media, its logical structure must be available.

2 Method

2.1 Participants

25‘000 blind and visually impaired people had been enquired in Austria, Germany and Switzerland. Feedbacks from 205 participants in the age from 18 to 80 years as well as in various occupational and educational situations were received. 55% male are accompanied by 45% female. The participants had been categorized into 6 categories of visual impairment according to the International Classification of Diseases (ICD) [8]. 27% of them are blind with no light perception, 22% with light perception and 16% with a visus worse than 0.05. 18% suffer from a severe visual impairment (visus worse than 0.1), 11% from a moderate visual impairment (visus worse than 0.3) and 6% from a mild visual impairment (visus worse than 1.0) [Fig. 3]. 61% of them are visually impaired by birth.

13% of all participants classified their computer literacy as little, 60% as good and 26% as very good [Fig. 4].

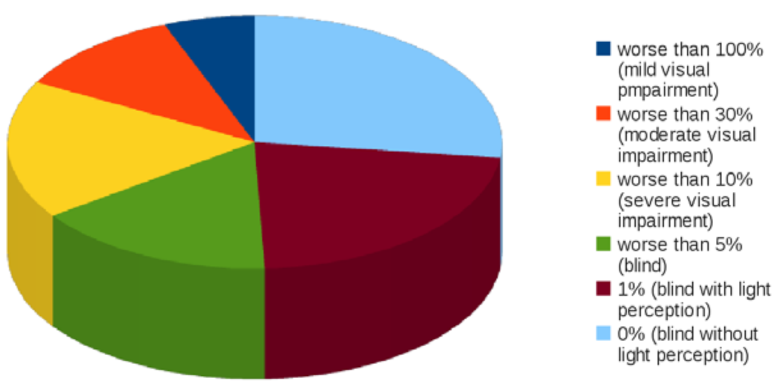


Fig. 3. Categories of visual impairment

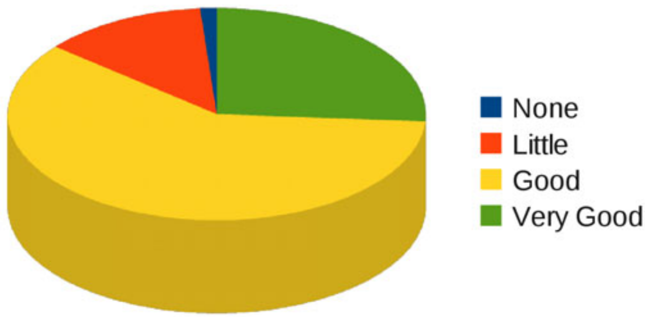


Fig. 4. Categories of computer literacy

2.2 Questionnaire

The questionnaire consists of 16 questions and covers the following areas of document accessibility. Most of the questions are multiple choices with a nominal or ordinal scale as well as a free text field for additional comments.

- Satisfaction
- Media and formats provided and wished
- Assistive technologies, input and other electronic devices in use
- Structural information
- Problems and solutions
- Wishes for a novel concept of an assistive technology
- Personal information: Visual impairment, occupational and educational situation and optionally contact details margin.

2.3 Procedure

Through organisations, online mailing lists, relevant forums and blogs as well as in paper based magazines blind and visually impaired people were invited to participate in this survey. In that invitation, the goal of the research was introduced to them. The participants could choose if they would like to take part online via the link provided or by phone in form of an interview. In the online survey, the quantitative aspect was promised while for people, questioned by phone, the qualitative aspect was centered. The answering took about 30 minutes. After the analysation of the results, four essential questions were resent to the 205 participants via email to get more detailed information.

3 Results

3.1 Digital Formats

The digital documents provided to blind and visually impaired people are a mix of different formats [Fig. 5]. The formats most used are HTML, Microsoft Word and PDF.

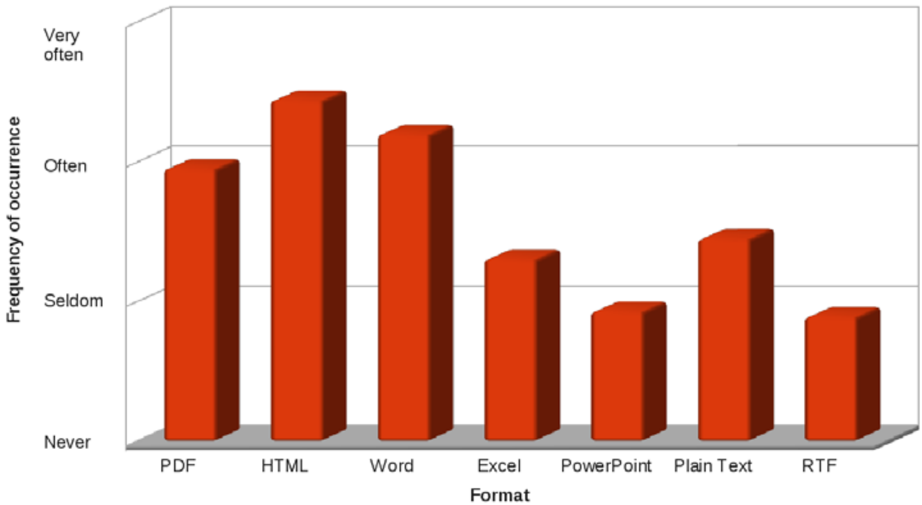


Fig. 5. Formats provided to blind and visually impaired

3.2 Issues

At the moment blind and visually impaired people are not very satisfied with their current situation as to the reading of text-documents. The problems and wishes mentioned lead to four key issues, where further research is needed.

Structure Overview. Speech can only provide one word at a given time. Braille can only show one line and people, working with heavy magnification can only see a single detail in a time. All of the participants wish a better overview over the structure

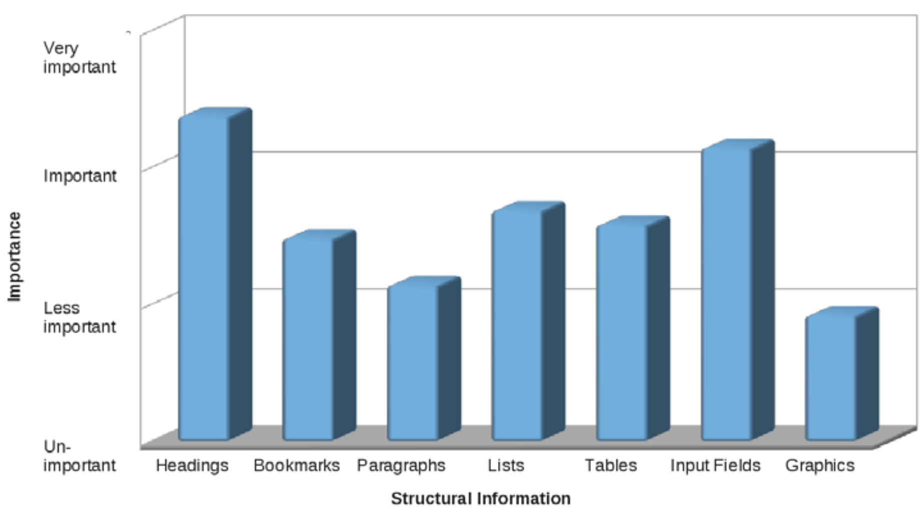


Fig. 6. Importance of Structural Informations

of a document. The most important structural informations are headings and input fields followed by lists and tables. The least important element for blind and visually impaired is graphics [Fig. 6].

Content Overview. For blind and visually impaired people it is not possible to skim read a text passage to get a fast overview over the contained information. A faster overview over the document's content more than over the document's structure is wished.

Usability of Assistive Technologies. In order to read an digital document efficiently many key sequences have to be known by heart and they differentiate between specific products. An exception is Apple VoiceOver [9]. Especially people with little computer literacy wish that the assistive technology is more easy, intuitive and standardised to use.

Accessibility. All of the participants mentioned that they get documents which they cannot access. The concrete problems depend on the visual acuity a specific person comes with [Fig. 7].

All of the participants wish better accessibility of digital documents.

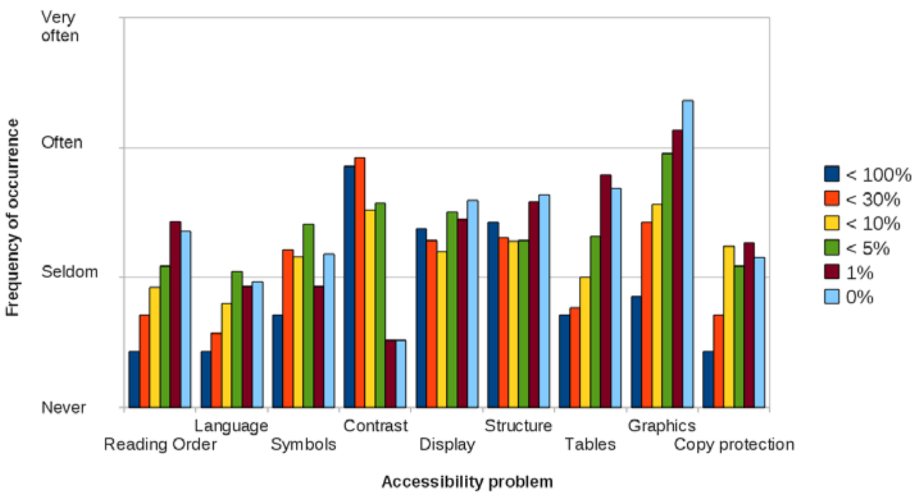


Fig. 7. Occurrence of accessibility problems

4 Discussion and Further Work

There is serious need for action. A concept should be made to display an overview over the document structure on other media such as speech and braille. In order to provide a content overview, the assistive technology needs comprehension over the document's content by using text mining. The usability may be improved by using alternative input devices like speech recognition or gestures. Accessibility is not a problem of missing theories rather than that they are not moved in the praxis. To improve this situation software may be developed which advises the author against potentially accessibility problems [10] or recognizes the document structure automatically [11].

References

1. W3C WAI Essential Components of Web Accessibility,
`http://www.w3.org/WAI/intro/components.php`
2. W3C WAI Authoring Tool Accessibility Guidelines (ATAG) 2.0,
`http://www.w3.org/TR/ATAG20/`
3. W3C WAI Web Content Accessibility Guidelines (WCAG) 2.0,
`http://www.w3.org/TR/WCAG20/`
4. W3C WAI User Agent Accessibility Guidelines (UAAG) 2.0,
`http://www.w3.org/TR/UAAG20/`
5. Microsoft Active Accessibility (MSAA),
`http://msdn.microsoft.com/en-us/library/dd373592%28VS.85%29.aspx`
6. Freedom Scientific - JAWS for Windows Screen Reading Software,
`http://www.freedomscientific.com/products/fs/jaws-product-page.asp`
7. Brugger, R.: Eine statistische Methode zur Erkennung von Dokumentstrukturen. PhD thesis no. 1251, University of Fribourg (1998)
8. WHO International Classification of Diseases (ICD),
`http://www.who.int/classifications/icd/en/`
9. Apple Accessibility VoiceOver,
`http://www.apple.com/accessibility/voiceover/`
10. Davishy, A., Hutter, H.P., Horvath, A., Dorigo, M.: A Flexible Software Architecture Concept for the Creation of Accessible PDF Documents. In: Miesenberger, K., Klaus, J., Zagler, W., Karshmer, A. (eds.) ICCHP 2010. LNCS, vol. 6179, pp. 47–52. Springer, Heidelberg (2010)
11. Ruemer, R., Miesenberger, K., Kummer, F., Gravenhorst, C.: Improving the Redigitisation Process by Using Software with Automatic Metadata Detection. In: Miesenberger, K., Klaus, J., Zagler, W., Karshmer, A. (eds.) ICCHP 2010. LNCS, vol. 6179, pp. 35–42. Springer, Heidelberg (2010)

Readable Image for the Visually Impaired

Sunil Kumar Kopparapu

TCS Innovation Labs Mumbai
Tata Consultancy Service,
Yantra Park, Thane (West), Maharastra 400601, INDIA
SunilKumar.Kopparapu@TCS.Com

Abstract. A picture is worth a thousand words is a well know adage and refers to the fact that any complex word description can be conveyed easily and quickly with a still image. The use of images, pictures and block diagrams to describe something is often used without a second thought. The use of images in any form becomes an accessibility issue for the visually impaired. With the concept of a document and the hyper text markup language (HTML) page blurring, W3C has come up with measures to make the images accessible on the web. The most recent being the use of `alt` attribute, which is designed to be an alternative text description for images on web pages and `longdesc` attribute which is a mechanism to give greater details of the image. In this paper, we propose an approach which enables accessibility of images. This paper has two parts, the first part describes a mechanism to build a *multi level description* of an image to enable accessibility or readability while the second part describes an user interface that enables navigation of the multi level description by hovering on the image.

1 Introduction

A picture is worth a thousand words is a well know adage and refers to the fact that any complex description of words can be conveyed, probably better, with a still image. The use of images to describe something is often used without a second thought and serves the purpose of compressing a large number of textual words. The use of images in any form becomes an accessibility issue for the visually impaired. Images can present a major obstacle to individuals that are blind or have low vision. With the proliferation of web usage the accessibility of a document has become synonymous with the accessibility of web document. In this paper we restrict ourselves to accessibility of images which are integral part of any document in general. There are several relatively simple techniques that can make an image accessible on the web [4]. "Section 508 Standard (a)" [15] of the US government addresses proper use of an image for accessibility. The world wide web consortium (W3C) [16] mandates the use of `alt` and `long-desc` attribute in hyper text markup language (HTML). While `alt` is designed to be an alternative text description for images the `long-desc` is a longish version of the alternative text description of the image. Most the discussion on accessibility have been restricted to making known what the image represents at a very

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 136–145, 2011.

broad level, typically a a crypt caption of the image. Accessibility of graphics specifically in technical documentation is discussed at length in [10]. However discussion on making any image accessible or readable has not been addressed in literature specifically. In this paper we develop a method that allows making any image accessible so that it can be read by the visually impaired.

The rest of the paper is organized as follows, in Section 2 we describe the problem with an example. We look at an overview of related literature in Section 3 and introduce the contribution of this paper. We describe our work on image readability (accessibility) in Section 4 and conclude in Section 5.

2 The Problem

Consider the image for example Fig. 1. This as a standalone picture or in a document or in a web page [7] is not completely accessible to the visually impaired. The only description of the image is through the caption (*Description 1.*) of the image.

Description 1. *Two People conversing*

Notice that there is certainly a longer description that is possible which the normal person would *read* in the image, for example, *Description 2*.

Description 2. *Two people, one in a black coat and a red tie with a black eye wear, balding, fair, $\cdots$ sitting on the left $\cdots$ and the other in a white shirt and left*

Fig. 1. `converse.jpg`: "Two people conversing" (from [7])

leg over the right leg, dark complexioned, with black hair and beard ··· sitting on the left of the person in coat ··· sitting on a brownish wooden bench set against a dark brown checkered wall bearing a caption WINE SALES written in Roman all capital Font in light brown with bushes of flowers colored red and white to their right and red and violet to their left with all the flowers in front and some more flowers handing from the top colored yellow and violet with green leaves surrounding them, <more description> conversing.

While *Description* 2. enables a person to *read* the image and make it more accessible, however it is not unique. Another possible description enabling accessibility of the image could be as shown in *Description* 3.

Description 3. *Two people, sitting on a brownish wooden bench set against a dark brown checkered wall bearing a caption WINE SALES written in Roman all capital Font in light brown with bushes of flowers colored red and white to their right and red and violet to their left with all the flowers in front and some more flowers handing from the top colored yellow and violet with green leaves, one person is in a black coat and a red tie with a black eye wear, balding, fair, ··· sitting on the left ··· while the other in a white shirt and left leg over the right leg, dark complexioned, with black hair and beard ··· sitting on the left of the person in coat ··· surrounding them, <more description> conversing.*

While *Description* 1. is cryptic and conveys an *overall* description of the image (a typical caption associated with an image), the *Description* 2. is more *informative* and captures finer details embedded in the image. The description a user infers when looking at this image lies probably somewhere in between *Description* 1. and *Description* 2. Depending on the personal interest, background and the context in which the user is looking at the image the longish yet sequential description (*Description* 2.) can change. One way is to describe the surroundings (about the flowers etc) in the image first and then tell about the people in the image, say as in *Description* 3. Notice that there is (a) no one sequential way of describing an image and (b) there is no limit to the amount of description, say in words, that one can associate with the image. Several researchers, for example, [2,5,12,9,13,11], have shown that the sequence in which the image is visualized by people depends on the cultural background of the person viewing the image. In this paper, given that the description of the same image could be very brief or very long and when description is long there is no one preferred sequence in which the image cab be described, we develop a method that allows an user to read the description of the image the way he desire, both in terms of the sequence and the density of description.

3 Related Literature

The issue of image accessibility has been discussed in literature sparsely and most of them are found in terms of patents. Accessibility of graphics in technical documentation [10] discusses a mechanism to read out block diagrams and the likes of it which are usually accompanied in a technical document, in some sense it uses the descriptions associated with the generation of the block diagrams.

A recent published patent application [3] describes a method for rendering annotations associated with dense and huge map images, typically of size 5 GB with limited viewing capability, say 1 MB. They describe an interface to enable panning and zooming of the image which has an annotations with respect to a location on the image. The annotations content can be an audio loop, narrative audio, text labels, etc. The essential aspect is that the huge image has been annotated at a single level and the method of zooming and panning essentially brings into viewing focus a smaller part of a huge image and hence making only the description associated with that that part of the image active. In short, the image is annotated at one level only and only that annotation that corresponds to the part of the image in focus becomes active.

In [14] a method and system to process a Digital Image is describe. It describes a method to embed an audio data within an image to provide an embedded image wherein the audio data is freely recoverable from the embedded image. This process embeds an audio stream into an image and constructs a new image. The patent also speaks of storing more or less audio data on the image in terms of the the audio quality of the embedded audio.

In [6] a method to capture an image and encode the audio data using markings in the image that are substantially imperceptible to an unaided eye of a human viewer is discussed. It talks about making *a priori* markings in an image which would be mapped to a corresponding audio embedded into the image file. This patent also speaks of embedding audio file into the image and making that portion of the audio active which corresponds to the image. Along similar lines, in the patent [1] a method that enables visually impaired users to navigate websites and hear high quality audio of narration and description of each website is disused. The system involves creating an audible website corresponding to an original website by utilizing voice talent to read and describe web content and creating audio files for each section of the original website and then assign these audio files to the respective sections of the website. Text, images, and other rich media content on the website are represented by audio files. However the description of images is restricted to the caption associated with the image or the `alt` tag associated with the web page.

While the attempt is towards making image accessible they do not address the problem stated in Section 2 namely that of allowing the user to choose the amount of description that they can see and in the sequence in which they desire to see the description. The main contribution of this paper is in identifying a method which allows the user to read a description of the image in the sequence they desire (*Description 2.* or *Description 3.*) and at the level of description (*Description 1.* versus *Description 2.*) they desire. In brief, we describe a method that specifically describes an image at multiple levels and allows the user to jump the description sequence and density asynchronously to read the description of the image.

4 Readable Images

There are essentially two parts in making images accessible. The first part is to build and organize the description associated with the image with the express view to enable readability or accessibility of the image. We describe a mechanism to annotate an image at several levels. At one end of the annotation spectrum, namely the finest level is the description of each and every pixel corresponding to the image while at the other end of the spectrum is a gross description of the complete image, namely all pixels. In between these two levels of description are several levels of descriptions. The second part is a user interface that enables a visually impaired person navigate the description of the image by moving the pointer device (a mouse) on the image. Specifically the position of the mouse on the image will make all the descriptions that have been annotated corresponding to that (x, y) location of the mouse active and the up or down scroll of the scroll button on the mouse will activate a description at the finer level or the coarser description of the image respectively. A text to speech engine (TTS) then reads out the description enabling the visually impaired person navigate and read the

```
<description>
    <image>
        <name> converse.jpg </name>
        <size> M x N </size>
    </image>

    <pixel, x, y>
        <level 1>
            <des> Black </des>
        </level 1>
        <level 2>
            <des> Hair </des>
        </level 2>
        <level 3>
            <des> Head </des>
        </level 3>
        <level 4>
            <des> A person </des>
            <des> Sitting to the right </des>
        </level 4>
        ...
        ...
        ...
        <level K>
            <des> Two people having conversation </des>
        </level K>
    </pixel, x, y>
</description>
```

Fig. 2. A Typical Multi level description (`converse.des`)

image he desires. The image (see Fig. 1, `converse.jpg`) could be a jpg, pgm, ppm, gif, bmp or any of the multitude of known image formats and the multi level description of this image is captured as a text file (see Fig. 2, `converse.des`). In this paper, we show (a) a mechanism (manual or semi supervised) to create the description file (Fig. 2), (b) representation of the description at different levels and (c) a method to access the multi level description of the image there by making the image accessible.

Consider an image I to be made up of $M \times N$ pixels, and to represent a pixel in the $(k,l)^{th}$ position we use the notation $I(k,l)$. Clearly k can take values from 1 to M while l can take values from 1 to N. Let there be K levels of descriptions, level K being the coarsest which could be the caption of the image (*Description 1*. , namely, "Two people conversing") and level 1 be the finest description of the image (every pixel described). Level n has a coarser descriptions than level $n+1$.

4.1 Building Multi Level Description

The method of creating multi level description of an image is semi supervised. The finest level details corresponding to each pixel can be captured automatically by identifying the color of the pixel so at `level` 1 there is a description for each pixel, meaning there are MN descriptions associated with the image at `level` 1. The next level, `level 2` description can be achieved by image segmentation [8] which allows grouping of pixels that have similar properties, say pixels having the same texture. In Fig. 1 it could be the region associated with the brick wall

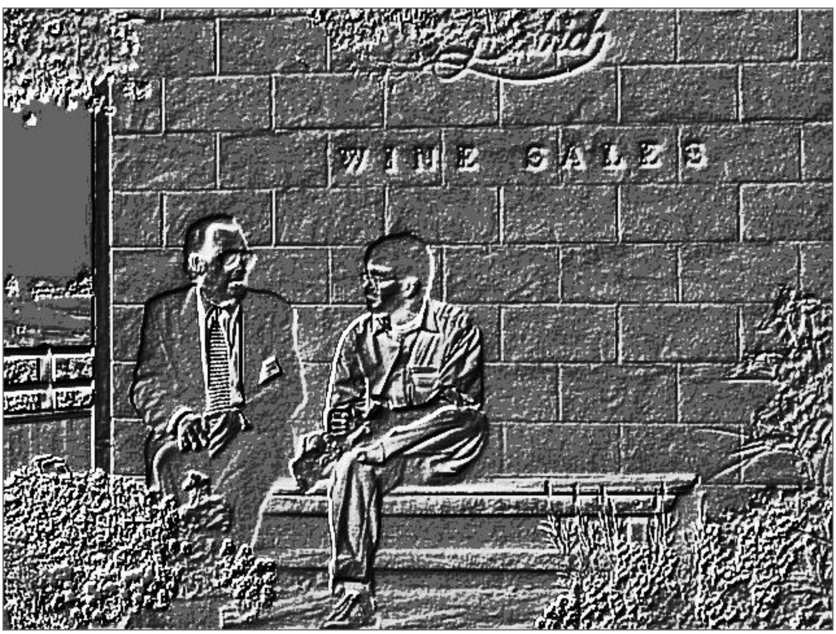


Fig. 3. Different level image annotation

behind the two people having a conversation. So a pixel in this area would have a `level 1` description as brown color while the `level 2` description would be brick wall. For example Fig. 3 has clearly very fewer details. For example, there are flower bushes, the brick wall, two people sitting next to each other, a bench. The annotation at this level would have only these broad description. Note that the same pixel will have multiple descriptions based on the level at which it has been annotated. For example a pixel on the head of the person sitting to the right (in white shirt) would have a `level 1` description of black a `level 2` description of hair, a `level 3` description of head, a `level 4` description of person to the right, sitting on a bench, in front of the wall, a `level 5` description of conversing with a person on the left and so on until `level` K description Two people conversing or the caption of the Fig. 1. A typical `converse.des` would have a structure, as shown in Fig. 2. The method does not embed the description into the image file as an audio stream as discussed in a couple of patents in Section 3 but keeps

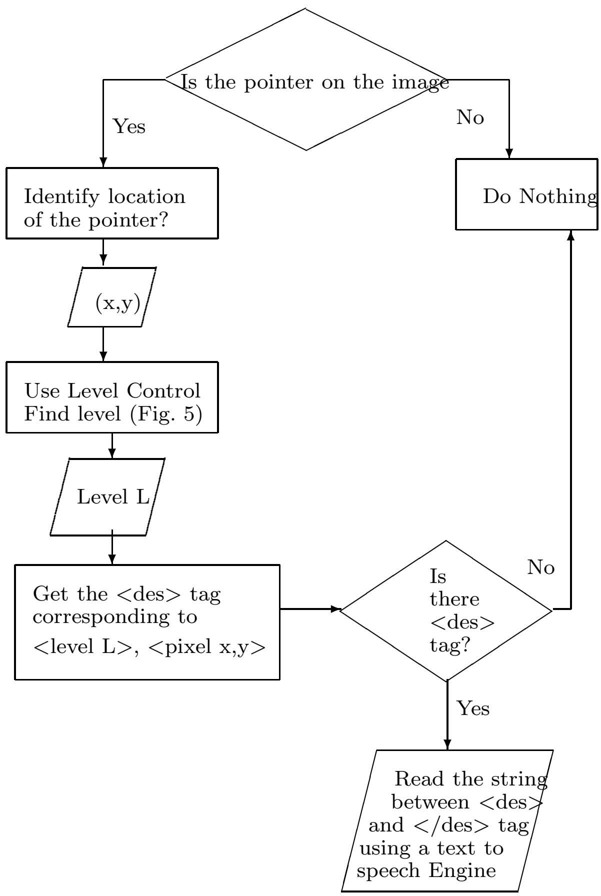


Fig. 4. Accessing multi level description of an image, given the image and the multi level description file

the image file (`converse.jpg`) and the description file (`converse.des`) separate, further the description is stored as text information and not as an audio file.

4.2 Accessing Multi Level Description

Accessing the multi level description associated with an image is shown as a flow chart in Fig. 4. A user interface has the capability of rendering the image and also determining the spatial and the scroll position of the pointer on the rendered image. As shown in Fig. 4, the accessibility of the image based on the location of the pointer (typically a mouse pointer) on the image. When the pointer is on the image, the location of the pointer is determined (say (x, y)) and the level identified (say `level` L). Then the string between the `<des>` and `</des>` tags corresponding to the pointer location (x, y) and `level` L is extracted from the description file (`converse.des`). This string is then read out using a text to speech engine or a screen reader. The accessibility of the image at different levels is possible through level sensing, which is enabled through the scroll button the mouse. Fig. 5 shows the control which allows the selection of the level of description by moving the scroll button of the mouse up or down. The level control makes sure that the `level` remains between 1 and maximum number of levels (say, K).

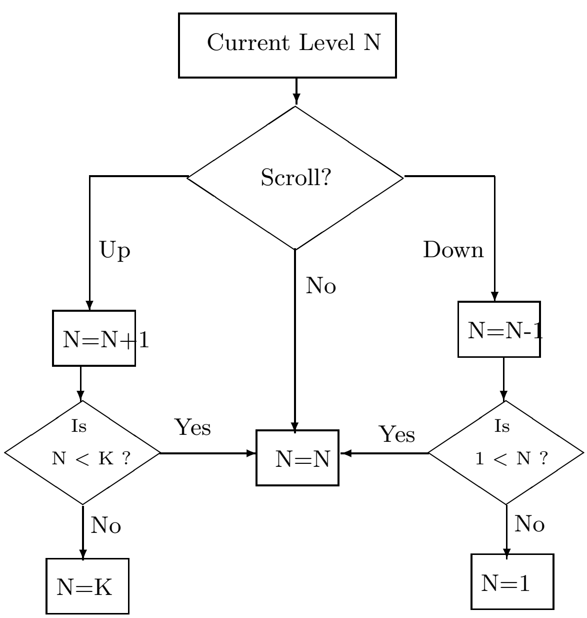


Fig. 5. Level Control to find `level` and K is the maximum number of levels at which an image is described

5 Conclusions

Use of images in documents or otherwise is very common in technical literature. Any kind of image in a document fuels accessibility issue to the visually impaired. While attempts have been made to make images accessible in technical documents by providing longish descriptions there is however no attempt to the best of the authors knowledge to make an image in general accessible. In this paper we described a method which allows images to be read (and thus accessible) by the visually impaired. We described a method that can be used to capture description of the image at multiple levels and showed how this multi level description associated with an image can be accessed by the visually impaired using a simple to use user interface. The proposed method not only enables a method to access image but also allows access information about the image to the desired depth and in the order in which the user is interested to read the image. This aspect removes the restriction that the user can read the image in only the sequence in which the image description has been captured by the creator of the description. The key contribution of this paper are (a) a method to annotate images at different levels, (b) a method to access information about the image at different levels, (c) a method to capture annotation in a description file at different levels and (d) a method that makes accessible an image in an non-sequential fashion.

Acknowledgments

The author would like to thank several members of the TCS Innovation Labs - Mumbai who have been instrumental in refining the idea of making an image readable.

References

1. Bradley, N.T.: Method and apparatus for website navigation by the visually impaired. US Patent Publication Publication No. US 7653544 B2 (2010), http://ip.com/patent/US7653544
2. Chua, H.F., Boland, J.E., Nisbett, R.E.: Cultural variation in eye movements during scene perception. Proceedings of the National Academy of Sciences of the United States of America 102(35), 12629–12633 (2005), http://www.pnas.org/content/102/35/12629.full.pdf+html, doi:10.1073/pnas.0506162102
3. Cohen, M.: Rendering annotations for images. US Patent Application No. US 2010/0085383 A1 published on (April 8, 2010), http://ip.com/patapp/US20100085383
4. Creating accessible images, http://www.doit.wisc.edu/accessibility/online-course/standards/images.htm
5. Evans, K., Rotello, C.M., Li, X., Rayner, K.: Scene perception and memory revealed by eye movements and receiver-operating characteristic analyses: Does a cultural difference truly exist? The Quarterly Journal of Experimental Psychology 62(2), 276–285 (2009), http://www.informaworld.com/10.1080/17470210802373720

6. Inness, G.: Picturs with embedded data. US Patent Publication No. US 2005/0068589 A1 published on (March 31, 2005), http://ip.com/patapp/US20050068589
7. IVCNZ 1998 (1998), http://www.citr.auckland.ac.nz/~ivcnz98/pictures/wineHouse_8.jpg
8. Kang, W.X., Yang, Q.Q., Liang, R.P.: The comparative research on image segmentation algorithms. In: International Workshop on Education Technology and Computer Science, vol. 2, pp. 703–707 (2009)
9. Miellet, S., Zhou, X., He, L., Rodger, H., Caldara, R.: Investigating cultural diversity for extrafoveal information use in visual scenes. Journal of Vision 10(6) (2010), http://www.journalofvision.org/content/10/6/21.abstract
10. Murphy, S.: Accessibility of graphics in technical documentation for the cognitive and visually impaired. In: Proceedings of the 23rd Annual International Conference on Design of Communication: Documenting & Designing for Pervasive Information, SIGDOC 2005, pp. 12–17. ACM, New York (2005), http://doi.acm.org/10.1145/1085313.1085320
11. Associated Press: In Asia, the eyes have it, http://www.wired.com/culture/lifestyle/news/2005/08/68626
12. Rayner, K., Castelhano, M.S., Yang, J.: Eye movements when looking at unusual/weird scenes: Are there cultural differences? Journal of Experimental Psychology: Learning, Memory, and Cognition 35(1), 254–259 (2009)
13. Roach, J.: Chinese, Americans truly see differently, study says, http://news.nationalgeographic.com/news/2005/08/0822_050822_chinese.html
14. SIM, W.H., HII, Toh Onn, D.: Method and system to process a digital image. World Intellectual Property Organization Publication No. WO/2005/05983 (2005), http://www.wipo.int/pctdb/en/wo.jsp?WO=2005059830
15. USGovernment: Resources for understanding and implementing Section 508, http://www.section508.gov/
16. World wide web consortium: Accessibility, http://www.w3.org/standards/webdesign/accessibility

Camera Canvas: Image Editing Software for People with Disabilities

Christopher Kwan and Margrit Betke

Image and Video Computing Group, Department of Computer Science,
Boston University, 111 Cummington St, Boston, MA 02215, USA
{ckwan,betke}@cs.bu.edu

Abstract. We developed Camera Canvas, photo editing and picture drawing software for individuals who cannot use their hands to operate a computer mouse. Camera Canvas is designed for use with camera-based mouse-replacement interfaces that allow a user with severe motion impairments to control the mouse pointer by moving his or her head in front of a web camera. To make Camera Canvas accessible to as wide of a range of movement abilities as possible, we designed its user interface so that it can be extensively tailored to meet individual user needs. We conducted studies with users without disabilities, who used Camera Canvas with the mouse-replacement input system Camera Mouse. The studies showed that Camera Canvas is easy to understand and use, even for participants without prior experience with the Camera Mouse. An experiment with a participant with severe cerebral palsy and quadriplegia showed that he was able to use some but not all of the functionality of Camera Canvas. Ongoing work includes conducting additional user studies and improving the software based on feedback.

Keywords: Assistive Technology, Camera Mouse, Human Computer Interaction, Image Editing, Mouse Replacement System, Photo Editing, Video-based Interface.

1 Introduction

Individuals who cannot speak and cannot use their hands to operate a computer mouse are extremely limited in their means of communication. The goal of our work was to design image-editing software that these individuals can use for communication and as a canvas for expression. The process of developing Camera Canvas helped us gain general knowledge about interface design for people with severe motion impairments. We gained insights and developed techniques that may be applied to future projects for this user group.

We built upon our experience in developing software for camera-based mouse-replacement interfaces; in particular, the Camera Mouse input system [2], which enables a user to control the mouse pointer by moving his or her head in front of a camera. The Camera Mouse [2] is software that was developed at Boston College and Boston University and is freely available on the web for download at www.cameramouse.org. With this version of the Camera Mouse, only left-click events can be simulated. A left-click event is registered when the user "dwells," or

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 146–154, 2011.

keeps the mouse pointer within a small radius of the item to be selected, for a certain amount of time, e.g.: one second. We suggest that Camera Canvas can also be used with other camera-based mouse-replacement interfaces [5], [10].

There have been various other kinds of applications specially designed for use with Camera Mouse [1]. In particular, software has been developed to enable people with motor impairments to create drawings: Eagle Paint [1] is a program designed for use with the Camera Mouse that allows users to draw freeform lines, EyeDraw [7] is a drawing program designed for use with an infrared eye tracker, and VoiceDraw [6] is a drawing program that allows users to draw freeform lines by making different sounds with their voices. There has also been work to create customizable [9] and automatically generated [4] user interfaces for people with motor impairments.

Our goal was to develop a program that gives users with motor impairments photo-editing and drawing capabilities and that has a user interface that can be extensively customized to meet the needs of each user. Many interactions in image-editing involve complex mouse actions, such as clicking and dragging, which are not possible with the Camera Mouse input system. To make such interactions possible, we needed to redesign such actions completely. We also had to take into account the wide range of movement abilities of individuals with motor impairments. Some users have better control of their movements along certain axes [3], some users have good control of their movement but only within a certain range [11], and some users can only click buttons of a certain minimum size. We also considered how our software can remain usable for users whose abilities degrade over time. To meet the needs of as many users as possible, we designed our software to be highly configurable.

2 Methods

2.1 Sliding Toolbar

The main user interface element of Camera Canvas is the Sliding Toolbar. It consists of two panels: a tool menu panel containing specific image editing tools and a navigation panel containing navigation buttons. When a tool from the tool menu panel is selected, if that tool has a submenu, the buttons of that submenu will replace the current buttons in the tool menu. The user can get back to the previous menu of buttons by pressing the Back button in the navigation panel or go back to the topmost tool menu by clicking the Main Menu button in the navigation panel.

The Prev and Next buttons in the navigation panel allow the user to reposition the tool panel by sliding it sideways. This sliding ability addresses the problem of some users only having good control within a certain range. If users cannot reach a button on the edge of the screen, they can use the Prev and Next buttons to slide the buttons towards the center. The direction of movement is from the perspective of the button currently in the center position of the toolbar (above the space between the Prev and Next buttons). Pressing the Prev button will cause the button in the previous position to the center position to slide to the center position. Similarly, pressing the Next button will cause the button in the next position after the center position to slide to the center position. As long as the user keeps the mouse cursor on top of the Prev or Next button, the toolbar will continue to automatically slide on an adjustable interval.

2.2 Configuration

Camera Canvas has three configurable settings in the Settings menu: toolbar placement, button size, and toolbar sliding speed. These settings can all be changed at run-time using tools within the application. The tools are designed to be easy to use so that the user can actually modify the configuration himself. The placement and orientation of the toolbar can be changed to four settings: Horizontal-Top (Fig. 1, left), Horizontal-Bottom, Vertical-Left, and Vertical-Right (Fig. 1, right).

Fig. 1. Camera Canvas in Photo-editing mode with a Horizontal-Top layout and smaller buttons (left) and in Drawing mode with a Vertical-Right layout and larger buttons (right). The second set of Prev and Next buttons signify that there are more buttons off-screen.

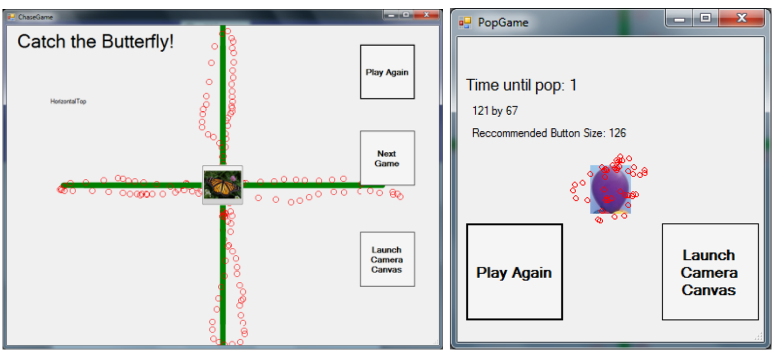


Fig. 2. The "Catch the Butterfly" game recommends which axis and area of the screen are best for the user by having her follow a butterfly (left). Green lines show ideal mouse trajectory, red circles show actual trajectory. The "Pop the Balloon" game recommends a button size for the user by having her try to keep the mouse cursor still within a small area (right). The balloon is the ideal area; red circles show the actual mouse movement area.

Each setting aims to constrain movement primarily along a single axis and in a single area of the screen to address the challenges of users having better movement abilities along different axes and users being able to reach different areas of the screen more easily than other areas. The size of all buttons in the application can be made smaller and larger to address the challenge of different people being able to click buttons of different minimum sizes. Finally, the interval at which the toolbar buttons slide can also be changed so that the buttons slide faster or slower. The Settings menu also contains a configuration wizard for Camera Canvas in the form of two simple, easy to understand games (Fig. 2). These games look at a user's performance using the Camera Mouse and recommend settings for Camera Canvas that would make it the most usable for that user.

2.3 Photo-Editing Tools

For the Photo-editing mode of Camera Canvas, we developed several interaction techniques to make common photo-editing tasks possible with camera-based interfaces. The Rotate tool uses a custom user interface component called a Choice Form (Fig. 3, left) that is an alternative to components such as sliders or small increment arrows, which are challenging for users who have difficulties controlling the mouse pointer. The middle of the Choice Form contains a preview of the rotated image so that the user can see the effects of the rotation before actually committing the change. The Choice Form is also used by many other tools in Camera Canvas.

The Move and Zoom tools place four translucent arrows in the middle of the screen. To pan around the image, the user puts the mouse pointer over one of the arrows and the image automatically moves until the user moves the mouse pointer off of the arrow. No matter the size of the image, the user only needs to make small movements between the arrows to pan, instead of having to physically move the mouse pointer around the entire image.

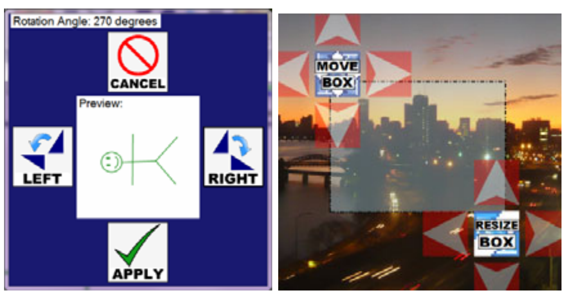


Fig. 3. Rotating a drawing using a Choice Box (left). Selecting a portion of an image (right).

Instead of the traditional click and drag method of selecting a portion of an image, the Select tool uses two sets of arrows similar to the ones used in the Move and Zoom tools. When using Select, a translucent blue rectangle (representing the selection) and two sets of arrows appear in the center of the image. The set of arrows on the left control the position of the top-left corner of the selection box and the set of arrows on the right control the position of the bottom-right corner of the selection box. By moving the mouse cursor in each of the arrows, the user can control the position and size of the selection box. Once the selection box is of the desired position and size, the user can then cut, copy, paste, or crop the selection. The two sets of arrows never change positions so no matter the size of the selection, the user can control it using only small movements between the two sets of arrows.

2.4 Drawing Tools

The Camera Canvas interaction for drawing straight lines and geometric shapes was inspired by the drawing process in EyeDraw [7]. To address the "Midas touch" problem for drawing (how to differentiate looking at the picture versus actually drawing the picture) the researchers of EyeDraw created a system where if the user looked at one spot for some amount of time, the cursor would change colors to signify that drawing was about to begin; if the user was just looking and did not want to actually start drawing they would just need to look elsewhere.

In Camera Canvas, to start drawing, the user must first dwell on the area where she would like to place the starting point of her drawing. After a click is registered, a green helper box appears where she clicked to signal that drawing is about to begin. If the user would actually like to start drawing at that point, she keeps the mouse cursor in the green Helper Box long enough for another click to register and then drawing begins. If the user does not want to place the starting point at that location, she only needs to move the mouse cursor out of the green Helper Box and it disappears, resetting the process. As the user is drawing, the line or shape is continuously redrawn with the ending point at the current position of the cursor. When the user wants to end the drawing, she dwells where she would like to end the drawing and a red Helper Box appears. If she would in fact like to place the end point of the drawing at that point, she just needs to keep the cursor inside of the red Helper Box. If she does not want to place the end point there and instead wants to continue drawing, she just needs to move the cursor out of the red Helper Box and it disappears. The sizes of the Helper Boxes are the same size as the toolbar buttons and will change if the button size is changed. The drawing process is outlined in (Fig. 4, left).

Instead of using a traditional color palette which relies on sliders or clicking of a precise point in a color wheel, we implemented a simple color palette that is much more usable with Camera Mouse but still gives users a fair amount of color variety. The color menu (Fig. 4, right) first displays a set of primary colors: black, white, brown, red, orange, yellow, green, blue, and violet. When the user clicks on a primary color, nine different shades of that color are then automatically generated for the user to choose from.

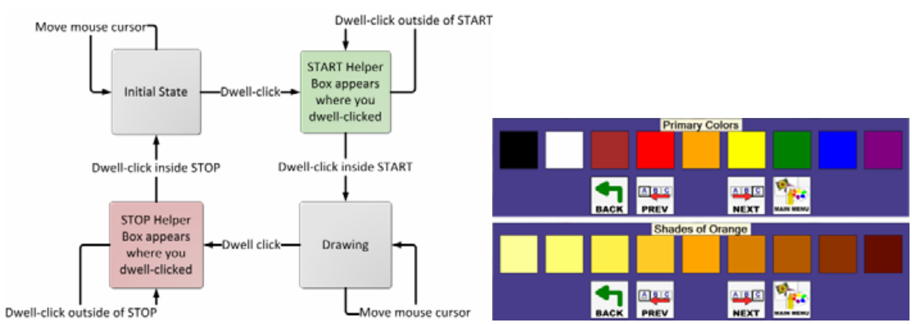


Fig. 4. The drawing process in Camera Canvas (left). The Camera Canvas color palette generating different shades of orange (right).

3 Experiments and Results

We conducted several user studies with a total of 28 users without disabilities. Their age ranges included elementary school, middle school, high school, and college age ranges. All users had never used the Camera Mouse input system before. The goal of these studies was to obtain a qualitative assessment of the program to see if it was easy to understand and use.

We asked the users to use various Photo-editing tools to manipulate a photo, use various Drawing tools to draw a shape, and then to play around with changing different configuration settings. There was no strict test plan; users were given freedom to explore the different features of the program as we observed them.

In general, the participants of our experiments found the software easy to understand and use even without prior experience using the Camera Mouse. With a little experimentation time, users were quickly able to start drawing shapes and manipulating images. We found that nearly all users enjoyed the drawing tools the most and spent most of their time with the program drawing (Fig. 5). The users gave us valuable feedback on which features needed improvement and also what features they wanted to see in future versions, common suggestions were a fill tool and clip art stamps.

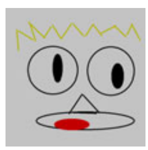


Fig. 5. A drawing created by a user without disabilities

We recently conducted experiments with a non-verbal user with severe cerebral palsy and quadriplegia (Fig. 6). This user had participated in experiments with the initial version of Camera Canvas [8]. In the prior experiments, the user was excited about the prospect of manipulating images but was unable to use the majority of the features [1]. In our test procedure, we first explained the general purpose of the program to him. We explained how the toolbar worked and that we could change its position, button size, and sliding speed if necessary. From an earlier experiment with different software during that session, we found that this user had a difficult time reaching buttons at the top of the screen. We therefore changed the Camera Canvas toolbar to have the Horizontal-Bottom layout and made the button size larger.

In our experiments, we asked the participant to try out the different modes of the program. Originally we had a detailed test procedure planned for him but the session turned into more of a qualitative exploratory session where we observed which features he was able to actually use.

After we explained the Sliding Toolbar to the user, he understood how the interface worked and was able to use the Prev and Next buttons to slide the toolbar buttons he wanted towards the middle of the screen. The Prev and Next buttons had mixed results. Although the user could slide to buttons that he wanted to reach, often times he would slide the toolbar too much and overshoot the button he wanted or would accidentally activate the Prev or Next button when trying to click on a button in the tool menu, causing his intended target to shift. To address this problem, we tried to slow down the sliding speed setting, but the user still hit the Prev and Next buttons by accident because of their proximity to the tool menu buttons.

In general, the user kept accidentally clicking neighboring buttons to his intended button because the buttons were too close together. This suggests that the buttons should be spaced farther apart or that this setting should also be adjustable. A particularly frustrating experience for the user was accidentally hitting the Main Menu button when he was in the middle of trying to apply an effect to the image. Accidentally hitting the button would take the user all the way to the Main Menu of the program and then he would have to click on Photos, slide down to Effects, and then click on the effect again. This happened multiple times and eventually we took control of the mouse in order to get the user back to the Effects menu again.

The observations that the user had to keep sliding to reach buttons near the edge of the screen and that he kept hitting buttons accidentally, suggest that the user might benefit from the toolbar having fewer buttons. A greater amount of buttons on the toolbar increases the chance for error and also may be cognitively overwhelming for someone using the program for the first time. Perhaps a more hierarchical approach (more levels with fewer buttons at each level) would be more usable for this user.

In general, even with larger buttons, it was difficult for this user to stay on top of one button long enough for the click to register. Shortening the time required for a click in the Camera Mouse settings helped reduce the problem slightly but in general it still persisted. In the future, we may experiment with adjusting the dwell radius setting in Camera Mouse to more closely match the button size in Camera Canvas. It is possible that even though the user placed the pointer on top of the button, he may not have kept it within a small enough radius for Camera Mouse to register a click.

The ability to configure the user interface of Camera Canvas was very important in experiments with this user. We used all three of the configuration options: toolbar placement, button size, and sliding speed to try to make the most usable layout for the user. The user also played the configuration games. He was able to understand and complete both the butterfly (toolbar layout configuration) and the balloon (button size configuration) games, although the layouts of the buttons in both games could be improved or ideally made configurable. The games recommended a Vertical-Right layout with buttons of size 160 by 160 pixels for the user. The user was satisfied with these settings and chose to keep them for the remainder of the experiment. Although the user liked these settings, we do not know if there were settings that could have made the program even easier for him to use because we stopped trying different settings after the user indicated he was satisfied.

In the Photo-editing tools, the user was able to successfully use the Move and Zoom features to zoom the image to a greater magnification and then pan the image so that a particular portion was centered on the screen. Even though the arrows of the Move feature were a fixed size, the user was still able to activate them because they were activated whenever the mouse entered the region rather than forcing the user to stay in the region for an amount of time, as is the case with buttons. This suggests that a boundary crossing or mouse touch approach instead of buttons might be more usable for this user. The user was also able to apply the Invert Colors feature to the image and then undo the change.

In the Drawing mode, the user was able to select different shapes and then draw rectangles around the image [Fig. 6, left]. We are not sure if the user intended to draw something specific or was just experimenting with the tool, as this may have been the first time that this user had interacted with a drawing interface.

We learned a great deal from our experiments with this user. We were able to see problems with the interface that did not arise when testing the software with users without disabilities. While there were many features that the user had trouble with, could not use, or did not try, in general the experiment was a major improvement over the experiment with the initial version of the software with this user.

Fig. 6. An image edited by a user with severe cerebral palsy. He was able to rotate the image (presented to him upside-down) and experiment with drawing several shapes on the image (left). A user with severe cerebral palsy interacting with Camera Canvas using the Camera Mouse (right).

4 Ongoing Work

Our ongoing work involves conducting additional user studies with people with motor impairments. We are also working on improving the program based on observations from our experiments with the user with severe cerebral palsy. We hope to make modifications to the software to make it more usable for this user and conduct additional experiments with him. We are also adding additional features to the program, such as the much requested fill and clip art stamps and continuing to look into using simple games to recommend user interface settings.

Acknowledgements. We would like to thank the participants of our user studies for their valuable feedback. We would also like to thank Mikhail Breslav, Samuel Epstein, Nathan Fuller, John J. Magee, Eric Missimer, Isaac Paquette, Ashwin Thangali, Diane Theriault, Gordon Towne and Zheng Wu for their support throughout the development of this program and their assistance during user studies. We gratefully acknowledge funding from the National Science Foundation (HCC grants IIS-0910908, IIS-0713229, and IIS-0855065) and from the Boston University Undergraduate Research Opportunities Program.

References

1. Betke, M.: Intelligent interfaces to empower people with disabilities. In: Nakashima, H., Augusto, J.C., Aghajan, H. (eds.) Handbook of Ambient Intelligence and Smart Environments, pp. 409–432. Springer, Heidelberg (2009)
2. Betke, M., Gips, J., Fleming, P.: The Camera Mouse: visual tracking of body features to provide computer access for people with severe disabilities. IEEE Transactions on Neural Systems and Rehabilitation Engineering 10(1), 1–10 (2002)
3. Connor, C., Yu, E., Magee, J., Cansizoglu, E., Epstein, S., Betke, M.: Movement and recovery analysis of a mouse-replacement interface for users with severe disabilities. In: 13th International Conference on Human-Computer Interaction (HCI International 2009), San Diego, CA (2009)
4. Gajos, K.Z., Weld, D.S., Wobbrock, J.O.: Automatically generating personalized user interfaces with Supple. Artif. Intell. 174(12-13), 910–950 (2010)
5. Gorodnichy, D., Dubrofsky, E., Ali, M.: Working with computer hands-free using Nouse perceptual vision interface. In: Proceedings of the International CRV Workshop on Video Processing and Recognition (VideoRec 2007). NRC, Montreal (2007)
6. Harada, S., Wobbrock, J.O., Landay, J.A.: VoiceDraw: a hands-free voice-driven drawing application for people with motor impairments. In: Proceedings of the 9th International ACM SIGACCESS Conference on Computers and Accessibility, pp. 27–34. ACM, Tempe (2007)
7. Hornof, A.J., Cavender, A.: EyeDraw: Enabling children with severe motor impairments to draw with their eyes. In: Proceedings of ACM CHI 2005: Conference on Human Factors in Computing Systems, pp. 161–170. ACM, New York (2005)
8. Kim, W.-B., Kwan, C., Fedyuk, I., Betke, M.: Camera Canvas: image editor for people with severe disabilities. Department of Computer Science Technical Report BUCS-2008-010, Boston University (2008)
9. Magee, J., Betke, M.: HAIL: Hierarchical adaptive interface layout. In: Miesenberger, K., Klaus, J., Zagler, W., Karshmer, A. (eds.) ICCHP 2010. LNCS, vol. 6179, pp. 139–146. Springer, Heidelberg (2010)
10. Manresa-Yee, C., Varona, J., Perales, F.J., Negre, F., Muntaner, J.J.: Experiences using a hands-free interface. In: Proceedings of the 10th International ACM SIGACCESS Conference on Computers and Accessibility, pp. 261–262. ACM, New York (2008)
11. Paquette, M., Betke, M., Magee, J.: IWeb Explorer: A web browser designed for use with an eye controlled mouse device. Computer Science MA Thesis Report, Boston University (2005)

Exact Compensation of Color-Weakness with Discrimination Threshold Matching

Rika Mochizuki[1,2], Satoshi Oshima[2], Reiner Lenz[3], and Jinhui Chao[2]

[1] NTT Cyber Solutions Laboratories, 1-1 Hikarinooka, Yokosuka-shi, Kanagawa, 239-0847 Japan
[2] Dept. of Science and Engineering, Chuo University, 1-13-27 Kasuga, Bunkyo-ku, Tokyo, 112-8551 Japan
[3] Dept. Science and Engineering, Linköping University, SE-60174 Norrköping, Sweden

Abstract. In this paper we describe a novel compensation algorithm for color-weakness based on a new, objective criterion to compare normal observers and color-weak observers, using Riemann geometric properties of color spaces. The criterion is to match the color discrimination thresholds of average, normal observers and a colorweak observer. The method uses local and global isometry theory and provides the two groups of observers with the same color-difference experience. A one-dimensional compensation and simulation of color-weakness is shown as an application of the general approach to the Brettel color-blind model. The 2D and 3D compensations and simulations are illustrated in chromaticity planes and full color spaces.

1 Introduction

Today visual media has become an indispensable part of communication, social activity and daily life. With the rapid development of IT technology, most digital contents consists of large volume datasets, containing visual information. On the other hand, it is known there are more than 5 to 8 percent of humans suffer visual impediments and especially color vision defects. Adaptation of the visual information and the compensation of the color vision of color-weak observers is one of the most important and challenging tasks in universal design and barrier-free IT technology. Currently used methods either enhance color contrast between symbols and background and thus they do not apply to natural images, or they need to define certain criteria taken into some often conflicting visual restrictions to find compensation colors by optimization. The obtained solutions are not unique and often difficult to justify.

One of the difficulties in the compensation of color-weak vision comes from the problems in the observation or the measurement of human vision, resulting in a lack of an objective criterion for compensation. Moreover the degree and the properties of color-weakness vary widely among individuals and different color stimulations, which demands compensation taking into account these personal variations.

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 155–164, 2011.

In this paper, we propose a new criterion for color-weakness compensation based on the matching between discrimination thresholds of color-weak and color-normal observers. An exact compensation is then possible in the sense that subjective color differences between every pair of colors in an image are the same between both color-weak and normal observers.

Applying this theory to Brettel's color-blind model, we show a color-weak map for simulation and compensation of color-weakness. The approach is then extended to high dimensional color spaces.

2 The Riemann Geometry of Color Spaces

Color vision is subjective varying between different observers and one of the few observable data to characterize color perception are color differences. The most accessible and reliable measurement are small or local color differences. In particular, the so-called just noticeable-difference (jnd) thresholds or the discrimination thresholds are the minimal color-difference a human observer can differentiate. These threshold data are measured at different points in color space and provide a measure of local distance in the color space. For a test color x and a color vector close to x say y, the discrimination threshold is an ellipse/ellipsoid centered at x, defined by the following equation:

$$y^T G(x) y = 1 \tag{1}$$

Here the positive definite matrix $G(x)$ defines the local geometry depending on the test color x. The local distance around x can be expressed as

$$\| dx \|^2 = dx^T G(x) dx. \tag{2}$$

Such a space, with a smoothly defined local distance or the matrix $G(x)$ (called as Riemann metric) is called a Riemann space[5]. The fact that a color space is a Riemann space rather than a Euclidean space is known from Helmholtz [4].

Another yet even more important quantity in color perception are large color differences, which however is more subjective and hence harder to deal with. In fact, the color difference between two points in a color space can also be expressed as distance between the two points. In a Riemann space, the distance between points x and x' is defined as the length of the shortest curve (geodesic) connecting the two points.

In order to present the same color stimuli to a color-weak observer as perceived by color-normal observers, the most natural way is to transform the color space of the color-weak observer so that it has the same geometry, and therefore the same color differences between every pair of colors as in the color space of color-normal observers. Such a color-difference preserving mapping is called an isometry in Riemann geometry.

It can be shown that two locally isometric spaces are also globally isometric. This means that the distance between any pair of points in one space is equal to the distance between the corresponding pair of points in the other space. In other words, under a map between two color spaces, matching of discrimination threshold will guarantee matching of large color differences, in the form of geodesics, between the two spaces[6].

3 Compensation by Discrimination Threshold Matching

3.1 Measurement and Estimation of Thresholds

First we measured the discrimination thresholds of both, color-normal and color-weak observers, The observer sees a square of size $14cm \times 14cm$ from a distance of 80cm by using a 10 degrees field of vision. They were obtained from 45 college students (37 male, 8 female including 2 color-weak persons) in CIEXYZ coordinates. The discrimination thresholds in the uv plane, of a protanopic color-weak observer and an average of color-normal observers (the inner ellipses) are shown in Fig.1.

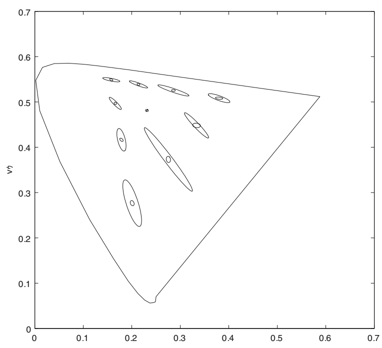


Fig. 1. Ellipsis of the color-weak vs normal observers on $u'v'$ plane

3.2 New Compensation Criterion and Algorithms

We define the color difference preserving map which transforms C_w of the color-weak observer to C_n of color-normal observers, the "color-weak" map, as follows:

$$w : C_n \longrightarrow C_w, \qquad x \longmapsto y = w(x) \tag{3}$$

The color-weak map for an observer can be determined by threshold matching between the color-normal and the color-weak. In particular, assume that a color stimulus x perceived by color-normals is mapped by w to $y = w(x)$ perceived by a color-weak. Using discrimination thresholds $G_n(x)$ of the color-normal and the corresponding thresholds $G_w(y)$ for color-weak, the Jacobian matrix D_w of w such that $dy = D_w dx$ can be obtained by the threshold matching condition:

$$G_w(y) = (D_w)^T G_n(x) D_w \tag{4}$$

The compensation of the color-weakness is given by inverse map of the color-weak map w. On the other hand, applying w to the input image and showing it to color-normal observers will provide them with the same experience, ie. the simulation of the color-weak observer.

4 Colorweak Model and Compensation on Confusion Lines

4.1 A Linear Color-Weak Model and Compensation

Human color perception depends on the existence of the three L,M, S cones on the retina. It is known that a model of color-blind vision can be represented by a projection in LMS space along the L axis for protanopia and the M axis for deuteranopia. Since color-blind or weak observers cannot discriminate between colors on the L axis for protanopia and the M axis for deutanopia, the straight lines parallel to the L or M axis are called the confusion lines. The color blind model by Brettel et al [1] determines the color stimuli perceived by color blind observers, which consists of the two planes in LMS space. The two planes are spanned by three invariant hues which are perceived equally by both, color-normal and color-blind observers.

We first consider a linear color-weak model and compensation based on Brettel's model. (Fig.2). We denote $\boldsymbol{Q}$, $\boldsymbol{Q'}$ as perception of color-normals and color-blinds, $\boldsymbol{Q'}$ is projection onto the color-blind plans along the L axis for protanopia and the M axis for deutanopia, which are in correspondence to the confusion lines. For every 3D color vector $\boldsymbol{R}$, only 1D direction along the confusion line passing through $\boldsymbol{R}$ is of our interest, thus we will use the scalar R to represents $\boldsymbol{R}$ as its position on the confusion line.

In our linear color-weak model, every stimulus Q is mapped by a color-weak map w towards the color blind plane however without reaching it. Assuming that the color weak map w is a linear map, then Q'' as the perception of color-weak observer can be described as the image of Q under w :

$$Q'' = w(Q) = \omega Q' + (1-\omega)Q \qquad (0 \le \omega < 1) \tag{5}$$

Here ω is called the color-weak index which indicates the degree of color-weakness. The observer is completely color-blind if $\omega = 1$ and color-normal if $\omega = 0$. As shown in Fig.4, Q' is moved the reduced distance $(Q' - Q)$ by a factor of $1 - \omega$.

$$Q'' = w(Q) = Q' + (1-\omega)(Q - Q') \tag{6}$$

The color map w simulates color-weak vision when applied to the original image, the compensation is obtained from the inverse map of w:

$$P = w^{-1}(Q) = Q' + \frac{1}{(1-\omega)}(Q - Q') \tag{7}$$

In fact, substituting the above color P into (5) one can confirm that the color-weak observer actually perceives the same color as the color-normals. Unfortunately the color-weak index ω contained in the above color weak model and color-weak map is not directly measurable. However, it can be determined by using the threshold matching criterion. First, the Jacobian of color-weak map w is

$$D_w = 1 - \omega. \tag{8}$$

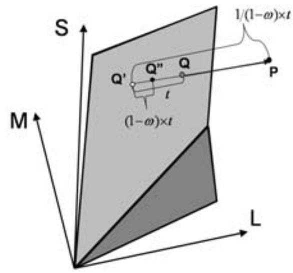


Fig. 2. Linear color-weak map along confusion lines

We denote α_n, α_ω as the thresholds of a color-normal and a color-weak. Thus the criterion (4) now becomes as described in Eq.(9) in the color-weak map.

$$\alpha_n = (1 - \omega)\alpha_w \tag{9}$$

Thus we can define ω as follows so that we obtain the color-weak map w.

$$\omega := 1 - \frac{\alpha_n}{\alpha_w} \tag{10}$$

4.2 A Rigorous Color-Weak Model and Exact Compensation

Figure 3 shows the thresholds of normal and color-weak observers along confusion lines. Therefore color-weakness has large variations not only among individuals but for different color stimuli as well. Since the assumption of linearity of the color-weak map is equivalent to that the color-weak index ω is a constant on a confusion line, the color-weak map is therefore in generally not a linear map. Below, we show a rigorous model of color-weak simulation and compensation (Figure 4).

From the threshold matching criterion, the Jacobian of the color-weak map needs to satisfy the following condition, where $\omega(Q)$ is the color-weakness of Q. Now the condition (4) becomes

$$D_w(Q) = \frac{\alpha_n(Q'')}{\alpha_w(Q)} =: 1 - \omega(Q) \tag{11}$$

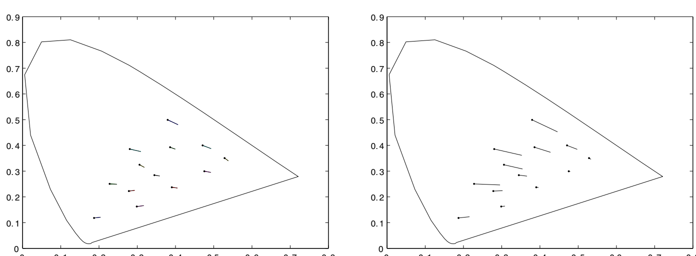


Fig. 3. Thresholds of normal and color-weak observers along confusion lines

The exact color-weak map is then the integral of Jacobian in C_w which generalizes (6):

$$Q'' = w(Q) = Q' + \int_{Q'}^{Q} (1 - \omega(x))dx. \tag{12}$$

The compensation map is the integral in C_n which generalizes (7):

$$P = w^{-1}(Q) = Q' + \int_{Q'}^{Q} \frac{1}{1 - \omega(y)} dy. \tag{13}$$

In this way, an exact correspondence between two color spaces of the color-weak observer and color-normals is obtained from the color-weak map (12) and its inverse map (13).

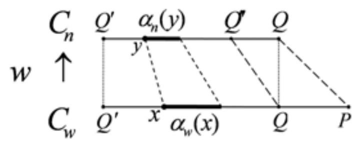


Fig. 4. Rigorous color-weak maps along confusion lines

To implement the color-weak map and the compensation map from the observed discrimination threshold data, choose the sampling points $\{x_0, x_1, x_2, ...\}$ on the confusion line in C_w:

$$x_0 = Q', \qquad x_{i+1} = x_i + \alpha_w^{(i)}$$

Here, $\alpha_w^{(i)} := \alpha_w(x_i), i = 1, 2, ...$ are the discrimination thresholds at sampled points $\{x_i\}$.

Similarly, the sampling points $\{y_0, y_1, y_2, ...\}$ on the confusion line of C_n are defined using the discrimination threshold data $\alpha_n^{(j)} := \alpha_n(y_j), j = 1, 2, ...$:

$$y_0 = Q', \qquad y_{j+1} = y_j + \alpha_n^{(j)}$$

Then under the color-weak map, one has a precise correspondence between x_i and y_j.

$$\begin{aligned} w(x_k) = Q' + \int_{Q'}^{x_k} (1 - \omega(x))dx &= x_0 + \sum_{i=0}^{k-1} \int_{x_i}^{x_{i+1}} (1 - \omega(x))dx \\ = y_0 + \sum_{i=0}^{k-1} \alpha_n^{(i)} = y_k \qquad & k = 0, 1, 2, ... \end{aligned}$$

The color weak index at the k-th interval ω_k is defined by using the discrimination threshold of color-normals $\alpha_n^{(k)}$ and of the color-weak observer $\alpha_w^{(k)}$.

$$\omega_k := 1 - \frac{\alpha_n^{(k)}}{\alpha_w^{(k)}} \tag{14}$$

Here $x_I \leq Q < x_{I+1}$ in C_w and $y_J \leq Q < y_{J+1}$ in C_n.

Assuming $\alpha_w(x), x \in [x_{k-1}, x_k]$ in the k-th interval of C_w is a constant equal to that on the right end of the interval $\alpha_w^{(k)}$ and the discrimination threshold in k-th interval in C_n is a constant equal to $\alpha_n^{(k)}$, then the color-weak map and the compensation map can be realized by a sum of the discrimination thresholds on the confusion line:

$$Q'' = \sum_{i=0}^{I}(1-\omega_i)(x_{i+1}-x_i) = \sum_{i=0}^{I}\alpha_n^{(i)} \tag{15}$$

$$P = \sum_{j=0}^{J}\frac{1}{1-\omega_j}(y_{j+1}-y_j) = \sum_{j=0}^{J}\alpha_w^{(j)} \tag{16}$$

5 High Dimensional Extensions

As being observed from experimental data, the discrimination threshold ellipsoids do not simply expand along the confusion lines. In fact, it is reported in [7] that severe color-weak observers have discrimination threshold ellipsoids that expand along the confusion lines. On the other hand, for the color-weak observers of mild severity, their discrimination threshold ellipsoids expand in a random way. Considering the main population of color-weak observers belongs to the latter group, this implies the necessity of simultaneous compensation in more than one direction or a spatial compensation.

Below, we show a high dimensional extension of the color weak map (Figure 5). Assume we have a set of sampling points in 3D C_w: $\{x_i = (x_1^i, x_2^i, x_3^i)^T, i = 1, 2, ...,$ corresponding to the set of the images of the sampling points into C_n: $y = (x_2, y_2, z_2)^T = w(x) \in C_n$, $\{y_i = (y_1^i, y_2^i, y_3^i)^T\}, i = 1, ..., N$.

The colorweak map $w : C_w \longmapsto C_n$ can be linearly approximated at the neighborhoods of each sampling point onto the neighborhoods of their images, by the Jacobian matrix $D_w^{(k)} = D_w(x_k)$. which defines the local affine map between the neighborhood of x_k and the neighborhood of its image $y_k = w(x_k)$

$$y - y_k = D_w^{(k)}(x - x_k) \tag{17}$$

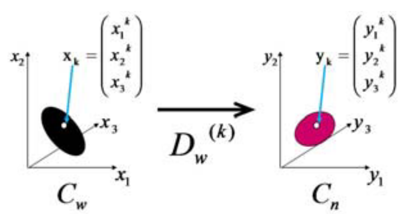


Fig. 5. Local color weak and compensation map

The Jacobian matrix $D_w^{(k)}$ of w is determined again by the local isometry or threshold matching condition (4):

$$R_n^{(k)} = (D_w^{(k)})^T R_w^{(k)} D_w^{(k)} \tag{18}$$

6 Simulation and Compensation

Figure 6 shows the simulation of color-weak vision, the original and the compensated image by the 1D color-weak map. Figure 7 shows the simulation of color-weak vision, the original and compensated image by the 2D color-weak map on the chromaticity plane. Here the color-weak map and compensation is described locally by 2×2 matrices and estimated from matching of threshold ellipses using (18). The 1D compensation and color-weak simulation images

Fig. 6. 1D color-weak simulation, original image and 1D compensation

Fig. 7. 3D color-weak simulation, original image and 3D compensation

are evaluated using the semantic differentiation (SD) test,a standard culture-independent procedure to quantitatively evaluate subjective impressions. First a selection of opponent-pairs of adjectives related to the test images were chosen by an individual group. Then the test images are scored by another group and the color-weak observer using a 7 points scale for each pair of adjectives. In particular, two sets of SD tests are averaged and compared with each other in Figure 8: The left between the original image evaluated by color-normals and for the compensated image by the color-weak, the right between the "color-weak simulation" of the original image evaluated by the color-normals and the original image evaluated by the color-weak. The comparisons show that the color-normals obtained similar impressions from the original image as the color-weak from the compensated image, and that the color-weak's impression on the original image is close to that of the color-normals from the "color-weak simulation" of the original image.

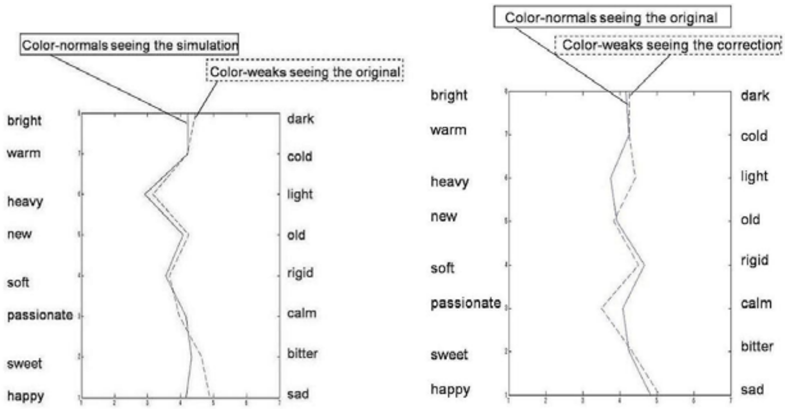


Fig. 8. SD evaluation of 1D compenation and color-weak simulation

7 Summary and Conclusions

We presented a novel color-weak compensation approach based on an objective criterion using the Riemann geometry of color spaces. The 1D compensation along confusion lines and its high dimensional extension are then proposed. These methods provide tools to simulate the perception of a given image by a color weak observer and to construct a color mapping that compensates the distortion of the color perception of a color weak observer in comparison to the perception of a color normal. Efficient measurement of discrimination thresholds data and fast implementation are among future research subjects.

References

1. Brettel, H., Vienot, F., Mollon, J.D.: Computerized simulation of color appearance for dichromats. Journal of Optical Society of America 14(10), 2647–2655 (1997)
2. Ichikawa, M., Tanaka, K., Kondo, S., Hiroshima, K., Ichikawa, K., Tanabe, S., Fukami, K.: Preliminary Study on Color Modification for Still Images to Realize Barrier-Free Color Vision. In: Proc. IEEE International Conference on Systems, Man and Cybernetics SMC 2004 (2004)
3. Troiano, L., Birtolo, C., Italiane, P.: Adapting Palettes to Color Vision Deficiencies by Genetic Algorithm. In: Proc. 10th Genetic And Evolutionary Computation Conference, in CD-ROM (2008)

4. Wyszecki, G., Stiles, W.S.: Color Science, 2nd edn. Wiley Classics Library, Chichester (2000)
5. do Carmo, M.P.: Riemannian Geometry. Birkhauser, Basel (1992)
6. Chao, J., Osugi, I., Suzuki, M.: On definitions and construction of uniform color space. In: Proceedings of CGIV 2004, The Second European Conference on Colour in Graphics, Imaging and Vision, April 5-8, pp. 55–60 (2004)
7. Regan, B.C., Reffin, J.P., Mollon, J.D.: Luminance noise and the rapid determination of discrimination ellipses in colour deficiency. Vision Research 34(10), 1279–1299 (1994)
8. Mochizuki, R., Nakamura, T., Chao, J., Lenz, R.: Color-weak correction by discrimination threshold matching. In: Proceedings of CGIV 2008, 4th European Conference on Color in Graphics, Imaging, and Vision, June 9-13, pp. 208–213 (2008)
9. Chao, J., Lenz, R., Matsumoto, D., Nakamura, T.: Riemann geometry for color characterization and mapping. In: Proceedings of CGIV2008, 4th European Conference on Color in Graphics, Imaging, and Vision, June 9-13, pp. 277–282 (2008)
10. Ohshima, S., Mochizuki, R., Chao, J., Lenz, R.: Color-reproduction using Riemann normal coordinates. In: Trémeau, A., Schettini, R., Tominaga, S. (eds.) CCIW 2009. LNCS, vol. 5646, pp. 140–149. Springer, Heidelberg (2009)

Collaborative Editing for All: The Google Docs Example

Giulio Mori[1], Maria Claudia Buzzi[1], Marina Buzzi[1],
Barbara Leporini[2], and Victor M. R. Penichet[3]

[1] CNR-IIT, via Moruzzi 1, 56124 Pisa, Italy
{Giulio.Mori,Claudia.Buzzi,Marina.Buzzi}@iit.cnr.it
[2] CNR-ISTI, via Moruzzi 1, 56124 Pisa, Italy
Barbara.Leporini@isti.cnr.it
[3] Computer Systems Department, University of Castilla-La Mancha, Albacete, Spain
victor.penichet@uclm.es

Abstract. Collaborative software tools allow people to share documents and knowledge via Internet, in a simple, economic and efficient way. Unfortunately collaborative software often relies heavily on visual features and dynamic technologies with user interfaces that are difficult to use via screen reader, or are sometimes even inaccessible for the blind.

In this paper we illustrate and discuss results of an accessibility inspection of the main collaborative functions of Google Docs using the JAWS screen reader. Results highlight several difficulties encountered when interacting with elements of the Google Docs interfaces. Content perception is often incomplete, since many elements or changes occurring in the collaborative environment are not intercepted by the screen reader and announced to the user. In addition, the behavior of the collaborative functions analyzed (as well as the rendering) changes from one web browser to another. Some general guidelines are discussed, for designing user interfaces of collaborative editors that are more usable when interacting via screen reader.

Keywords: groupware, collaborative editing, accessibility, blind, screen reader.

1 Introduction

Collaborative editing tools allow people to work together via Internet in a simple, economic and efficient way. However, their user interfaces are not always easy to use, nor do they consider the needs of differently-abled persons, thus excluding a considerable number of potential users.

To create a valuable product, both accessibility and usability must be considered in the design phase. Accessibility is a prerequisite that permits users to perceive online content and interact, while usability enhances the quality of the interaction, which should be simple, efficient and satisfying. Studying principles of web design to allow easy Internet use for each type of disability is difficult, since disabilities are numerous and heterogeneous. Nevertheless, many user studies suggest that totally blind users encounter more difficulty than people with other sensory disabilities (such as low vision, or motor or hearing impairment) when executing specific tasks [1], [2], [3], so we focus our research on them.

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 165 – 174, 2011.

Blind people usually interact with computers via screen reader, voice synthesizer and keyboard. The sequential interaction may lead to serious problems, such as content serialization and information overload. The screen reader adds a level of complexity to the interaction, when designing for blind users it is essential to consider the overall interaction, involving the perceptual, motor and cognitive systems of the Human Processor Model [4].

In this paper we specifically analyze several important collaborative features when interacting via screen reader. Unfortunately, collaborative tools mostly use visual techniques to provide information (e.g., who is working simultaneously on the system, what parts are being modified, etc.). In this context, we take into account the main features of collaborative tools in order to understand how they could be made more usable for blind users. Specifically, we consider the collaborative features available on an editing tool by analyzing the popular Google Docs (http://docs.google.com) as a collaborative editing system. At the beginning of our study we analyzed the main accessibility problems of the Google Docs environment when a blind user interacts via screen reader and voice synthesizer [5]. In this paper we specifically discuss accessibility and usability issues for collaborative functions available in Google Docs. Typical collaborative features of groupware environments include collaboration, cooperation, coordination, communication, information sharing, awareness, time and space.

To verify the accessibility of typical groupware, we inspected specific functions of Google Docs involving interactions with other persons [6]: inviting people, getting a link to share, seeing who has access, email with attachment, sending messages to collaborators, publishing a web page, uploading files, reviewing history and comments. The test aimed to evaluate the degree of accessibility of collaborative features of Google Docs when interacting via screen reader (we used JAWS, the most commonly used screen reader in the Italian blind community [7]), and was performed by the three sighted authors of these paper (with computer screen turned off), and by the blind author of this study, who is proficient in using JAWS.

Our results offer an overview of several difficulties encountered when interacting with elements of the Google Docs interfaces. Content perception is often incomplete, since many elements or changes occurring in the collaborative environment are not announced by the screen reader. As an additional problem, the functioning (and rendering) of the collaborative functionalities analyzed could differ depending on the web browser used. Finally, we will suggest some basic guidelines for designing more effective, efficient and satisfactory UIs for collaborative editing.

2 Accessibility and Usability of Google Docs Collaborative Features

In this section we only analyze specific functions of Google Docs that involve collaborative interactions [6]. The degree of accessibility of these functions have been verified using the screen reader JAWS (versions 10 and 12), and the Internet Explorer 8 and Firefox 3.6.13 browsers. Interaction with these two browsers has shown significantly different behaviors that may negatively impact on the user. In the following the issues encountered for each main collaborative function will be described.

2.1 Invite People

Considering the main interface of Google Documents shown in Fig. 1, it is possible to share a document with other people in three ways:

- Accessing the "share" pull-down menu of the main interface (Fig. 1);
- Selecting the checkbox associated to a document and accessing the "share" item menu, which appears after pressing the simulating *"right mouse button"* key on the keyboard (Figure 1);
- Opening a document and accessing the "share" pull-down menu from the editor (Fig. 2).

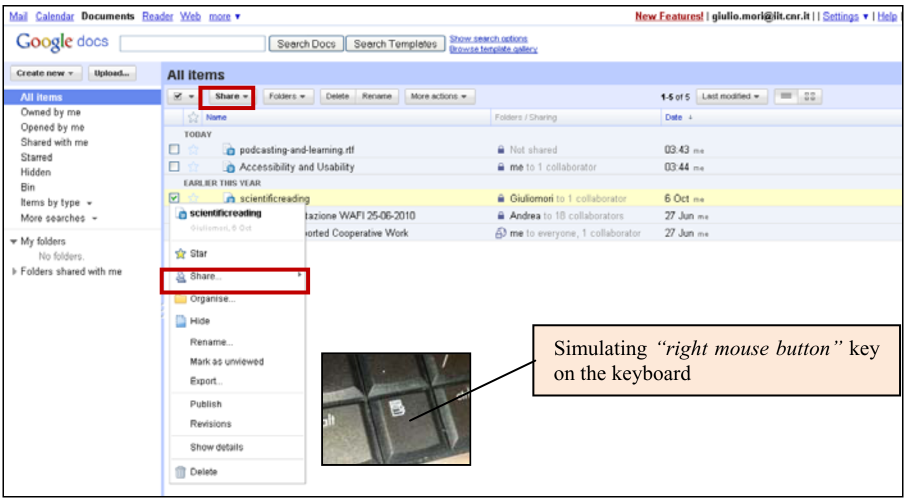


Fig. 1. Main interface of Google Docs – Documents

Verifying the accessibility of these three modalities for inviting people we observed:

1) The *"share"* pull-down menu is not accessible on both Internet Explorer and Firefox. 2) The checkbox which is associated with a document is only selectable on Firefox. The checkbox is not selectable using Internet Explorer, so it is not possible to access the menu functions related to a selected document using that browser. In terms of usability, the association between a checkbox and a document is not intuitively perceived by a user interacting via screen reader and keyboard.

Besides, 3) the "share" pull-down menu on the editor interface is not very accessible (after pressing the ESC key in the editing area), because the pull-down menu is announced as a general "button menu". Interaction improves a little using JAWS v. 12. Selecting the "share" pull down or item menu it would open the "sharing settings" window (Fig. 3a) where it is possible to insert collaborators' addresses, but this only happens using Firefox.

This page also presents some accessibility problems: a) many labels of the elements (for example “Add people”, etc.) are not announced by the screen reader, so the blind user cannot know what information to insert in that field; b) sometimes we detected the loss of the window focus during interaction when the “Sharing Settings” window is opened; c) in case no people are added, the “Share” button is correctly non-clickable, but after closing the “sharing settings” window (using the “Cancel” button), a warning message (Fig. 3b) appears and it is not announced by the screen reader; d) after closing the window for sharing a document, focus goes to the “mail” link, on the top of the main interface (Fig. 1), probably provoking user frustration.

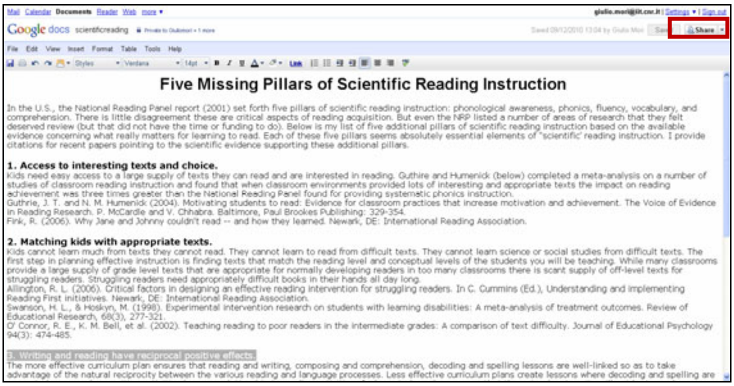


Fig. 2. Editor interface of Google Documents

(a)

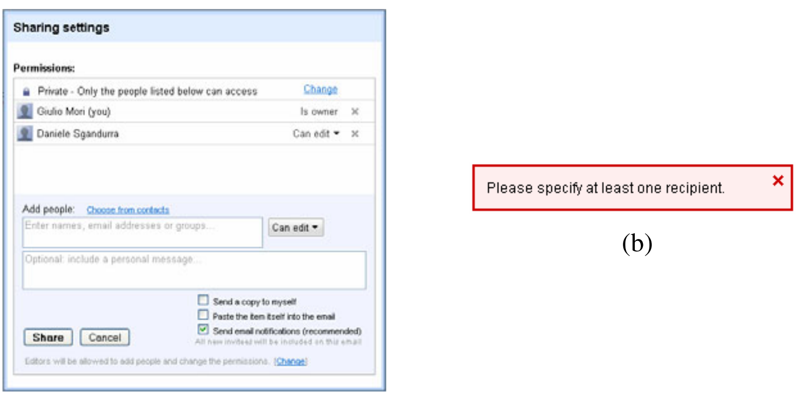


Fig. 3. a) The sharing settings window; b) No added people warning message

2.2 Get the Link to Share/Email as Attachment/Send Message to Collaborators

These functions are not accessible through the *“share”* pull-down menu (Fig. 1) both on Internet Explorer and Firefox. On the contrary, the share functionalities are accessible only using Firefox pressing the key simulating the right mouse button (after the selection of the associated checkbox to a document – Fig. 1), or using the *"Share"* pull down menu on the editor (Fig. 2).

The document can be shared with other collaborators sending an email (message) containing the link to the document (Fig. 4b) or an email containing the document as an attachment (Fig. 4a). These window interfaces are reachable on Firefox but present these issues (Fig. 3a): a) labels are not announced by the screen reader; b) windows sometimes lose focus; c) warning message (Fig. 4c) when no email addresses are inserted is not announced; d) after closing the "Email as Attachment" and "Send messages" windows (Fig. 4a, 4b), focus goes to the "mail" link, on the top of the main interface (Fig. 1).

Our test interacting via JAWS screen reader and keyboard emphasized some difficulties inserting email addresses on the "Email as Attachment" window.

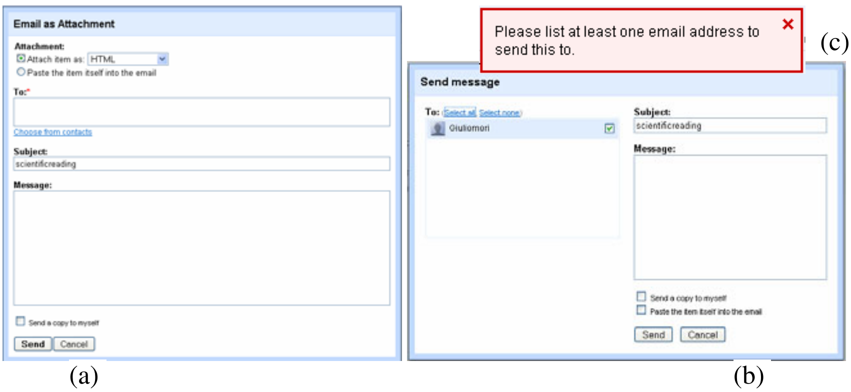


Fig. 4. a) Email as Attachment window; b) Send message window; c) warning message

2.3 See Who Is Accessing

Considering the editor, the screen reader is not able to perceive who is working at the same time on the same document, as appears visually (Fig. 5). Sometimes, when two or more collaborators are modifying the same part of a document, a warning message appears (Fig. 6). This message is accessible, but user awareness (i.e. a user's knowledge about the actions that other users are performing in the system) depends on whether the focus is on the window.

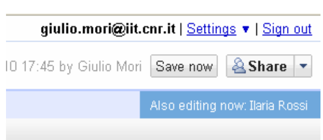


Fig. 5. Feedback about who is accessing

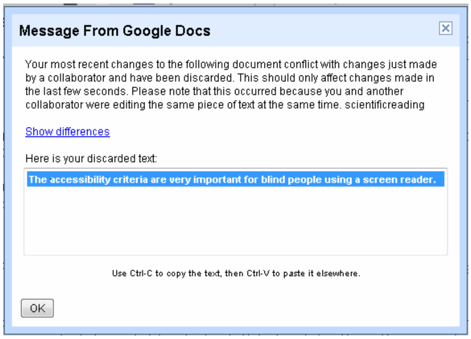


Fig. 6. Warning for collaborators modifying the same part of a document

2.4 Publish as Web Page

This function is reachable only using Firefox, selecting the check box associated with a document and pressing the key simulating the right mouse button (Fig. 1). Interactive elements of the publishing interface are accessible, but text is not automatically announced (Fig. 7).

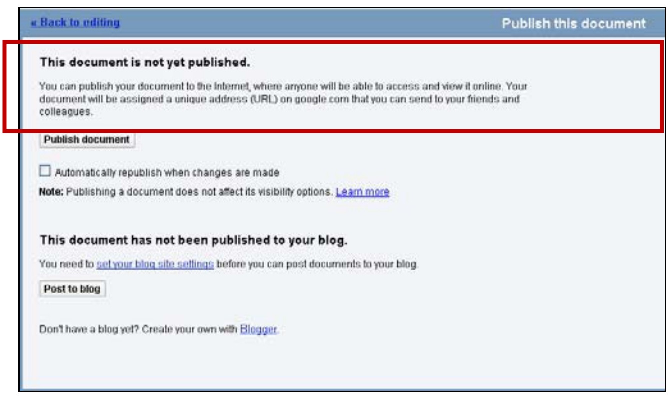


Fig. 7. Publishing interface

2.5 Upload

The *"Upload"* function is not accessible using JAWS v. 10. Using Internet Explorer the button on the main interface (Fig. 1) is announced, but it is not clickable; with Firefox the button is not announced and then is not clickable. In contrast, using JAWS v. 12, the *"Upload"* function is accessible. However, after accessing the upload interface (Fig. 8), the *"Select files to upload"* link is only accessible using Firefox (although it is announced as *"Browser"*). The *"Start upload"* button is accessible with both browsers (although it is useless on Internet Explorer, due to the inaccessibility of the *"Select files to upload"* link).

2.7 See Revision History

This function is reachable only using Firefox, in three ways: 1) selecting the check box associated to a document and pressing the key simulating the right mouse button (Fig. 1), 2) after the selection of the checkbox, through the *"More actions"* pull-down (difficult for the user), 3) in the editor by means of the menu *"File"* (although it works only with JAWS v. 12), which is very hard to access (Fig. 9).

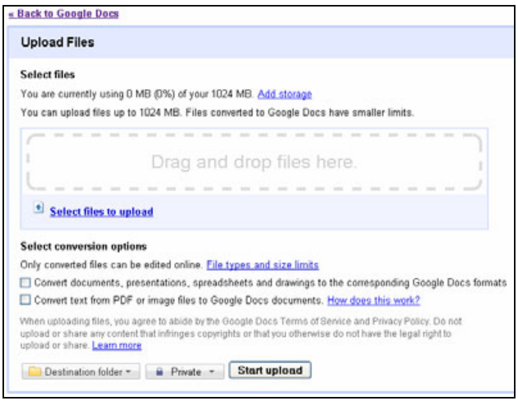


Fig. 8. Upload interface

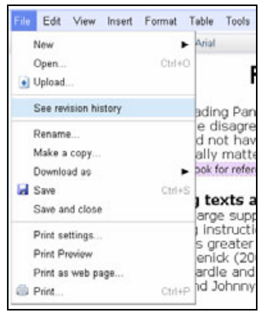
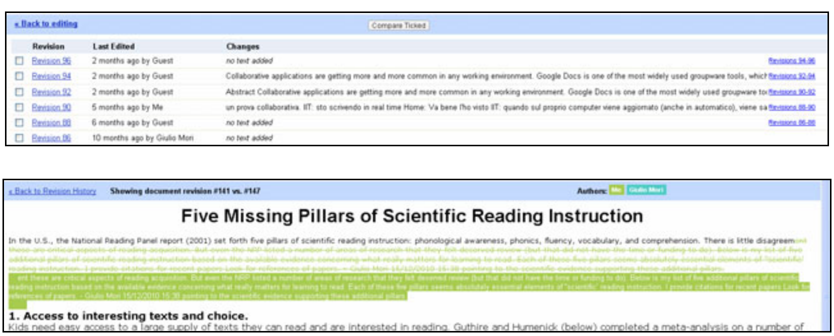


Fig. 9. Revision History interface

Looking at the "Revision History" interface (Fig. 10a), links of different revisions are accessible, but the text of each revision is not announced (also exploring via arrow keys). The revision item on the menu appears only after selecting the checkbox of one document. The "Compare Ticked" button allows to compare two or more different versions of the same document, but the use of this functionality is not simple, because the user must make the extra effort to select at least two checkboxes (each one associated with a revision), and then reach back sequentially to the "Compare Ticked" button. Besides, after pressing the "Compare Ticked" button, the revision differences are not announced by the screen reader (Fig. 10b).



(a)

(b)

Fig. 10. a) Revision History interface; b) Compare Ticked interface

2.7 Comment

A comment can be inserted into a document only using the Firefox browser, through the "Comment" item of the *"Insert"* menu in the editor page (hard to access), so this task is quite difficult to complete for a blind user (Fig.11).

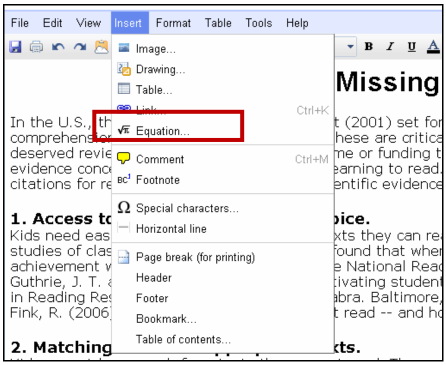


Fig. 11. Not accessible (via screen reader) to insert "Comment" item

However, although the "Comment" function is not easy to complete, as a test we forced the comment, observing that it is not possible to exit from the comment field using the keyboard (Fig. 12). Besides, a blind user cannot understand the difference between a comment and the text contained in the document, since the screen reader does not announce any difference, and it reads it in the same way (Fig. 12).

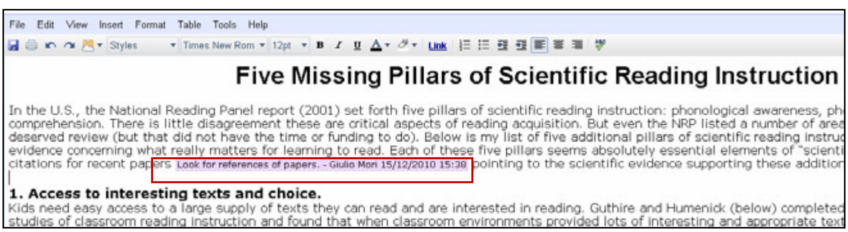


Fig. 12. A comment into a document

3 Some Suggestions for Solving Accessibility Problems

Making the important features and properties in a tool for collaborative editing accessible and usable for blind users presents various peculiarities. A previous study [5] highlighted possible problems arising when accessing Google Docs and writing a document. In this paper, more specific aspects of the collaborative functions available in document sharing are described. Based on the examined case study we suggest a few specific aspects to be keep in mind when thinking of more detailed guidelines or criteria for a more usable editing environment. These criteria should be applied in addition to the general W3C guidelines for Web Content Accessibility [8]. In brief, when developing a collaborative editing environment, it is important to make:

- **Editing operable** – Main editing functions must be operable via keyboard. The edited text has to be readable char by char, word by word, and line by line using an editing cursor. Selection should be possible through the standard operating system shortcuts (e.g. Ctrl+Shift+arrow keys) and highlighted by using standard color combinations (i.e. while conserving the compatibility with the assistive technologies). The focus should be easily switchable between the main areas, such as editing, toolbars and any other menu or interactive elements available for important functions (e.g. Save, Close, etc.). Opening the editor in a new window could greatly improve usability via screen reader.
- **Awareness perceivable** – Information on other connected users should be provided through a suitable method in order to appropriately provide the information to the assistive technologies. Status and actions concerning other collaborators should be made accessible to the assistive technology. Information on who is online or offline, or what they are doing, should be available and easy to obtain at any time.
- **Co-editing understandable** – When two or more users are working on the same document together, information on portions that are free to be edited, on content blocks that are being edited by other users at the same time, and on which parts have been modified, should be adequately provided to the assistive technology so that the user can quickly understand.
- **Dynamic messages and instantaneous feedback** – Feedback and short messages or alerts should be made clear and easy to read. Alerting should be promptly provided to the assistive technology so that users can be readily informed. Short sounds or audio feedback could improve user perception of given events, such as failure or success, or outcomes for specific performing results (e.g., a sound for bold style application, another one for italics, and so on).

Based on these principles to consider, specific guidelines could be better detailed. ARIA-suite, the W3C-WAI Accessible Rich Internet Applications Suite [9], a valuable technical support for implementing more accessible and usable UIs, would also be applicable to collaborative editing environments.

4 Discussion

Collaborative tools are increasingly adopted in environments such as working group activities, e-learning systems, distance education and social networks. Collaboration is important, and increases our opportunities: reinforcing our knowledge, learning more, sharing ideas, getting feedback. For this reasons user interfaces of collaborative software applications should be accessible and usable for all.

In this paper, we have shown issues of interacting via screen reader with Google Docs user interfaces, focusing only on its collaborative features. Manual inspection with the JAWS screen reader has highlighted some major accessibility issues with Internet Explorer, but Firefox also presents some usability deficits. The latest version of JAWS offers benefits when using Firefox, allowing more satisfying interaction for the blind, but JAWS is a commercial product so not all blind users can update immediately to the latest version. Additional effort is required to guarantee easy, effective and satisfying access via screen reader, especially from Google application designers and software engineering, for implementing accessible and usable user interfaces, conforming to the W3C WAI Web Content Accessibility Guidelines v.2.0 and WAI-ARIA.

Last, some basic suggestions have been offered in order to make collaborative environments more useful for blind users. The progress of collaborative environment usability is valuable in general for all users, and the benefits of accessibility will be reflected in the wider diffusion and use of these kinds of tools and applications.

References

1. Craven, J., Brophy, P.: Non-visual access to the digital library: the use of digital library interfaces by blind and visually impaired people. Technical report, Manchester: Centre for Research in Library and Information Management - CERLIM (2003), `http://www.cerlim.ac.uk/pubs/index.php`
2. Ivory, M. Y., Yu, S., Gronemyer, K.: Search result exploration: a preliminary study of blind and sighted users' decision making and performance. In: Extended abstracts of CHI 2004, pp. 453–456 (2004)
3. Petrie, H., Hamilton, F., King, N.: Tension, what tension?: Website accessibility and visual design. In: Proc. 2004 International Cross-disciplinary Workshop on Web Accessibility (W4A), pp. 13–18 (2004)
4. Card, S.K., Moran, T.P., Newell, A.: The Psychology of Human-computer Interaction, pp. 29–97. Lawrence Erlbaum Associates, London (1983)
5. Buzzi, M.C., Buzzi, M., Leporini, B., Mori, G., Penichet, V.M.R.: Accessing Google Docs via Screen Reader. In: Miesenberger, K., Klaus, J., Zagler, W., Karshmer, A. (eds.) ICCHP 2010. LNCS, vol. 6179, pp. 92–99. Springer, Heidelberg (2010)
6. Garrido, J.E., Penichet, V.M.R., Lozano, M.D., Buzzi, M.C., Buzzi, M., Leporini, B., Mori, G.: Analysing Google Documents as groupware tool and an improvement proposal. In: XI Congreso Internacional de Interacción Persona-Ordenador (2010)
7. Leporini, B., Andronico, P., Buzzi, M., Castillo, C.: Evaluating a modified Google user interface via screen reader. The Universal Access in the Information Society (UAIS) 7(3), 155–175 (2008)
8. W3C. Web Content Accessibility Guidelines 2.0. (December 11, 2008), `http://www.w3.org/TR/WCAG20/`
9. W3C. WAI-ARIA Technical Specification, `http://www.w3.org/TR/wai-aria/`

Acoustic Modeling of Dialogue Elements for Document Accessibility

Pepi Stavropoulou[1,2], Dimitris Spiliotopoulos[1],
and Georgios Kouroupetroglou[1]

[1] Department of Informatics and Telecommunications
National and Kapodistrian University of Athens
Panepistimiopolis, Ilisia, GR-15784, Athens, Greece
[2] Department of Linguistics, University of Ioannina,
GR-45110, Ioannina, Greece
{pepis,dspiliot,koupe}@di.uoa.gr

Abstract. Document-to-Audio accessibility assumes that all meaningful presentaion elements in the document, such as bold, italics, tables or bullets, should be properly processed and acoustically modeled, in order to convey the intended meaning to the listeners in a complete and adequate manner. Similarly, several types of documents may contain reported speech and dialogue content signaled through punctuation and other visual elements that require further processing before being rendered to speech. This paper explores such dialogue elements in documents, examines their actual indicators and their use, and investigates the most prominent methods for their acoustic modeling, namely the use of prosody manipulation and voice alternation. It further reports on a pilot experiment on the appropriateness of voice alternation as means for the effective spoken rendition of dialogue elements in documents. Results demonstrate a clear listener preference for the "multiple voice" renditions over the ones using a single voice.

Keywords: Acoustic modeling, document accessibility, dialogue, reported speech, Text to Speech synthesis, voice alternation, Document-to-Audio.

1 Introduction

Work on universal access to documents aims towards making document content accessible to the widest possible range of end users including people with disabilities. Proper adjustments to document layout and text formatting as well as utilization of different modalities are key means to accommodating users "with different abilities, requirements and preferences in a variety of contexts of use" [26]. Accordingly, Text to Speech (TtS) systems transfer document content to the acoustic modality making it accessible to the visually impaired, in eyes-busy situations, in spoken dialogue applications and so forth. For the transfer to be effective document metadata must be utilized and visual elements must be properly rendered to speech as part of a complete Document-to-Audio (DtA) process.

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 175–184, 2011.

2 DtA - Acoustic Modeling of Visual Elements in Documents

In transferring documents from the visual to the aural modality, elements optimized for vision need to be correctly identified, appropriately processed and subsequently vocalized in a manner that improves naturalness, aids comprehension, and minimizes listening effort. Visual presentation elements such as tables, paragraphs, headings and bullets convey semantic and pragmatic information critical for understanding the intended meaning of the text, thus necessitating the use of appropriate acoustic modeling of the underlying logical association between visual structure and the meaning itself. How does one read tables? What is an appropriate prosodic specification for bold or italics? How can bullets or quotation marks be acoustically perceived? Take the minutes of a meeting for instance. A brief inspection of a sample "meeting of minutes" document (Figure 1) reveals several visual elements that on the one hand pose certain challenges for the effective acoustic rendition, while on the other hand provide important information for accessing, interpreting and subsequently communicating the underlying semantic and pragmatic content to the listener.

Minutes of Meeting Title

<table>
<tr><td>Subject</td><td colspan="2">Project Planning</td></tr>
<tr><td>Date</td><td colspan="2">2011-01-24</td></tr>
<tr><td rowspan="3">Participants</td><td>Company A Name</td><td>Company B Name</td></tr>
<tr><td>George P.</td><td>Michael S.</td></tr>
<tr><td>Mary B.</td><td></td></tr>
</table>

Agenda

- Agenda Item 1
- Agenda Item 2

Discussion, Decisions, Assignments

Agenda Item 1
… George P. suggested that deployment should be completed by **Monday** the latest…

Agenda Item 2
….

Follow Up Actions

….

Fig. 1. Sample "minutes of meeting" document

More specifically, these elements are:

- 1. Headings and subheadings indicating basic discourse segmentation and topic hierarchy. Discourse structure in general is marked through variation in prosodic parameters such as pitch range, intensity and preboundary lengthening, as suggested in a number of previous studies [7, 14, 27]. However, most synthesizers today disregard structure beyond the utterance level, making no use of meta-information on paragraph breaks or document sectioning.

- 2. Table structures. A significant amount of work has been devoted to the extraction and meaningful reconstruction of the logical relations that are implicit in the tabular layout [2, 3, 4, 11, 15, 18, 20, 28]. Once the underlying logical relations are extracted and reconstructed in a way suitable for non-visual modalities, the resulting semantic representation may be utilized through proper parameterization of Text to Speech synthesis [22, 23]. Previous work has demonstrated that appropriate prosody control and the use of earcons and spearcons improve naturalness, acceptance and listening effort especially in the case of complex table structures in which simple linearization techniques are proven inadequate [24].
- 3. Types of list structures such as bulletin. In a study of spoken lists [19] showed that appropriate use of rhythm and pitch decreases cognitive effort and aids recall. Xydas et al. [29] demonstrated listeners' preference for a combination of modifications in volume, pitch and use of earcons for the discrimination among bold letters, italics and bulletin respectively. In general, use of non speech sounds such as tones or beeps is considered a particularly effective way for introducing list items and vocalizing bullets [5].
- 4. Bold formatting indicating emphasis. In a way similar to the aural rendition of other metadata and visual emphatic events, pitch modification on the bold word and increase in volume have been utilized in cueing bold letters [26, 29].
- 5. Reported speech elements. While our example involves an instance of indirect reported speech ("George P. suggested that…") that is not typographically set off, there are many types of documents such as interviews, transcripts, proceedings and narratives that include an abundance of visual elements cueing the presence of direct speech and dialogue. Such elements and their subsequent effective vocalization constitute the focus of this paper.

In particular, this paper examines the appropriateness of synthetic voice alternation as a means for transferring reported direct speech and dialogue content in written documents to the acoustic modality. In the following sections, we first briefly analyze the basic types of reported speech along with the visual components and other cues used in documents for denoting them. Next, the most appropriate means for successfully rendering them to the aural modality are presented, namely prosody parameterization and voice alternation. A pilot psychoacoustic experiment comparing listeners' perception of a single voice rendition to their perception of the one using voice alternation is presented in sections 3 and 4. Major findings and directions for future work are discussed last.

3 Speech and Dialogue Elements in Text

Following Sinclair [21] there are two main ways for reporting one's words when writing, namely *quote structures*, also referred to as *direct speech*, and *report structures*, also referred to as *indirect speech*. In the first case, the speaker's exact words are reported, while in the second case there is no exact reproduction involved, rather certain changes apply on the original utterance's grammar structure and content. Intuitively, the action of reporting another utterance constitutes a universal communicative and linguistic phenomenon, while at the same time a clear cut distinction between direct and indirect speech is found encoded in a good number of languages around the world [8].

Though indirect speech is indicated in written text solely on a lexico-syntactic basis through the use of reporting phrases such as "John said" along with changes in personal, temporal or locative references, direct speech is, in addition, cued through certain visual components applied directly onto the text or in the form of meta-information embedded in the source document, These components provide visual cues to the existence of direct speech and other dialogue elements such as dialogue turns. Following are the most common and widely used visual indications to speech and dialogue elements within written discourse:

Quotation mark pairs are most often used to delimit the beginning and end of direct speech. Depending on specific language conventions quotation marks may come in different forms such as single ('') or double ("") inverted commas, double angle quotes («») or corner brackets. Speakers may also alternate between different forms, in order to denote the presence of a nested quote (a quote within a quote). The accompanying reporting phrase is placed outside the quotation marks and can be positioned at the beginning, at the end or within the quote structure; when the reported "voice" has already been established in context, the reporting phrase is often omitted.

When dialogue is reported, turn taking is indicated through the use of line or paragraph breaks. In dialogue inverted commas are often replaced by a quotation dash corresponding to a single dialogue turn. Furthermore, the name of the speaker may be used followed by colons (:) or a quotation dash (.-). This explicit mentioning of speakers' names greatly simplifies the task of matching each speaker to the correct turn. Finally, in interviews dialogue is sometimes indicated through the use of bold letters for representing the interviewer's utterance. Again turn taking is cued through line or paragraph breaks. Table 1 summarizes quoting styles and visual cues to reported speech providing examples for each one.

Table 1. Reported speech indicators

	Indirect Speech	Direct Speech
Deictic references and syntactic dependence	He complained ***that*** nobody came to ***his*** party	–
Reporting phrases	***He complained*** that nobody came	"Nobody came", ***he complained***
Quotation marks	–	"..'' / «…» / '…' etc
Quotation dash	–	**-** Nobody came, he said.
Colons	–	**JUDGE:** Can the defendant please rise?
Line breaks	–	Line or paragraph breaks indicating speakers' turns

At this point it should be noted that, while there is an abundance of visual cues that are more or less consistently used, the functional load of each element (e.g. the use of colons before enumeration or the use of inverted commas to denote irony and so forth) as well as the complexity of certain dialogues may pose several challenges to the identification of dialogue within text, including the identification of dialogue participants and the correct assignment of each turn to the respective participant. Addressing these issues, however, is beyond the scope of this paper.

Punctuation and other visual cues within written text often serve as a substitute of prosody in speech [9]. Subsequently, visual markers of reported speech such as quotes or line breaks should *in a broad sense* correspond to prosodic markers of reported speech in spoken discourse. Accordingly, in a study of informal conversations examining the relationship between reported speech and prosody in English, Klewitz and Cooper [13] found instances of reported speech to correlate with shifts in pitch range, intensity, speech rate and perceptually isochronous rhythmic patterns, as well as paralinguistic expressive qualities such as breathy or nasal voice. Their data further suggests that in dialogue distinct prosodic marking may be assigned to the different "voices" (i.e. interlocutors) reported, facilitating the hearer's task of keeping track of *who is speaking now*. In addition, Jansen [12] reported on a statistically significant expansion of overall pitch range of direct speech compared to both surrounding narrative segments as well as instances of indirect speech. Furthermore, direct speech was found to be more often preceded by stronger intonational breaks. In contrast, no statistically significant differences were reported between indirect speech and surrounding narration with regards to the prosodic parameters attested. Similar effect of pitch range has also been demonstrated for direct speech in Brazilian Portuguese [17]. In short, taking into account that prosody has been shown to function as a significant marker of discourse structure in general [1, 7, 10, 14, 16] (among others), and reported speech instances constitute more or less clearly demarcated discourse segments with particular discourse functions, the appropriate manipulation of prosodic parameters such as pitch range, intensity, lengthening and pausing is expected to play a key role in successfully transferring reported speech back to the aural modality.

In addition to prosody manipulation, in the process of rendering documents to speech, TtS systems can provide further means for signaling reported speech and dialogue. Intuitively, voice alternation – switching between different synthetic voices over the course of the interaction – is expected to be the most appropriate other medium for signaling multiple voices in written dialogue. In a similar vein, voice alternation has been used for landmarking and context setting in automated spoken dialogue systems. Association of different synthetic voices to specific dialogue states may improve system navigation, and increase user confidence and awareness of dialog progress.

The study presented in this paper focuses on the use of voice alternation for the effective acoustic rendition of direct speech and dialogue in particular. Direct speech was preferred over indirect speech as it is more explicitly related to the existence of a different – other than the narrator's – voice, bringing forth a distinct speech situation and reproducing the perspective of the original utterance. As such, it is better suited for switching to a different voice. In contrast, indirect speech is more tightly integrated within the embedding context, maintaining the perspective of the narrator and lacking syntactic independence and expressive properties (e.g. use of exclamation marks). Accordingly, it is not distinguished typographically and speakers do not tend to prosodically mark it [12]. Furthermore, dialogue was suited for testing more than two voices as well as assessing the effect of voice alternation on facilitating comprehension. Written dialogue often lacks explicit "tagging" of interlocutors through the use of reporting phrases (e.g. "he said"). In effect line breaks are often the only means left signaling turn taking, rendering their effective transfer to the acoustic modality crucial for determining *speakership*, i.e. understanding "who is speaking now".

4 Experimental Setup

In accordance with the above a pilot experiment was carried out comparing two versions of synthetic speech renditions, one using voice alternation and one in which only a single synthetic voice is used. The two renditions were compared in terms of both objective and subjective criteria. As part of the objective evaluation, participants were asked to answer a set of questions aiming to assess their degree of comprehension and appropriate alignment of turns to interlocutors. For the subjective evaluation participants were asked to assign a score on a Likert scale ranging from 1 to 5 evaluating overall impression, acceptance, ease of comprehension and naturalness of each rendition.

Materials consisted of three texts rendered in both – single voice and multiple voices – conditions. The first text was a narrative passage with two characters, the narrator being one of the characters. Dialogue turns were marked by quotation dashes, line breaks and reporting phrases. Reporting phrases were either omitted or placed right before the quote. The second text was also a narrative passage with two characters; only this time the narrator was not a character in the story, and thus there were three voices involved in total. Dialogue turns were marked through double inverted commas, line breaks and reporting phrases. The reporting phrases were either omitted, placed in the beginning, at the end or in the middle of the quote. In the later case, the reported speaker's dialogue turn was interrupted by the narrator's voice. The third text was a "transcripts of trial" document with three voices and no narrator. Dialogue turns and speakers' identity were marked explicitly through the use of the speaker's name followed by colons at the beginning of each turn. Thus no ambiguity regarding speakership arose. For the multiple voice rendition of the third text, each speaker's name was mentioned only once when it first occurred and was afterwards omitted. In other words, once the mapping between speaker and synthetic voice had been established, the identification function performed by the mentioning of the speaker's name at the beginning of each dialogue turn was now fulfilled through switching to a different voice, in an attempt to achieve faster and still intelligible interaction.

The materials were presented to two groups of four first-time listeners. Each group was presented with the materials in reverse order to ensure that order or memory effects were factored out and did not bias the results. For the narrative passages the following procedure was followed: After having listened to one condition, subjects answered a set of comprehension questions regarding text's content. For each answer they further provided a degree of confidence ranging from 1 to 5, with 5 being "absolutely certain that my answer is correct" and 1 being "completely unsure". In addition, they again assigned a score on a scale from 1 to 5 grading acceptance, naturalness and listening effort of the rendition as a whole. Next, they listened to the second condition. At the end they declared their preference between the two renditions, ranked each rendition for overall impression and made any other comments that they considered helpful. For the trial transcripts no comprehension questions were asked, as the structure of the document unambiguously determined speakership.

5 Results

Overall, subjects preferred the multiple voice renditions to the single voice ones. They considered the former to be more appropriate and easier to understand. More specifically, the multiple voice renditions were preferred in 21 out of 24 cases in total. One particular listener declared a preference on the use of a combination of features found in each rendition separately, namely the explicit use of the interlocutor's name before each turn along with the switch to a different voice.

Figure 2 illustrates listeners' ranking of each rendition as far as overall impression, listening effort and naturalness are concerned. Ranking scale ranged from 1 to 5, 5 being the optimal rank. As can be seen from figure 2, the multiple voice renditions scored higher, achieving an average 0.8 points improvement on all criteria examined. Furthermore, subjects proclaimed a higher degree of confidence with regards to their feedback in the comprehension task.

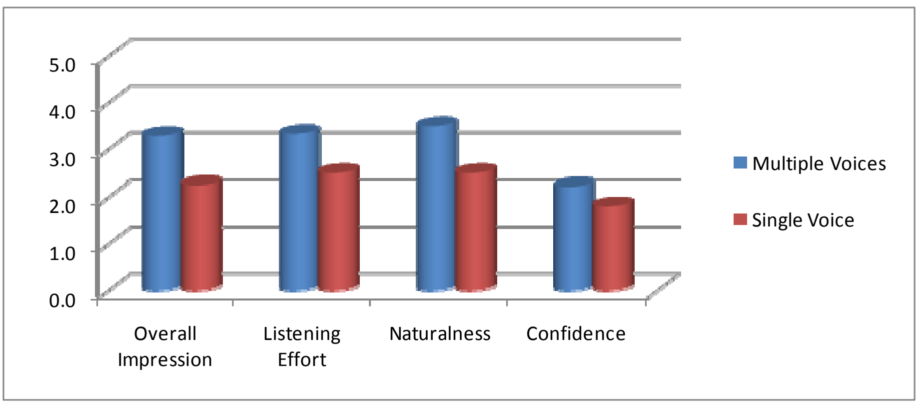


Fig. 2. Subjective evaluation criteria

Subjective evaluation results were indeed corroborated by subjects' performance on the comprehension task. The subjects' answers were 94% accurate in the case of the multiple voice renditions contrary to the single voice renditions for which accuracy was limited to 75%. Response times were also 10-20% faster for the multiple voice renditions. Still, even the multiple voice renditions were considered inadequate in some instances: more specifically, 25% of the total answers to the question "Do you consider this type of rendition acceptable?" were *NO*. For the single voice renditions, there was only one *YES* answer.

6 Discussion – Future Work

The results of the experiment presented here confirm the main hypothesis that voice alternation can effectively be used to model dialogue elements in documents, minimizing listening effort and facilitating comprehension. The utilization of different synthetic voices is particularly useful for determining speakership, that is the alignment of interlocutors to particular turns. Speakership itself is a particular instance of footing [6] that places a significant load on cognition.

Nevertheless, voice alternation alone sometimes fails to meet user expectations and reach level of acceptance. The latter necessitates the use of other means such as prosody control for improving the quality and legibility of the rendition. As noted, human speakers use pausing, alter pitch or employ paralinguistic devices – among others – to signal reported speech. Thus, further experimentation is required for determining the most appropriate combination of prosodic parameterization and voice alternation.

On a final note, modeling of reported speech may also prove to be a promising area for the study of emotional speech as part of affective computing. In particular, several reporting phrases used in quote structures may indicate the manner of speaking as well as the speaker's emotional state. Verbs and phrases such as "scream", "yell", "whisper", "plead", "reply angrily" or "cry" can serve as cues for detecting and simulating emotions in an effort for more natural and expressive interaction. In fact, the human paradigm suggests that human narrators employ certain paralinguistic devices in order to convey the reported speaker's emotional state and attitude [13].

7 Conclusion

The empirical evidence presented in this paper favors the use of voice alternation for the acoustic modeling of dialogue elements in written documents, in an attempt to make document content universally accessible. When speech reported within written text is rendered back to the aural modality, there should be certain acoustic cues to the beginning and end of the reported speech stretch as well as to each dialogue turn, in order for the listener to correctly identify and comprehend the intended dialogue structure. Voice alternation is one of them. Appropriate prosody modeling is another medium that certainly calls for further research. In any case, prior to any acoustic modeling, the correct identification of reported speech stretches within text as well as the appropriate assignment of "speakership" require extensive document preprocessing that constitutes a demanding and important task in its own right.

References

1. Den Ouden, H., Noordman, L., Terken, J.: The prosodic realization of organizational features of texts. In: Proc. Speech Prosody 2002, pp. 543–546 (2002)
2. Chen, H.-H., Tsai, S.-C., Tsai, J.-H.: Mining tables from large scale html texts. In: Proceedings of the 18th International Conference on Computational Linguistics, Saarbrucken, Germany (2000)
3. Embley, D.W., Hurst, M., Lopresti, D.P., Nagy, G.: Table-processing paradigms: a research survey. Int. J. Document Analysis 8(2-3), 66–86 (2006)

4. Filepp, R., Challenger, J., Rosu, D.: Improving the Accessibility of Aurally Rendered HTML Tables. In: Proc. ACM Conf. on Assistive Technologies (ASSETS), pp. 9–16 (2002)
5. Fröhlich, P.: Increasing Interaction Robustness of Speech-enabled Mobile Applications by Enhancing Speech Output with Non-speech Sound. In: Proc. ROBUST 2004, COST278 and ISCA Tutorial and Research Workshop (ITRW) on Robustness Issues in Conversational Interaction, Norwich, England (August 2004)
6. Goffman, E.: Forms of talk. Basil Blackwell, Oxford (1981)
7. Grosz, B., Hirschberg, J.: Some intonational characteristics of discourse structure. In: Proceedings of the 2nd International Conference on Spoken Language Processing, Banff, Canada, pp. 429–432 (1992)
8. Haberland, H.: Reported Speech in Danish. In: Coulmas, F. (ed.) Direct and Indirect Speech. Trends in Linguistics, Studies and Monographs, vol. 31. Mouton de Gruyter, Berlin (1986)
9. Halliday, M.A.K.: Spoken and written language. Deakin University Press, Geelong (1985)
10. Herman, R.: Intonation and discourse structure in English: Phonological and phonetic markers of local and global discourse structure. PhD Thesis (1998)
11. Hurst, M., Douglas, S.: Layout & Language: Preliminary Experiments in Assigning Logical Structure to Table Cells. In: Proc. 4th Int. Conf. Document Analysis and Recognition (ICDAR), pp. 1043–1047 (1997)
12. Jansen, W., Gregory, M.L., Brenier, J.M.: Prosodic correlates of directly reported speech: Evidence from conversational speech. In: Proceedings of the ISCA Workshop on Prosody in Speech Recognition and Understanding, Red Banks, NJ, pp. 77–80 (2001)
13. Klewitz, G., Couper-Kuhlen, E.: Quote-unquote? The role of prosody in the contextualization of reported speech sequences. Pragmatics 9(4), 459–485 (1999)
14. Lehiste, I.: Some Phonetic Characteristics of Discourse. Studia Linguistica 36(2), 117–130 (1982)
15. Lim, S., Ng, Y.: An Automated Approach for Retrieving Hierarchical Data from HTML Tables. In: Proc. 8th ACM Int. Conf. Information and Knowledge Management (CIKM), pp. 466–474 (1999)
16. Nakatani, C., Hirschberg, J., Grosz, B.: Discourse Structure in Spoken Language. Studies on Speech Corpora (1995)
17. Oliveira, M., Cunha, D.A.C.: Prosody as Marker of Direct Reported Speech Boundary. In: Speech Prosody 2004, Nara, Japan (March 23-26, 2004)
18. Oogane, T., Asakawa, C.: An Interactive Method for Accessing Tables in HTML. In: Proc. Intl. ACM Conf. on Assistive Technologies, pp. 126–128 (1998)
19. Pitt, I., Edwards, A.: An Improved Auditory Interface for the Exploration of Lists. ACM Multimedia 1997, 51–61 (1997)
20. Pontelli, E., Gillan, D., Xiong, W., Saad, E., Gupta, G., Karshmer, A.: Navigation of HTML Tables, Frames, and XML Fragments. In: Proc. ACM Conf. on Assistive Technologies (ASSETS), pp. 25–32 (2002)
21. Sinclair, J.: Collins Cobuild English Grammar. Harper Collins, London (2002)
22. Spiliotopoulos, D., Xydas, G., Kouroupetroglou, G.: Diction Based Prosody Modeling in Table-to-Speech Synthesis. In: Matoušek, V., Mautner, P., Pavelka, T. (eds.) TSD 2005. LNCS (LNAI), vol. 3658, pp. 294–301. Springer, Heidelberg (2005)
23. Spiliotopoulos, D., Xydas, G., Kouroupetroglou, G., Argyropoulos, V., Ikospentaki, K.: Auditory Universal Accessibility of Data Tables using Naturally Derived Prosody Specification. Univ. Access Inf. Soc. 9(2), 169–183 (2010)

24. Spiliotopoulos, D., Stavropoulou, P., Kouroupetroglou, G.: Acoustic Rendering of Data Tables using Earcons and Prosody for Document Accessibility. In: Stephanidis, C. (ed.) UAHCI 2009. LNCS, vol. 5616, pp. 587–596. Springer, Heidelberg (2009)
25. Stephanidis, C., Akoumianakis, D., Sfyrakis, M., Paramythis, A.: Universal accessibility in HCI: Process-oriented design guidelines and tool requirements. In: Stephanidis, C., Waern, A. (eds.) Proceedings of the 4th ERCIM Workshop on User Interfaces for All, Stockholm, Sweden, October 19-21 (1998)
26. Truillet, P., Oriola, B., Nespoulous, J.L., Vigoroux, N.: Effect of Sound Fonts in an Aural Presentation. In: 6th ERCIM Workshop, UI4ALL, pp. 135–144 (2000)
27. Wightman, C., Shattuck-Hufnagel, S., Ostendorf, M., Price, P.: Segmental durations in the vicinity of prosodic phrase boundaries. Journal of the Acoustical Society of America 91(3), 1707–1717 (1992)
28. Yesilada, Y., Stevens, R., Goble, C., Hussein, S.: Rendering Tables in Audio: The Interaction of Structure and Reading Styles. In: Proc. ACM Conf. Assistive Technologies (ASSETS), pp. 16–23 (2004)
29. Xydas, G., Argyropoulos, V., Karakosta, T., Kouroupetroglou, G.: An Experimental Approach in Recognizing Synthesized Auditory Components in a Non-Visual Interaction with Documents. In: Proc. Human-Computer Interaction - HCII 2005 (2005)

Part III

Universal Access in Complex Working Environments

Seeing the Wood for the Trees Again! SMART - A Holistic Way of Corporate Governance Offering a Solution Ready to Use

Fritz Bastarz and Patrick Halek

rep. by Infomedia Services GmbH
Rosenbursenstraße 4, 1010 Vienna, Austria

Abstract. SMART is a technology-supported solution for holistic corporate governance in a complex environment. It quickly delivers practical operational processes and information objects ready to use. This makes it possible to solve complex tasks immediately and effectively. By means of integrated information and process management, processes can be evaluated quickly and integrated easily into everyday business on the grounds of actual needs. For using SMART no special skills are needed. The principle of SMART is the stakeholder based systemic designing, analysing and seizing of topics or tasks. This is assured by considering all relevant factors of influence and the possibility of the quick, direct and easy implementation of results. They can be executed immediately and practically within assessable processes ready for use within the organisation. The goal of SMART is to obtain a holistic understanding of complex tasks in order to gain a better and more sustainable result.

Keywords: Corporate Governance, Corporate Performance, Stakeholders, Lateral Thinking, Systemic Thinking, Complexity.

1 The Challenge

It is a long accepted fact that circumstances in society and economy have become more and more complex and interconnected today (Vester 2007). For many of our traditional ways of thinking and acting are based on linear and hierarchical approaches they are not suitable for solving complex and multidimensional tasks successfully anymore (Heftberger/Stary 2004).

The knowledge society is creating new opportunities but also calls for a new thinking and acting (Halek/Nyiri 2002). While the transparency of social and economic processes can be increased through modelling, it is difficult to understand coherences and to predict or shape trends (Senge 2006). Context is getting lost while focussing on details. Decision makers see the trees, but not the wood anymore. But how can organisations remain competitive sustainably in such an environment at all (Kiessling/Bable 2007)? Companies have to take into account many more factors than they had to yesterday. They have to balance a turbulent environment, moving markets, different stakeholders, new social rules and patterns of thinking and behaviour

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 187 – 194, 2011.

(Haas/Oetinger/Ritter/Thul 2007). Only shifting data does not create value when striving for entrepreneurship (Nyiri 2007). It requires creative processes resulting in innovative and competitive positions. But what are the methods and tools to respond to these challenges fairly?

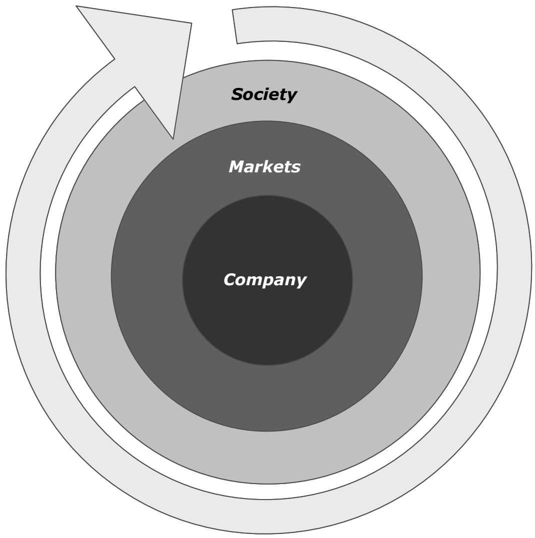


Fig.1. Interdependence of company, markets and society

One-dimensional approaches result in methods and tools that cannot meet the multi-dimensional requirements (Ossimitz 1995). Newer approaches, such as Lateral Thinking and Systemic Thinking can develop high quality solutions in a very pragmatic way. But humans have limited capabilities in dealing with abstract thinking (Glasersfeld 1995, 1997). Therefore, organisational development requires a tool that can be used pragmatically to deal with these approaches effectively.

Decision makers have to master capturing, analysing and designing of complex tasks in order to promote a focused and productive development. Organisational development means improving the efficiency in a defined direction with clearly defined tasks (Grimberger 2009). It requires clarity, agility, transparency and compliance to actually lead to more effectiveness and efficiency in business operation. Consequently, the goals of organisational development are:

- improving the quality of decisions
- improving the efficiency of communication
- saving the setup of actions
- accelerating task implementation
- flexible adaptation to changing circumstances

In the following we present a solution actually created for being used in business development without distracting running processes.

2 SMART

2.1 Towards a Practically Relevant Solution

The crucial step in organisational development is to use a tool allowing to display and evaluate complexity and building links as far as possible, since complexity can never be fully represented (cf. Luhmann 2006, 2008). Therefore a solution needs to recognise the context of processes and activities. The latter is also required to evaluate details properly. In SMART priorities can be set in a flexible, fast and easy way, depending on the business context. SMART is the technology-based implementation of a holistic organisational governance approach that links the levels of the organisation and its relevant stakeholders, and allows representing all related factors.

This creates an important basis for decisions. SMART is a lateral way to learn, sense and act succsessfully in a complex environment. Therefore, the most important requirement is practical relevance.

The abbreviation *SMART* has been created using the respective first letter of the following factors:

- <u>SYSTEMIC</u>: holistic, in order to allow the basic understanding of a situation and recognising of principles, patterns and dependencies.
- <u>MODEL-DRIVEN</u>: a structured description and interpretation of relevant realities.
- <u>ACHIEVABLE</u>: feasible, thus to be implemented effectively.
- <u>RELATED</u>: not only linear hirarchies but relations and their impacts are part of improvement. Not only the formal, certified processes are considered, but also the informal and human motives of relevant stakeholders determining success or failure.
- <u>TIMELY</u>: up to date and meeting today's requirements. Networking in companies, even across company boundaries, is facilitated through planning, implementation, and communication support.

Therefore, SMART is based on the grounds of the following considerations:

- Persons have not become less capable, the environment has become more complex.
- There is often a lack of differentiation of the terms complicated and complex.
- Mostly the overall view, i.e. the "Big Picture" is missing.
- Complex situations mostly have several solutions (solution scenarios) – the quality of solutions is the key to further potentials or deficits that arise from it.
- The basic requirement of organisational development is clearly defined objectives.
- Working on different levels of abstraction creates a cognitively familiar image (i.e. 3D model) of the requirements. This "image" can then be viewed and analysed from different perspectives - an isolated view of individual levels does not allow the intellectual reproduction of the actual structure.
- The implementation is supported by the practically relevant integration of an IT-tool which has been tested in every day business.

1.2 "smart4sense2act"

SMART follows a very specific way of thinking and acting: we call it "*smart4sense2act*" (read: *smart-for-sense-to-act*). This approach is inextricably linked to the principles of SMART. The focus is on the terms "sense" and "act", as they refer to all states from capturing the situation and defining all relevant stakeholders to executing the solution in an integrated way.

SENSE
SENSING means the capturing, perceiving and recognising of important events, factors and stakeholders – it means to feel and to interpret. In the corporate meaning it is like putting out the feelers in order to understand the environment, their own company and its stakeholders and to properly interpret their actions (Halek 2009).

ACT
ACTING means acting as a result of being in need for action that arises from different perceptions (sensing) and calls for clearly defined activities. The quality of the action depends on the stakeholder abilities and the functionality of the chosen tools.

1.3 The SMART-Principle

Based on this approach, the SMART-principle can be described as follows. Depending on each specific task, step 1 defines the topic space (which defines what has to be focussed on) und all areas of knowledge and stakeholders to be involved (we call it domain, i.e. the knowledge of a special field). Hence, step 1 results in the *levels of domains*.

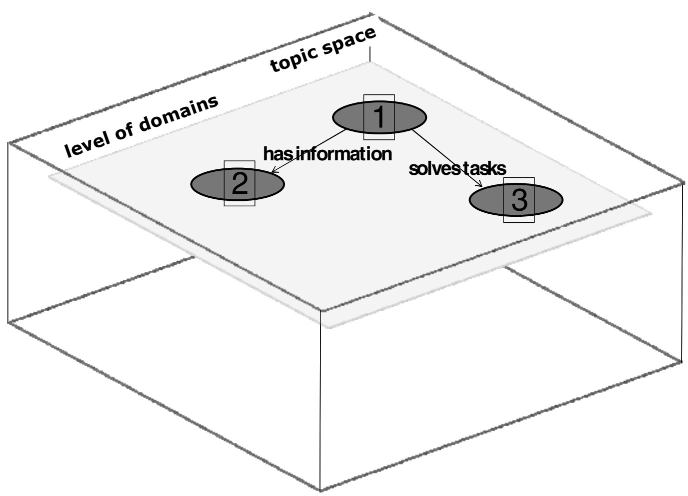


Fig. 2. The SMART-Principle, level of domains (step 1)

Step 2 defines all entities which are related to the domains (entities of all involved domains). Based on these definitions, now all involved sections can be identified (we call them elements). It is crucial to define only those tasks that are involved and relevant for the structures. This way the amount of information is manageable.

After having described the relevant topic space, now the tasks and their linked actions and activities can be linked to the elements (step 3). During this step different perspectives must be considered to obtain a valid structure. This creates new knowledge that will require to adapt the structures of domains again. This iterative procedure makes it possible to gather all relevant factors and link them.

Now the respective computer model can be evaluated according to various aspects. It provides the linked parts of the overall model for the task. In this way all necessary

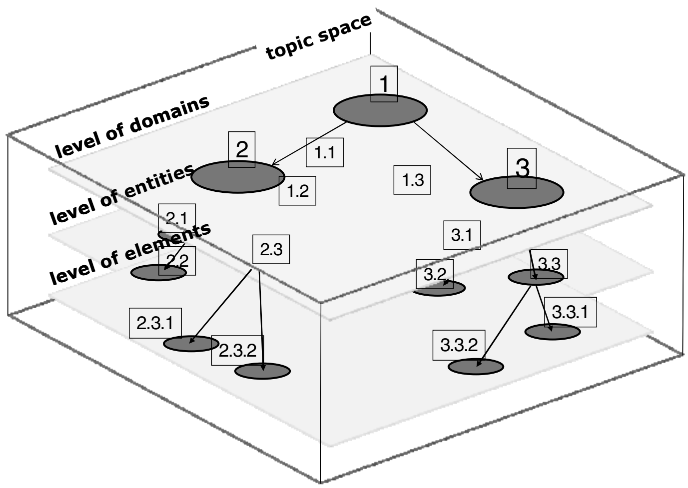


Fig. 3. The SMART-Principle, levels of domains, entities and elements (step 2)

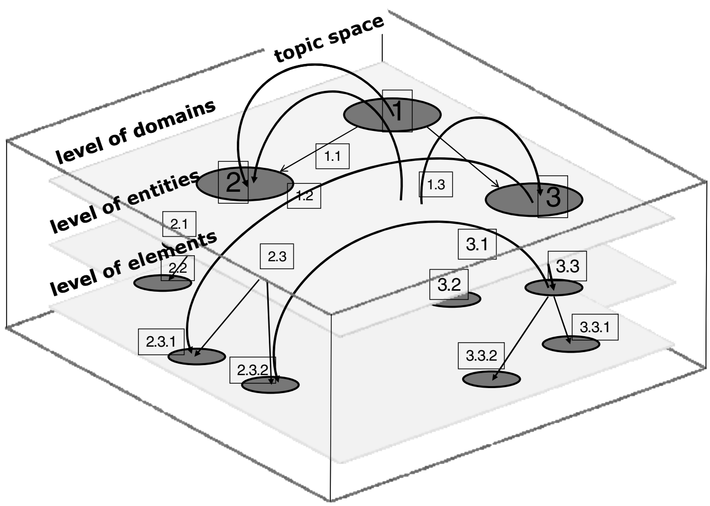


Fig. 4. The SMART-Principle, linking all levels (step 3)

steps, the roles involved and the actions and activities as well as the detailed descriptions of structures become available. These detailed descriptions are given concrete form through the creation of documents. The resulting filtering allows to show complex structures at different levels of granularity (including the meta-level) and to display it in an intelligible way. Now the created network represented by various stakeholders can be analysed for weaknesses and cross-linked with the help of the *Value-Network-Analysis (Allee 2003)*.

At critical points which could not be identified in the overall model so far, a more detailed and filtered analysis can be performed. Adaptions based on computer visualisation and integrated documentation can be done in far better quality and much safer than with the support of conventional methods. As a result of this analysis necessary adaptions can be represented immediately.

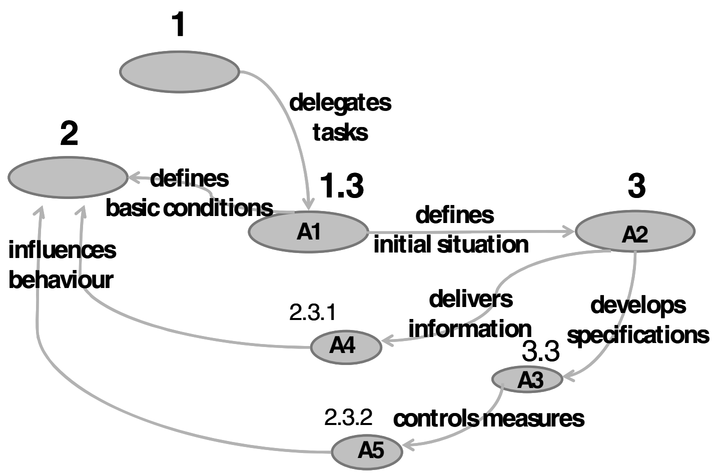


Fig. 5. The SMART-Principle, cross-linking

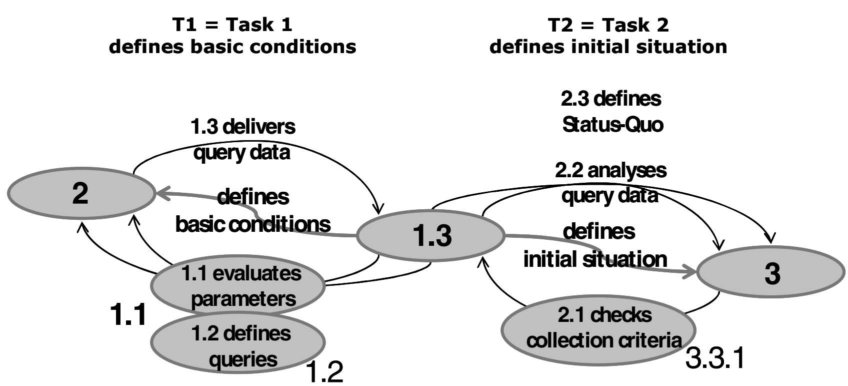


Fig. 6. The SMART-Principle, filtering

3 Summary

SMART was created based on findings that the knowledge society cannot be implemented with previously lived patterns of thinking and acting (Yeo 2009). Multidimensional tasks which are subject to the principles of networking and complexity cannot be solved with the tools of a hierarchically structured and linear world (Pullen/Beech/Sims 2007). SMART is a novel method development. It provided specific tools for networks and complex environments, not only to meet complexity and enable knowledge-orientation, but especially to facilitate organizational changes in every-day-business. SMART ensures seeing the wood for the trees again, as it works in a context-sensitive (systemic) while focussed way. It takes into account only elements that can be implemented in every-day-business, as they lay ground to competitive advantages (Varela/Thompson/Rosch 1995). SMART allows not only responding to new challenges, but also creating networks, and recognising complex environments as knowledge pool quickly and without barriers.

Therefore, the key to define processes is to focus on the stakeholders and their stakes involved (Nel 2008). Organisations are not only driven by formal structures and processes, but also strongly by informal processes, values and the implicit knowledge of the people who are in touch with these organisations (Henriksen 2010). The latter allow discovering and developing organisational potential. SMART offers specific IT-supported tools to implement the method step by step. The focus of the development is on barrier-free accessibility and integration into everyday business because the needs of today demand a straightforward procedure. Processes can be designed as a part of daily business and adapted flexibly and quickly to meet new requirements – without the need of discussing the complete set of business processes.

References

Allee, V.: The Future of Knowledge: Increasing Prosperity through Value Networks. Butterworth-Heinemann, Burlington (2003)

Fischer, H., Fleischmann, A., Obermeier, S.: Geschäftsprozesse realisieren – Ein praxisorientierter Leitfaden von der Strategie bis zur Implementierung. GWV-Fachverlag, Wiesbaden (2006)

Glasersfeld, v. E.: Über Grenzen des Begreifens, Benteli, Bern (1995)

Glasersfeld, v.E.: Wege des Wissens. Konstruktivistische Erkundungen durch unser Denken. Carl-Auer-Verlag, Heidelberg (1997)

Grimberger, G.: The High-IQ Company: The Development of the Organisational IQ. VDM Verlag, Saarbrücken (2009)

Haas, B., Oetinger, R., Ritter, A., Thul, M.: Nachhaltige Unternehmensführung. Verknüpfung wirtschaftlicher, sozialer und gesellschaftlicher Forderungen, Hanser, Munich (2007)

Halek, P., Nyiri, A.: Das Heartbeat-Modell. Der Brückenschlag zwischen strategischem Marketing und Wissensmanagement, Gabler, Wiesbaden (2002)

Halek, P.: Die Marke lebt! Das All-Brand-Concept. Die Marke als Kern nachhaltiger Organisationsführung. Facultas.wuv, Vienna (2009)

Heftberger, S., Stary, C.: Partizipatives organisationales Lernen, Ein prozessbasierter Ansatz. GWV-Fachverlag, Wiesbaden (2004)

Henriksen, D.T.: A little More Conversation, a little Less Action, please: Rethinking learn-ing games for organisational development and adult education. Lambert Academic Publishing, Saarbrücken (2010)

Kiessling, W., Bable, F.: Corporate Identity. Strategie nachhaltiger Unternehmensführung. Ziel- Zentrum F. Interdis, Solothurn (2007)

Luhmann, N.: Einführung in die Systemtheorie, Auflage 5. Carl-Auer-Verlag, Heidelberg (2008)

Luhmann, N.: Organisation und Entscheidung, Auflage 2. VS-Verlag, Wiesbaden (2006)

Nel, R.: Puppets Or People: People And Organisational Development: An Integrated Approach. Juta Academic, Claremont (2008)

Nyiri, A.: Corporate Performance Management. Ein ganzheitlicher Ansatz der Unterneh-menssteuerung. Facultas.wuv, Vienna (2007)

Ossimitz, G.: Entwicklung systemischen Denkens. Theoretische Konzepte und empirische Untersuchungen. Profil Verlag, Munich (2000)

Pullen, A., Beech, N., Sims, D. (eds.): Exploring Identity: Concepts and Methods. Palgrave Macmillan, Basingstoke (2007)

Senge, P.: The Fifth Discipline: The Art & Practice of The Learning Organization (rev. ed.). Crown Business, New York (2006)

Varela, F., Thompson, E., Rosch, E.: Der mittlere Weg der Erkenntnis: die Beziehung von Ich und Welt in der Kognitionswissenschaft - der Brückenschlag zwischen wissen-schaftlicher Theorie und menschlicher Erfahrung. Goldmann Wilhelm, Munich (1995)

Vester, F.: The Art of Interconnected Thinking. Ideas and Tools for tackling complexity. MCB-Verlag, Munich (2007)

Yeo, R.K.: Organisational Development in the 21st Century: Learning for Success - Les-sons from Singapore's Learning Organisations. VDM-Verlag, Saarbrücken (2009)

Using Human Service Center Interfaces and their Information to Foster Innovation Management

Klaus-Peter Faehnrich, Kyrill Meyer, and Benjamin Udo Strehl

University of Leipzig – Research Group Service Science & Technology
Johannisgasse 26, D-04103 Leipzig
{Faehnrich,meyer}@informatik.uni-leipzig.de, b.strehl@usu.de

Abstract. Doubtless innovation is a key for long-term success of any cooperation. However, empirical research indicates limited success of innovation management in practice. Not surprisingly companies are reinforcing their efforts and set a strong focus on customer integration since it has been identified as a major aspect of successful innovation management, current examples being open innovation, integration of social communities and learning communities. Most of these methods are directed towards a special group of customers, the so-called lead-users. This selected integration already led to noticeable progress in some companies and industries, but includes some critical drawbacks, especially since it neglects the majority of a company's human client interactions, which happen in the often centralized service centers, e.g. call centers. Our analysis has identified that these contacts can make a significant contribution to improving a company's innovation management, e.g. through evaluating existing customer-related information. Therefore our research focuses on the adoption of current methods of client integration for a wider audience via service centers. This paper presents the results of expert interviews, case studies and state-of-the art analysis that we conducted in this context. After a discussion of restrictions, benefits and potential results for this form of client interaction, a requirement framework for a support system will be introduced. Subsequently a service blueprint for a support system solution will be presented which was developed based on the requirement framework via quality function deployment in cooperation with reference companies, followed by an overview of further related research questions and information on planned implementations of the support system.

1 Introduction

Numerous authors and studies identified innovation management as one of the major success factors for an organization ([3]; [13]; [28]; [34]; [46]). In particular for a resource-scarce country such as Germany innovations are indispensible ([12]; [45]), since they are the source for competitiveness and prosperity [60]. Therefore, one finds professionalized innovation structures in most mid-to-large-sized companies. One major characteristic of this professionalization is the increased use of information technology.

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 195–204, 2011.

Nevertheless, success factors of innovation management solutions are unsatisfying. Globally just about six percent of all corporations are regarded as effective in usage of innovation management [36]. Former studies reveal that only 0.6 to 2.0 percent of all ideas are successfully introduced to markets. Even those introduced to markets have success rates of between 21 to 26 percent [51]. Moreover, current results even indicate deterioration ([17]; [30]). Though, companies undergo immense efforts through their Customer Relationship Management (CRM) or market research the lack in predictability of market and customer behaviors could not be overcome.

Based on this discrepancy a large variety of practical methods and tools has been developed to actively incorporate customers as well as their information into innovation management activities. The importance of the topic can be seen in the increased innovation budgets, which despite the recent economic crisis have on a global scale risen by 4.2 percent p.a. and by 2.3 percent p.a. in Europe ([1]; [36].

2 Problem Statement

With these new solutions some companies integrate customers as partners into development and exploration activities. Partly, the innovative ability has been improved through this opening of activities. Examples are the incorporation of the wisdom of crowds [55], solutions for requirement engineering [21], collaborative work, computer-aided design toolkits ([44]; [59]), mass collaboration [39] communities of practice or learning communities, idea market places or in general all open innovation approaches ([13]; [38]).

Almost exclusively these solutions are directed towards a special customer group able to cooperate on eye height with companies. This leads to major challenges. First, companies already communicate with many customer groups, including lead-users. Seldom, these customer interactions are used for innovation purposes. As a result current lead-user integration approaches have difficulties with the identification of appropriate customers already in regular contact with the enterprise [53].

Second, this selective approach excludes the majority of customers from innovation activities [56]. Though, customer understanding, especially of the average customer or the masses of demands, and the ability to incorporate customer views determine the later success of products or services ([14]; [57]). Customers that are neglected often can be characterized by one or more of the following qualities: low or no product affinity ([52]), high satisfaction, slight but no strong dissatisfaction, little switching costs [40], little revenue potential per single user [56] and ordinary application range of a company's offer [23]. Hence lead-user focused approaches could have difficulties with its general representativeness, its significance for operative areas such as pricing [23] and limited validity for mass customization markets [56]. Additionally, offers on the mass market become increasingly diverse ([5]; [59]), but the strong focus only on sophisticated customers can hardly support the growing complexity the markets. Consequently, a broadening of active customer integration on a wider, more representative base is of great importance. Also, research findings in this area are seldom translated into successful practical usage which underlines the necessity of more practicable solutions [51].

Service Centers represent a promising starting point for an expansion of the active customer integration. Corporations already bundle their customer interactions in centralized departments, often called customer service center. These centralized units already handle the largest part of customer contacts in many companies [54]. Surprisingly these interactions and the according information are to a large extent not used for innovation activities [11], although it is used for sales or efficiency purposes.

This complex of problems has grown over the last decade. First, time for reproduction of products has declined, leading to shorter amortization periods for new developments ([14]; [43]). Second, markets have matured and therefore a differentiation, specifically for companies from developed markets, can only be achieved through innovative products ([22]; [28]). Third, the recent economic crisis has increased the pressure to cut also innovation costs ([14]; [28]). As a result, companies are confronted with the challenge of faster development cycles, with lower error margins at reduced resources. As a consequence, they focus increasingly on existing resources and information. Therefore also existing customer interactions have to be exploited to the full potential. Interestingly, hardly any appropriate approach, holistic concept or technical solution exists on how existing customer interaction and the according information can be used for innovation purposes [6].

In order to develop a solution for a support system that addresses the shortcomings of the current approaches we assessed the special characteristics of the customer interactions in service centers, especially in the light of their integration into innovation management. Based on the advantages, disadvantages and the circumstances in practice we identified a requirement framework which we later translated into a service blueprint with functionalities that fulfills these requirements. The results and our concrete approach will be presented in the following paragraphs.

3 Background and Methodology

A considerable amount of research has already been conducted for the subjects of innovation management, customer interaction and service center already. Additionally there are numerous interfaces with other close related research areas such as CRM, knowledge management, market research and organizational learning, but also with more distant disciplines like psychology. Our state-of-the art analysis of the existing literature and research [27] identified a large variety of relevant literature and research. Surprisingly a substantial number of recent research, especially quantitative research, exists which deals with different aspects of the topic ([8]; [15]; [19]; [33]; [41]; [53]). This quantitative research ranges from different types of customer integration [37] over motivation factors for participating customers to success of customer ideas and from approaches in innovation management over the variety of IT-solutions [7] to the concrete success of innovative products.

However, no research could be identified that combines these findings in the light of active customer integration in service centers. In order to develop such a holistic understanding and to analyze its special characteristics a detailed analysis is needed which compares the already available existing findings. Moreover, the level of detail in the existing quantitative studies is mostly inappropriate to draw direct conclusions for an application as targeted in our research goal. Therefore, the first focus is on qualitative research which delivers a more detailed perspective [50]. Additionally, due to the explanatory research question, single subject matters are appropriate [61]. The complexity of the topic, the interfaces with other disciplines and the lack of in-depth information underline the necessity of qualitative methods such as expert interviews and case studies to develop a first understanding and first solutions in this area [26]. This approach is based on the grounded theory by Glaser and Strauss [31] in which close to reality theory is developed based on empirical data.

Guideline-based interviews [9] have been conducted with industry experts. By starting with a fixed questionnaire it was assured that all topics were covered in each interview and that answers have a high degree of comparability by at the same time allowing enough freedom to identify new aspects [61]. Starting point for the expert selection was the structure of the German industry. In order to maximize the relevance of our findings even with a smaller number of interviews we chose experts from the four largest German economic sectors representing 92 percent of all employees [29]. Additionally, we choose experts from different companies in each sector ranging from small-sized companies with less than 50m EUR in revenues 2009 to companies with over 5bn EUR in revenues to prevent biased results. Also, we included officers in charge of service centers, innovation management and customer contacts to get a preferably more complete picture. In total fourteen experts from eight companies have been interviewed in the first phase, each in a two hour interview. Results have been documented and confirmed back with the interviewees.

The same approach was used to identify requirements for a support system. Guideline-based interviews were used to identify a list of requirements which was then combined with all requirements identified throughout the state-of-the-art analysis. Then, the single requirements were evaluated by the experts. In particular two characteristics were evaluated, necessity and impact. In the first round each expert had one vote with which an unlimited number of items could be marked as indispensible. Items with a majority vote were labeled as basic requirements. In a second step the additional value provided by the fulfillment of each of the remaining items was estimated. No further breakdown of the term additional value was provided since each expert and company has a different set of values demanded to innovation solutions. Some companies put faster processes in the middle of their attention, some aim at more accurate forecasts and others target resource savings. The individual assessments were aggregated and based on this view we derived three further requirement categories reaching from requirements of high, medium to low priority.

In the next step the method of Quality Function Deployment (QFD) very common in engineering science and quality management was used to derive a set of functions that fulfill the identified requirements ([15]; [32]). By applying the House of Quality, a tool for translating requirements into functionalities [2], we derived in close cooperation with experts from two companies a set of functionalities. Many of the developed functionalities were identified in the state-of-the-art analysis and in the expert interviews. Several iterations have been conducted to arrive at a complete blueprint. The blueprint as the result of this process covers all identified requirements with at least one function. The sufficiency of this coverage shall be evaluated in further research. An overview of the findings will be provided in the next chapter.

4 Results

Out of the vast variety of available literature and the state-of-the-art interviews a list of disadvantages and set of opportunities has been developed. This process and its result have been presented in earlier publications [20]. Potential restrictions can be divided into two categories, restrictions on the customers' and on the companies' side.

An important disadvantage linked to customers is the limited knowledge and experience of customers [55], especially of non lead-users [45], and missing technical language [58] that could result in e.g. communication issues. Additionally, motivating customers to cooperate could be another concern [4]. Moreover, increased expectations of participating customers are a potential thread too ([48]; [59]).

Within corporations one can differentiate between three kinds of barriers, the barriers of not-wanting, not-permitting and not-knowing [18]. In some companies there is a tendency to not accepting external input in general. In other cases employees do not have enough resources to conduct innovation activities ([6], [49]). Also, innovation success is blocked by the inability of internal staff to fulfill related tasks ([18]; [47]) or a lack in existing structures [42].

Nevertheless, the analysis of potential restrictions has neither identified any issue that is ultimately contradicting the research goal nor any issue that in general prevents a successful solution development [20]. However, all concerns have to be addressed by a potential solution. Whether all barriers can be overcome by a solution has to be evaluated by developing, testing and analyzing potential solutions. Besides downfalls there is also a variety of benefits supporting the continuation of the research.

4.1 Appropriateness of Service Center Contacts for Innovation Management

Our analysis revealed a set of detailed aspects that are supporting the capability of a service center's customer contacts to improve the innovation ability of corporations [20]. In total these aspects can be comprised in ten categories. At first, the customer interactions in service centers provide a multiplicity of types of innovation-relevant information [25], the usage and evaluation of this existing information can already be beneficial. The representativeness of these contacts is another benefit [42]. Moreover, the depth of information [16] and its wide spectrum [52] can be beneficial, especially in the important area of customer behavior [24]. Additionally, information retrieval in service centers limits information losses through third parties [57] and can be faster due to existing efficient structures in most service centers [13]. The reoccurring nature of service center contacts allows precious iterations with customers. Interestingly, these customer interactions are also said to provide valuable input for all different innovation types [59] as well as for each step of the innovation process [8].

Based on the potential innovation areas by Ulwick [57] we analyzed whether customer center contacts are able to support different innovation areas. In cooperation with the industry experts each of Ulwick's four areas was analyzed in light of whether available customer center information could provide additional value to support innovation success. Eventually for each category several application areas were identified in which customer center contacts and their information could be beneficial. Table 1 presents some examples of this additional value.

Table 1. Examples for added value by customer center contacts for different innovation areas

Innovation Areas	Examples for Service Center supporting the specific Innovation Area
Product/Service Innovation	Identification of weaknesses in product offerings, e.g. incident reports Development of offer based on weaknesses in consultation with clients
Process Innovation	Identification of gaps in value performance, e.g. customer complaints indicate late delivery, slow response times etc. Improvement in value performance gaps in close consultation with clients
New Market	Recognition of under or overachieved customer requirements through analysis of customer behavior, which allow targeting new customer groups Overachieved aspects signaled by frequent usage problems and customer requests inappropriate usage of products could signal new application areas
Disruptive Innovation	Identification of overfulfilled customer requirements by analyzing usage figures, e.g. features that are not or hardly used and mentioned by customer in any kind of interaction might not be needed in a next product or service

4.2 Requirements Framework

Through the complete process of requirement identification a list of 43 different items has been developed. However, the items were initially not compared or ranked against each other. For this purpose the before described evaluation process was conducted. Based on this assessment, four requirement categories were developed [20]. Basic requirements contain all items that are absolutely essential and can't be missed in any solution. High priority requirements all provide an in average high additional value if added to the solution and are in addition regarded as essential by parts of the expert group. The category of medium priority contains items with medium to high additional value that are not regarded as essential by any expert. In the fourth category, that of low priority, all items with a lower than average additional value are comprised. Figure 1 illustrates this requirements framework. For illustration purposes, we bundled the single items into thematic groups.

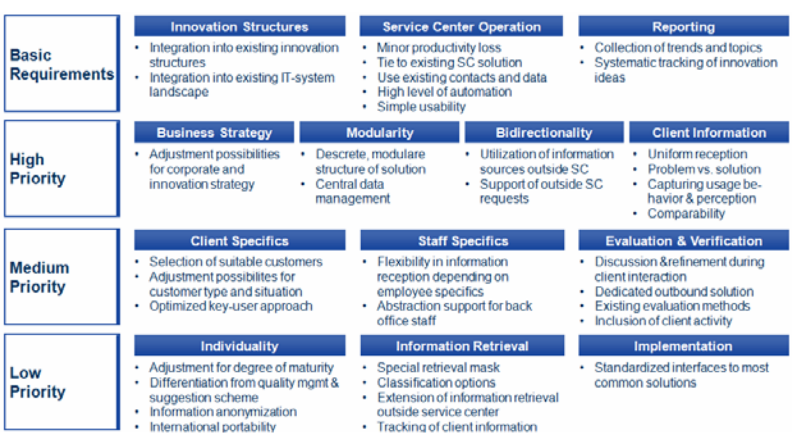


Fig. 1. Set of requirements for integrating service centers into innovation management

4.3 Functionalities for Support Solution

The complete set of requirements served as the starting point for the identification of necessary components and functionalities of any support solution. The method of QFD which was applied differentiates four levels, customer requirements, functionalities, quality characteristics and parts [2]. The House of Quality was used for the first transformation from requirements to concrete functions.

Resulting from the application of the House of Quality was a list of in total 83 single functionalities. The individual functionalities are characterized by a large discrepancy in complexity. Some of these functionalities are likely to be provided with comparable few resources, such items are automated spelling checks, tapping existing data sources and automated creation of data base entry. Other items might present a severe effort towards a successful solution development, examples are the assignment of the appropriate employee, support in abstracting available client information and referencing between various data sets.

The entire set of functionalities was then analyzed for interdependencies. Based upon this analysis the single items were set into chronological and hierarchical order. For this purpose the service system modeling technique by Boettcher [10] was applied. With this technique it is possible to consolidate numerous service items into a single comprising model. The underlying metamodel differentiates four different categories: resources, services, products and processes [10]. As a result the single function items were bundled and put into several task groups comprising several functionalities. Moreover, the eighteen categories were again aggregated to a set of five main services. The first two levels of the overall support model, the main tasks and the service groups, are shown in Figure 2. A specific presentation of each function attached to the task groups is not illustrated due to space constraints.

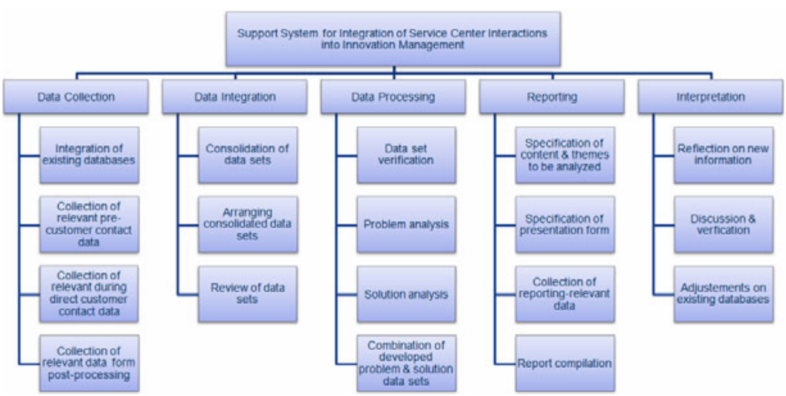


Fig. 2. First two levels of the support system model for service center integration

5 Conclusion and Outlook

Our analysis has shown the necessity of further examining customer interactions within service centers in the light of innovation. These contacts, i.e. the incorporated information, are identified to contain enormous potential to further improve existing innovation solutions. As shown, current market requirements foster effort towards an innovation management that builds on existing resources. Our research identified a variety of requirements and barriers that have to be overcome in order to implement a functioning support system solution in practice. These requirements have been translated to a set of functionalities directed at fulfilling the identified specifications.

In a next step the arranged functionalities have to be put into a complete process for the active customer integration that can be applied in various kinds of service centers. Therefore, each function has to be elaborated in detail, building on the already existing level of detail from the service system modeling. Additionally, the concrete development of an IT-solution is already ongoing. This solution, fulfilling all requirements, will be used to test the validity of both functionalities and requirements. Therefore, the solution will be implemented and tested in the service centers of two companies which will serve as real-life case studies. We are already in the planning phase with these two committed partner companies, one in the professional service sector and one machine construction company. Therefore the support system model will be integrated into the existing process structures and into system landscape of those companies. This procedure might lead to within the existing support system model or maybe to adjustments of the reference companies' internal process structure. In addition, our research will be accompanied by a continuous analysis of restrictions, e.g. of the legal perspective since call centers in particular have been set under strict regulations in Germany [35]. With this procedure a valid statement about the potential of service center customer interactions for innovation purposes should be possible.

References

[1] Accenture, Innovation – a Priority for Growth in the Aftermath of the Downturn. Accenture, Atlanta (2009)
[2] Akao, Y.: Quality Function Deployment. Verlag Moderne Industrie, Landsberg (1992)
[3] Albach, H.: Innovationsstrategien zur Verbesserung der Wettbewerbsfaehigkeit. Zeitschrift fuer Betriebswirtschaft 59(12), 1338–1352 (1989)
[4] Allen, R.C.: Collective invention. Journal of Economic Behavior and Organization 4, 1–24 (1983)
[5] Anderson, C.: The Long Tail. Hyperion, New York (2008)
[6] Auer, S., Faehnrich, K.-P., Riechert, T.: SoftWiki – Agiles Requirements-Engineering fuer Softwareprojekte mit einer großen Anzahl verteilter Stakeholder. GeNeMe' 06 – Gemeinschaft in neuen Medien, Dresden (2006)
[7] Ardilio, A., Auernhammer, K., Kohn, S.: Marktstudie Innovationssysteme - IT-Unterstuetzung im Innovationsmanagement. Fraunhofer IRB Verlag, Stuttgart (2004)
[8] Bessant, J.: High Involvement Innovation: Building and Sustaining Competitive Advantage Through Continuous Change. John Wiley, Chicester (2003)
[9] Bogner, A., Littig, B., Menz, W.: Das Experteninterview – Theorie, Methode, Anwendung, 2nd edn. VS Verlag fuer Sozialwissenschaften, Wiesbaden (2005)
[10] Boettcher, M.: Architektur integrierter Dienstleistungssysteme (Dissertation). University of Leipzig, Leipzig (2009)
[11] Bughin, J., Chui, M., Johnson, B.: The next step in Open Innovation. McKinsey Quarterly, 22–29 (Fall 2008)
[12] Bullinger, H.-J., Auernhammer, K.: Innovationen im Spannungsfeld von Kreativitaet und Planung. In: Warnecke, H.-J., Bullinger, H.-J. (eds.) Kunststueck Innovation, pp. 29–34. Springer, Berlin (2003)

[13] Chesbrough, H.W.: The Era of Open Innovation. MIT Sloan Management Review, 35–41 (Spring 2003)

[14] Christensen, C.M.: The Innovator's Dilemma, 1st edn. Collins, New York (2006)

[15] Cristiano, J.J., Liker, J.K., White, C.C.: Customer-driven product development through quality function deployment in the U.S. and Japan. Journal of Product Innovation Management 17(4), 286–308 (2000)

[16] Den Hertog, P.: Managing Service Innovation. Dialogic, Utrecht (2010)

[17] Duethmann, C.: Marken sind von Dauer. Lebensmittel Zeitung Spezial 2, 8–10 (2008)

[18] Enkel, E.: Chancen und Risiken von Open Innovation. In: Zerfaß, A., Moeslein, K.M. (eds.) Kommunikation als Erfolgsfaktor im Innovationsmanagement, 1st edn., pp. 177–194. Gabler, Wiesbaden (2009)

[19] Ernst, N., Zerfaß, A.: Kommunikation und Innovation in deutschen Unternehmen. In: Zerfaß, A., Moeslein, K.M. (eds.) Kommunikation als Erfolgsfaktor im Innovationsmanagement, 1st edn., pp. 57–83. Gabler, Wiesbaden (2009)

[20] Faehnrich, K.-P., Strehl, B.U.: Anforderungen an ein IT-gestuetztes Kundeninnovationsmanagement im Customer Service Center. In: Faehnrich, K.-P., Franczyk, B. (eds.) Informatik 2010. Service Science, vol. 1, pp. 82–88 (2010)

[21] Faehnrich, K.-P., Meyer, K., Boettcher, M.: Service Engineering IT-basierter Dienstleistungen – neue Perspektive durch ein integriertes Co-Design. IM – Fachzeitschrift fuer Information Management & Consulting 4, 19–25 (2008)

[22] Faehnrich, K.-P., Meiren, T.: Service Engineering. In: Spath, D., Faehnrich, K.-P. (eds.) Advances in Service Innovations, pp. 3–16. Springer, Berlin (2007)

[23] Fetterhoff, T.J., Voelkel, D.: Managing open innovation in biotechnology. Research-Technology Management 49(3), 14–18 (2006)

[24] Fiol, C.M., Lyles, M.: Organizational Learning. Academy of Management Review 10(4), 803–813 (1985)

[25] Franke, N., von Hippel, E.: Satisfying Heterogeneous User Needs via Innovation Toolkits – The Case of Apache Security Software. Research Policy 32(7), 1199–1215 (2003)

[26] Fredberg, T., Elmquist, M., Ollila, S.: Managing Open Innovation – Present Findings and Future Decisions. Chalmers University of Technology, Goeteburg (2008)

[27] Galliers, R.D.: Information Systems Research – issues, methods and practical guidelines. Blackwell Scientific Publications, Oxford (1992)

[28] Gassmann, O., Enkel, E.: Open Innovation. Zeitschrift Fuehrung + Organisation 3, 132–137 (2006)

[29] German Census Bureau. Arbeitsmarktstatistik 2009 (2010), `http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Grafiken/Arbeitsmarkt/Diagramme/Erwerbsbereiche,templateId=renderPrint.psml` (retrieved May 10, 2010)

[30] GFK, ConsumerScan Innovation Day: Launches und Relaunches als Motor der Wertschoepfung – Was ist Top, was ist Flop?. GFK, Nuernberg (2006)

[31] Glaser, B.G., Strauss, A.L.: The Discovery of Grounded Theory – Strategies for Qualitative Research. Aldine Transactions, New York (1967)

[32] Govers, C.P.M.: What and how about Quality Function Deployment (QFD). International Journal of Production Economics, 46–47, 575–585 (1996)

[33] Gruner, K., Homburg, C.: Does Customer Interaction Enhance New Product Success. Journal of Business Research 49(1), 1–14 (2000)

[34] Hauschildt, J.: Innovationsmanagement, 3rd edn. Vahlen, Munich (2004)

[35] Huettenegger, G.: Open Source Knowledge Management. Springer, Berlin (2006)

[36] Jaruzelski, B., Dehoff, K., Bordia, R.: Smart Spenders – The global innovation 1000. Strategy and Business 45, 46–61 (2006)

[37] Jokisch, M.: Active integration of users into the innovation process of a manufacturer – The BMW Customer Innovation Lab. Dr. Hut, Munich (2007)
[38] Kok, R.A., Hillebrand, W.B., Biemans, W.G.: What makes Product Development Market Oriented? Towards a Conceptual Framework. International Journal of Innovation Management 7(2), 137–162 (2003)
[39] Libert, B., Spector, J.: We are smarter than me. Wharton School Publishing, Upper Saddle River (2007)
[40] Luethje, C.: Characteristics of Innovating Users in Consumer Goods Field. Technovation 24(9), 683–695 (2004)
[41] Matthing, J., Sanden, B., Edvardsson, B.: New service development. International Journal of Service Industry Management 15(5), 479–498 (2004)
[42] Meffert, H.: Marketing, 7th edn. Gabler, Wiesbaden (1986)
[43] Meyer, K., Faehnrich, K.-P.: How to Engineer IT-enabled Services. In: Proceedings of First International Symposium on Service Science, pp. 137–148. Logos, Berlin (2009)
[44] Moschella, D.: Customer-driven IT. Harvard Business School Press, Boston (2003)
[45] Opitz, M.: Organisation integrierter Dienstleistungssysteme. Gabler, Wiesbaden (2009)
[46] Pearson, A.E.: Tough-minded ways to get innovative. Harvard Business Review 66, 99–106 (1988)
[47] Read, B.B.: Designing the Best Call Center for Your Business – a complete Guide for Location. In: Service, Staffing, and Outsourcing, 2nd edn. CMP, San Francisco (2005)
[48] Reichwald, R., Piller, F.: Interaktive Wertschoepfung - Open Innovation, Individualisierung und neue Formen der Arbeitsteilung, 2nd edn. Gabler, Wiesbaden (2009)
[49] Schumacher, J., Meyer, M.: Customer Relationship Management strukturiert dargestellt – Prozesse, Systeme, Technologien. Springer, Heidelberg (2004)
[50] Schwanninger, M.: Zur Zukunft der systemorientierten Managementforschung. Institut fuer Betriebswirtschaft, St. Gallen (1989)
[51] Schweisser, W., Mohnkopf, H., Hartmann, M., Metze, G.: Innovationserfolgsrechnung. Springer, Berlin (2008)
[52] Sohn, G.: Call Center World – Mangelhaftes Management der Kundenkontakte (2007), http://www.pressetext.at/news/070122020/call-center-world-mangelhaftes-management-der-kundenkontakte/ (retrieved September 10, 2010)
[53] Spann, M., Ernst, H., Skiera, B., Soll, J.H.: Identification of Lead Users for Consumer Products via Virtual Stock Markets. Journal of Production Innovation Management 26(3), 322–335 (2009)
[54] Surowiecki, J.: The Wisdom of Crowds. Abacus, London (2005)
[55] Tapscott, D., Williams, A.D.: Wikinomics: How Mass Collaboration Changes Everything, 1st edn. Penguin Group, New York (2008)
[56] Tidd, J., Bessant, J., Pavitt, K.: Managing Innovation, 3rd edn. John Wiley & Sons, Chichester (2005)
[57] Ulwick, A.: What Customers Want: Using Outcome-Driven Innovation to Create Breakthrough Products and Services. Mc Graw-Hill, New York (2005)
[58] Ulwick, T., Lawer, C.: Avoid the risks of customer-driven innovation. In: OMC (eds.) OMC Group Insight, Buckingham, pp. 1–5 (2007)
[59] Von Hippel, E.: Democratizing Innovation. MIT Press, Cambridge (2006)
[60] Warnecke, H.-J.: Innovationen in Technik und Gesellschaft. In: Warnecke, H.-J., Bullinger, H.-J. (eds.) Kunststueck Innovation, pp. 1–10. Springer, Berlin (2003)
[61] Yin, R.K.: Case Study Research – Design & Methods, 4th edn. Sage Publications, Thousand Oaks (2009)

Key Features of Subject-Oriented Modeling and Organizational Deployment Tools

Albert Fleischmann[1] and Christian Stary[2]

[1] Metasonic, Germany
[2] Universität Linz, Austria
Albert.Fleischmann@Metasonic.de,
Christian.Stary@JKU.at

Abstract. Subject-oriented Business Process Management (S-BPM) has been used in organizational and workflow development projects for several years. While the methodological support for the development process is important, the success of a project taking the subject-driven approach to organizational deployment also depends on the tool. The tool requires a set of key features, in particular an appropriate level of abstraction for behavior modeling, the representation of work-relevant relations among stakeholders, and the straightforward execution of business process specifications for participatory organization design. In this paper we analyze Metasonic, the most prominent tool for S-BPM, with respect to effective organizational change and interactive workflow design support. Metasonic integrates various BPM modeling constructs, and is suitable for the construction of stakeholder-oriented models and the subsequent deployment of business processes. However, the underlying organizational development activities should become more transparent when utilizing the tool features.

1 Introduction

Subject-oriented Business Process Management (S-BPM) has been practiced now for some years. It is developed continuously, both from a methodological, and tool-support perspective (Buchwald et al., 2010, Fleischmann et al., 2011). Overall, it allows for interactive structural flexibility of organizations, (re-)organizing internal procedures as well as arranging networks of organizations, based on business process specifications.

Various concepts have been proposed for organizational change management and learning (cf. Heftberger et al., 2004). The findings reveal two key elements: (i) the representation of (task) knowledge (according to individual mental models), and (ii) on its communication along collective reflection and processing. Hence, knowledge about work and its organization needs to be located, elicited, acquired, and explored (cf. Nonaka et al., 1999, Davenport, 1998, Senge, 1990) before being deployed in organizational settings.

Since individual learning has to be considered both, the starting and reflection point of Organizational Learning (OL) processes, stakeholders play an important role for triggering and performing those processes. Ideally, they should be drivers and controllers of change processes. However, there are few studies about stakeholder

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 205 – 214, 2011.

needs, and socio-methodological or –technological capabilities required for OL. When Herrmann has tried to qualify employees by letting them develop parts of business processes individually, he used a semi-formal workflow modeling language and a respective support tool (Herrmann, 2000). It allows stakeholders describing their particular view on work situations and tasks by means of a diagrammatic language, including individual procedures for task accomplishment.

S-BPM aims one step further: It not only should allow stakeholders describing work procedures in a diagrammatic way, but also executing their specifications in a non-distractive way (Fleischmann, 2010). However, it has still to be investigated, in how far S-BPM tool support facilitates organizational deployment in this way, since stakeholders have to explicate their mental models of task accomplishment (cf. Rouse, 2006). These models form relevant context when envisioning changes and designing organizations, and need to be shared, once being explicated (cf. Mehandijev et al., 2010). Consequently, we will follow an OL framework to investigate the effectiveness of key features of the S-BPM suite Metasonic (www.metasonic.de) concerning subject-oriented business process modelling and deployment.

We introduce the OL framework in section 2 and identify respective features categories. In section 3 we detail and reflect existing tool support. Section 4 concludes the paper, wrapping up the results and sketching future developments.

2 Deriving Requirements for Tool Support from Organizational Learning Design

In this section we describe how stakeholders trigger learning processes when they re-think and (re-)design their work (section 2.1) before developing stakeholder requirements for tool support (section 2.2).

2.1 Stakeholder-Centred Organizational Learning

For intertwining individual knowledge creation and organizational learning processes we follow the experiential learning cycle as detailed in Stary (2011) based on Heftberger et al. (2004). As shown in figure 1 (left side) individual knowledge creation serves as input to organizational learning processes. Individuals observe stimuli and their consequences from the environment. These observations are assessed partly consciously, partly unconsciously. On the basis of these assessments individuals form abstract concepts to react to the stimuli in a reflected way (termed *design*). The developed concepts are implemented in actual work situations. Recurrent observation and perception of the concepts in terms of effectiveness lead to iterations of the individual activities.

In the design phase stakeholders express their (role-)specific view onto information structures, work settings, critical situations, or business processes according to their task assignments and individual experiences. Hereby, an organizational learning step might be initiated. The resulting information structure or workflow specification can be embedded into actual work situations (*implement*). Interactive artefacts executing specifications, such as the S-BPM suite, enable hands-on-experience for task accomplishment. Stakeholders observe, in particular when interactive artefacts are used, possible effects the executed tasks have on their work and the organization. In case the results meet individual expectations or individual demands for change (*assess*), the concerned process serves as input for the learning process on the organizational level. If further process refinements or modifications are required the cycle starts again (*re-design*).

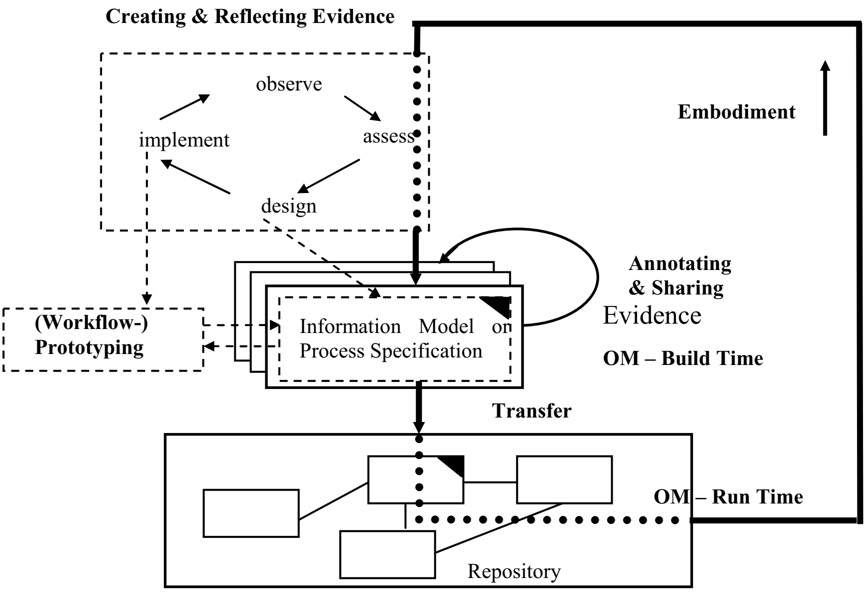


Fig. 1. The operational frame of reference for stakeholder-centred OL (see also Stary, 2011)

Figure 1 shows the fundamental structure, activities, components, and relationships not only on the individual, but also on the organizational level. The trigger to OL is creating evidence, i.e. information a stakeholder considers being of relevance for further organizational development with respect to his/her role and consequently, the organization. Once modelled, the evidence can be refined, modified or enriched through sharing different perspectives until a decision is made how to proceed on the organizational level.

The directed link from the individualized business information or process model refers to the entire Organizational Memory (OM), since all previous entries of the OM might be affected. The links between the information / business process model and (workflow) prototyping mean that experiences gained from using any interactive artefact are likely to influence the adaptation of content or business processes specifications, and vice versa, modified information structures and processes are likely to result in an alternative accomplishment of tasks. The organizational learning cycle might be iterated as soon as stakeholders have embodied the newly acquired and specified knowledge on the level of the overall organization.

2.2 Requirements for Tool Support

OL can be considered as a swinging pendulum between individuals and affected members of an organization, providing and exchanging individual or group perspectives on created information (structures). A learning step is considered complete (see also figure 1), once a modified information or business process model represents a commonly agreed basis for task accomplishment, and becomes effective on the organizational level through embodiment by its stakeholders.

Transforming individually elicited knowledge to become effective on the organizational level comprises several activities, with respect to modelling and deployment (cf. Chen et al., 2003):

- *Expressing work knowledge.* Basically, all created evidence, either in form of information structures or business process specifications, needs to be documented by its proposer(s) or a facilitator. It can then become available in an OM accessible to all stakeholders. Once stored, other stakeholders can express individual concerns and formulate additional inputs.
- *Experience process specifications.* Executing specifications helps visualizing and experiencing proposed changes in a straight-forward way. It also might initiate and focus discussions, leading to further modifications. Upon agreement novel content and/or processes can become effective on the organizational level.
- *Embody novel structures into the organization.* In order to become operative, process specifications need to be implemented in daily business. Behaviour specifications need to be integrated in the operation of the organization. They are embodied in individual work practices within the context of business processes.

All addressed bundles of activities need to be supported by a business process modelling and deployment tool.

3 Tool Support

In this section we revisit the S-BPM approach and its major support suite Metasonic (see also Fleischmann et al., 2011) by discussing each of the derived categories of requirements:

- Creating models from an individual stakeholder perspective, in order to allow individually expressing work knowledge (section 3.1)
- Ensuring automated execution, in order to support immediate experience of specifications, and to adjust individual inputs to existing behaviour sequences, leading to different (interactive) variants for task accomplishment (section 3.2)
- Keeping the history of individual and organizational developments, in order to embody the finally accepted variant into the organization (section 3.2).

3.1 Expressing Work Knowledge via Process Modeling

Business process models are assumed to facilitate the transfer of knowledge between individuals or groups with different attitudes and presuppositions, as they are boundary objects (Brown et al., 1998). According to their involvement in accomplishing work tasks, different stakeholders can be represented in a business-process model. For managing change business process models help different stakeholders and groups of an organization developing and propagating a common interpretation of accomplishing tasks. As such, they can be considered as trigger and enabler for individual and organizational learning processes.

In order to allow the various stakeholders and management to participate in OL processes through work process modelling and deployment S-BPM provides a notation and specification features that

- are capable of describing the organization of work from the perspective of each involved actor or role - in S-BPM stakeholders are represented by so called subjects and vice versa, since subjects are the active elements in sentences of all natural languages. Consequently, subjects are abstractions of concrete roles, agents or actors.
- focus on the interaction of subjects – in S-BPM task-relevant communication is described by messages exchanged between subjects.
- enforce the adjustment of subject behaviour specifications to create a complete representation of a business process – in S-BPM the behaviour specifications of all subjects involved to accomplish a certain task have to be transparent and mutually tuned.
- integrate business object representations into the flow of work – in S-BPM these objects are addressed and detailed in terms of their structure when messages are exchanged between subjects (see for instance, vacation request in figure 4).
- enable the seamless execution of business process specifications ensuring the direct interactive experience of work process models (see section 3.2)

For organizational development both levels need to be addressed by a tool, the individual and (cross-)organizational one, in order to get a complete understanding of the business processes and their anticipated advancement.

In figure 2 a sample subject-oriented model is shown. The screenshot contains the subjects involved in a process handling vacation requests of employees, and the

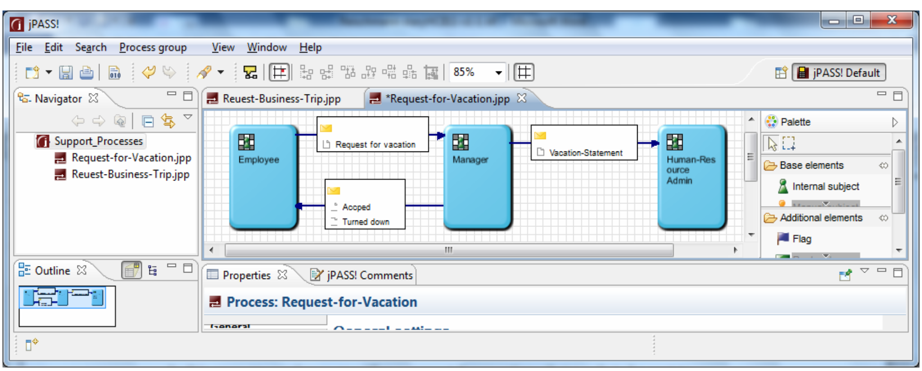


Fig. 2. Sample subject constellation

message types exchanged between these subjects. In that example three stakeholders ('Employee', 'Manager', 'Human Resource Admin') coordinate their work by exchanging messages.

Subjects execute three different types of activities: (i) Subjects send messages to other subjects, (ii) subjects receive messages from other subjects, and (iii) subjects execute some internal actions not visible externally. These activities are done in an order agreed between the stakeholders representing the various subjects. Figure 3 shows the behavior description of the subject 'Employee', sending messages to other subjects (i), and receiving messages from other subjects (ii), after filling in the vacation request form (iii).

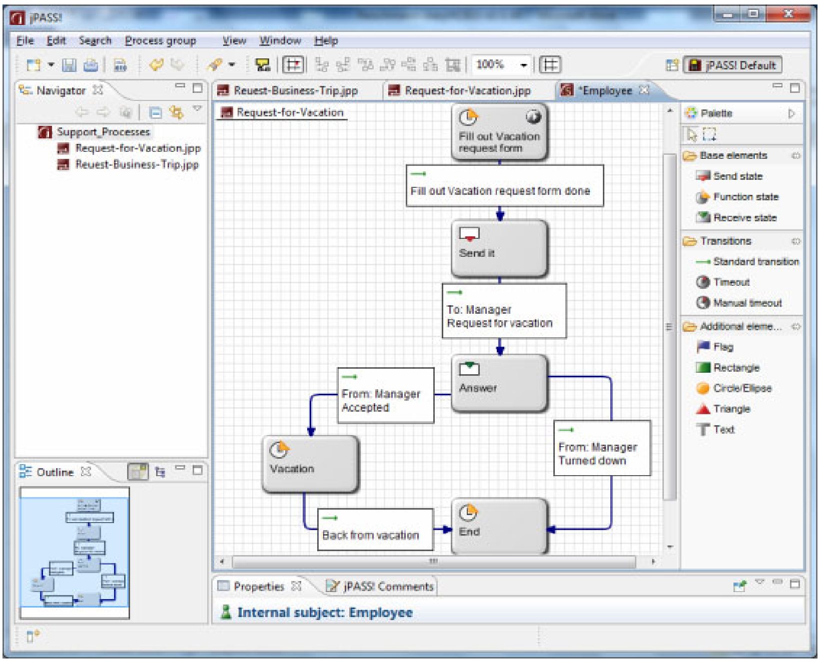


Fig. 3. Sample subject behavior ('Employee')

When an employee files a request for vacation, the request needs to be sent to the manager for approval. As S-BPM integrates the communication-oriented perspective with the (traditional) function-oriented one, both, message exchanges, *and* individual task activities need to be represented. In the example, sending the request refers to a typical send state, whereas receiving the answer is a typical receive state (explanation of symbols see right tool box in the screen shot of figure 3). In contrast, being on vacation is a function state in S-BPM. Of particular importance are the start and the end state ('Filling out vacation request form' and 'End' in the example), as they set the scope of the behavior description for a certain subject.

Depending on the content of the receive state 'Answer', different procedures lead to process completion, either to 'Vacation' or to 'End'. Finally, a behavior specification is enriched with state transitions, such as 'from manager accepted', in order to understand the trigger for the next state.

In the course of modeling messages need to be detailed, as they contain all task-relevant data, and thus, constitute business objects. In figure 4 the data structure of a vacation request is displayed. It is composed of properties detailing the request in terms of name, dates, and type of vacation.

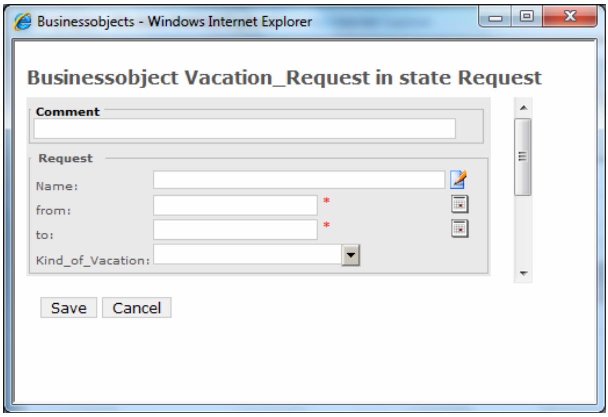


Fig. 4. Sample message details

However, (re-)thinking organizations in terms of communicating actors accomplishing tasks is not sufficient for deployment. It requires the recognition of an organization as a network of communicating actors – each actor is directly or indirectly connected to all others. Hence, lines of communications need to be defined for representing task accomplishment once different subjects are involved in a business process. A specification is considered complete and reflects the organizational perspective, once all involved subjects have been represented and their interfaces allow a coherent flow of interaction for task accomplishment.

Subjects in one process can communicate with subjects in other processes. In case subjects communicate with subjects involved in other processes, the subjects of the other processes are called external subjects. External subjects represent a process connected to the considered process.

3.2 Executing Models Enabling Interactive Deployment

Once process models can be executed, e.g., using a workflow system, subject-oriented process specifications can directly be experienced. They serve for reflection and negotiation among stakeholders in the course of change management. As it happens before actual changes are going to be implemented variants can be tested without effecting implementation, transformation, or development costs.

Up to now, only an abstract process model with subjects has been described. In addition to modeling, Metasonic allows for interactive deployment. In a next step it requires assigning concrete stakeholders to the identified subjects. Figure 5 exemplifies such an assignment: It is either Christian Stary or Albert Fleischmann who could execute the behavior of subject employee.

Once concrete stakeholders have been assigned they can follow the flow of work, as the modeling language and the corresponding representation scheme allow executing specifications without further transformation(s). The individual perspective of each stakeholder remains visible in the course of execution, however, embedded into the (required) organizational runtime context (i.e. the behavior of other involved stakeholders like 'Manager' and 'Human Resource Admin' for 'Employee'). Consequently, each involved stakeholder can interact with the workflow component of the Metasonic suite as if accomplishing a certain work task in daily business.

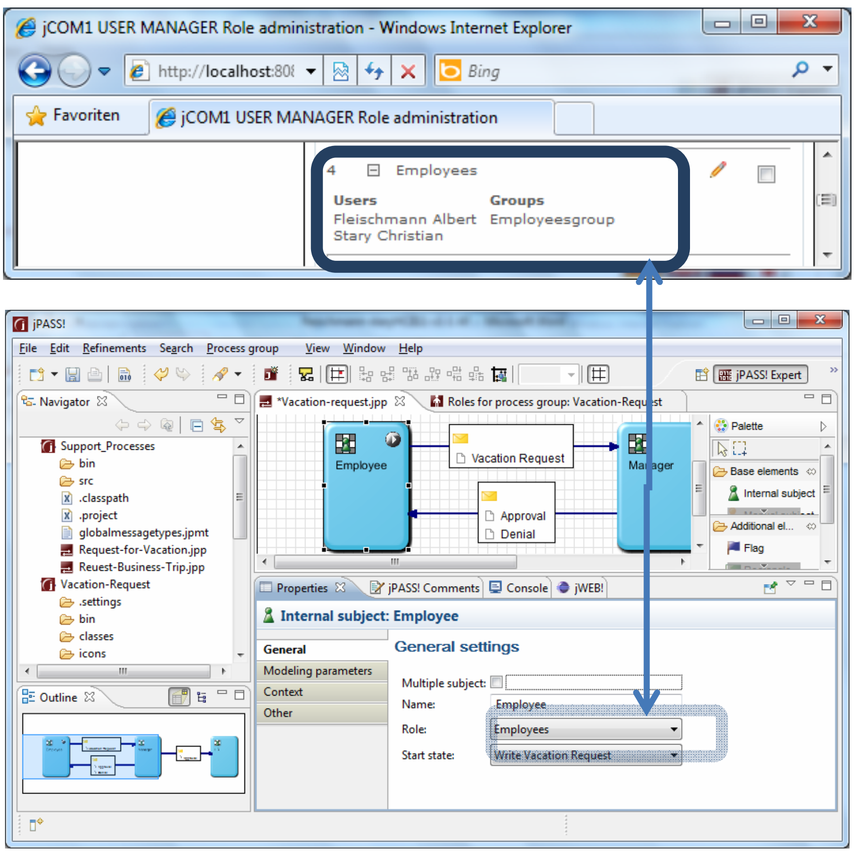


Fig. 5. Sample assignment of subjects to actual stakeholders

The swimlane view given in figure 6 allows tracing the execution according to the interaction perspective of all involved parties or systems. In this way, the organizational perspective is captured by the tool. Each swimlane corresponds to a single subject, such as to 'Employee', 'Manager', and 'Human Resource Admin'.

Moreover, the swimlane presentation of a process execution allows controlling the execution of a process, and analyzing the timeline for potential improvements. Consequently, a business process can be re-designed with respect to its effectiveness and efficiency before coming into operation. Of particular interest is the recognition of bottlenecks, which become visible through specific communication patterns.

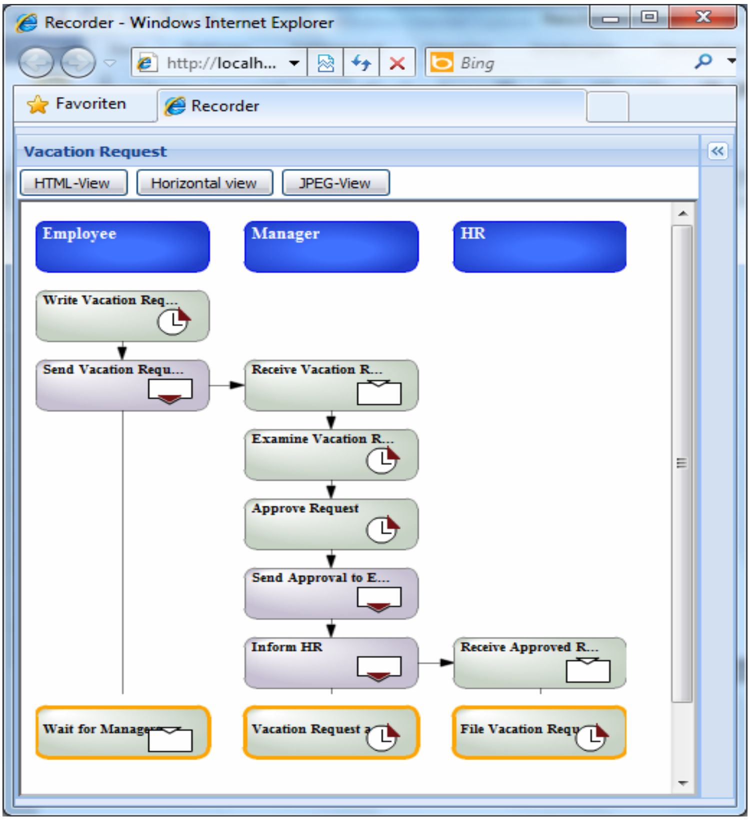


Fig. 6. Ensuring the organizational perspective

As each version can be stored separately, and re-executed at any time, Metasonic also provides some kind of OM. However, it would help to store (other) stakeholder inputs, such as annotations in text form, together with each variant of a process model according to the structure of the OL framework. Such a memory would allow tracing each organizational learning step, still focusing on process specifications when negotiating changes on the organizational level (cf. Stary, 2011).

4 Conclusion

Although Subject-oriented Business Process Modeling (S-BPM) has been used in organizational and workflow development projects for several years, its capabilities have not been explored accordingly. We have reviewed the methodological support for the development process, and the respective tool support when taking the subject-driven approach to organizational deployment.

The requirements for tool support, such as the appropriate level of task description for stakeholders, have been derived from an organizational learning framework. Metasonic, the most prominent tool for S-BPM and participatory organizational development, has been revisited with respect to a number of specific features required to achieve organizational change including workflow support.

While supporting the integration of communication-oriented modeling concepts with function-oriented ones, tool support needs to be more specific in each step of organizational development for stakeholders. On one hand, the various phases in organizational development should frame the individual modeling and deployment activities. On the other hand, active support of collective reflection when developing alternatives and reflecting proposed changes could comprise annotations of models, e.g., to capture comments from stakeholders. We envision some kind of OL-wizard to guide stakeholders in modeling, reflecting, and deploying alternative ways of task accomplishment.

References

1. Brown, J., Duguid, P.: Organizing Knowledge. California Management Review 40(1), 90–111 (1998)
2. Buchwald, H., Fleischmann, A., Seese, D., Stary, C. (eds.): Setting the Stage for Subject-Oriented Business Process Management, S-BPM ONE 2009. CCIS, vol. 85. Springer, Berlin (2010)
3. Davenport, T., Prusak, L.: Wenn Ihr Unternehmen wüßte, was es alles weiß. Verlag Moderne Industrie, Landsberg, Lech (1998)
4. Fleischmann, A.: What is S-BPM? In: Buchwald, H., et al. (eds.), pp. 85–106 (2010)
5. Fleischmann, A., Stary, C.: Whom to Talk to? A Stakeholder Perspective on Business Process Development. In: UAIS. Springer, Heidelberg (2011) (in press)
6. Fleischmann, A., Obermeier, S., Schmidt, W., Stary, C.: BPM 2.0. Subjekt-orientiertes Geschäftsprozessmanagement. Hanser, München (2011)
7. Chen, J.Q., Lee, T.E., Zhang, R., Zhang, Y.J.: System Requirements for Organizational Learning. Communications of the ACM 46(12), 73–78 (2003)
8. Heftberger, S., Stary, C.: Partizipatives organisationales Lernen – Ein prozessorientierter Ansatz. DUV, Wiesbaden (2004)
9. Herrmann, T.: Lernendes Workflow. In: Verbesserung von Geschäftsprozessen mit flexiblen Workflow-Management-Systemen (Band 4). Physica-Verlag, Heidelberg (2000)
10. Mehandjiev, N., Greven, P. (eds.): Dynamic Business Process Formation for Instant Virtual Enterprises. Springer, London (2010)
11. Nonaka, I., Takeuchi, H.: The Knowledge-Creating Company. Oxford Univ. Press, New York (1995)
12. Rouse, W.B. (ed.): Enterprise Transformation: Understanding and Enabling Fundamental Change. Wiley, Hoboken (2006)
13. Senge, P.: The Fifth Discipline, The Art and Practice of the Learning Organization. Doubleday, New York (1990)
14. Stary, C.: Perspective Giving – Perspective Taking: Evidence-based Learning in Organizations. Int. Journal of Information and Knowledge Management 10 (2011)

The Effect of a GPS on Learning with Regards to Performance and Communication in Municipal Crisis Response

Helena Granlund[1,2], Rego Granlund[1], and Nils Dahlbäck[3]

[1] Santa Anna IT Research Institute, Linköping, Sweden
[2] Swedish Defence Research Agency, Linköping, Sweden
[3] Linköpings Universitet, Linköping, Sweden
rego.granlund@santaanna.se, helena.granlund@santaanna.se, nilda@ida.liu.se

Abstract. This paper describes the basic learning aspects of crises management training in a computer simulated environment. A total of 108 professionals, 18 teams, contributed to the study where the impact of a GPS on C2 work was investigated. A comparison between professional groups on performance and learning show that the GPS has an impact that differed depending on the teams professional composition.

Keywords: Experiential learning, Microworld, Simulation.

1 Introduction

This paper present how a computer based simulation can be used to analyze the impact of a global position system (GPS) on command and control teams in terms of effects on performance. 18 Swedish municipal crisis management teams, a total of 108 professionals, participated in the project (Granlund et al, 2010). The key motive for the study was to gain knowledge on how teams differ depending on if they have GPS or if they have regular paper maps in their command post. The study was not designed to explore or optimize training, and had a passive attitude to perform training; meaning that training could be accomplished by the participating members themselves. The paper presents the learning approach, the training procedure, its underlying educational thoughts and a comparison between rescue service and municipal personnel's performance and communication.

1.1 Study Context

In Sweden the municipal organizations are accountable for crisis response within the borders of the municipality. The response preparedness has many aspects, including investments in technical support and training for the local crisis management organization.

The municipalities invest in information and communication systems, in order to gain performance and control in their work. Systems such as GPS support for

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 215–224, 2011.

command and control applications are often integrated with other technologies like multimodal communications and geographical information systems. They have real-time sensors for the units' position and state. They present their information unfiltered, without human intervention, strait in to the management level. Researchers in for instance computer supported cooperative work (Schmidt & Bannon, 1992) distributed cognition (Hutchins, 1995) and cognitive systems engineering (Hollnagel & Woods, 2005) emphasize the importance of evaluating the effect of new systems in practice. It is essential when organizations invest in support systems for command and control to take into account what the overall system, commanders, support and environment, does and performs (Hollnagel & Woods, 2005).

Lessons learned from empirical evaluations in practice can be feed back into the total command and control system during preparedness training. To empirically evaluate events and human experience in situations where geographically distributed commanders use technological systems under heavy workload is however difficult. Real crisis response operations are rarely reviewed in sufficient detail to gain insight for preparedness purposes. Simulations are one possible way to confront and analyze these situations and systems. The simulations that are relevant in this context are scenario-based simulations executed in real-time, and where the development of the tasks can be described as dynamic (Brehmer, 2005).

1.2 The Simulation Environment

The C3Fire simulation environment, used for the study, is designed for evaluation of command and control systems and can be understood as a micro world. Microworlds have complex, dynamic and opaque characteristics that represent the environments people encounter in real-life systems. In microworlds the system designer isolates chosen characteristics from the real world in order to study their effects on teams in a controlled manner (Brehmer, 2005; Brehmer and Dörner, 1993; Granlund, 2001). The system allows controlled studies of collaboration, decision-making, cultural differences in teamwork and effects concerning work processes and information communication tools in command and control (Artman and Wearn, 1999; Granlund et al., 2010; Johansson et al., 2010; Lindgren & Smith, 2006).

In this study C3Fire generated a dynamic forest fire fighting. Six management members needed to collaborate in a team. Their goal was to manage and extinguish forest fires and to protect homes and valuable areas. In order to handle the response the team needed to prioritize between different objectives and identify critical areas. They needed to create a plan and implement operations. All work was distributed, which means that the participants needed to exchange information within the group to execute the task although some participants were separately located. Organization, communication structure, resources, the participants' information systems and simulation environment are examples of properties that were configured in C3Fire to create appropriate learning and research scenarios.

2 Training Method

Members of the municipal crisis management organizations need training to handle a variety of costly, risky and often rare crises. Computer simulations is a means for the training. They offer opportunities to repeat the training process, which in itself is a prerequisite for learning when learning is seen as a continuous process (Kolb, 1984).

2.1 Experiential Learning for Team Training

Kolbs (1984) generally accepted model of experiential learning can be adapted for research and team training performed with computer-based simulations (Granlund, 2008). The two main components of the model are the four step learning cycle (Figure 1) and the idea of learning as a continuous process.

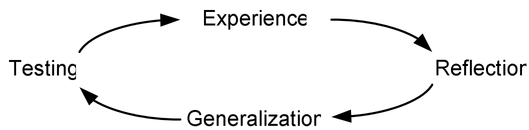


Fig. 1. Kolb's basic model of experiential learning

According to Granlund (2008) having an *experience* initiates the learning. For computer-based simulations, the instructor tries to direct the participant's experience towards the objectives of the education. The participants will have different experiences during the simulation depending on their previous knowledge and experience. During the *reflection* all participants communicate their own experiences from the simulation as well as listen to the other participants' in a shared reflection. The *generalization* takes place when thoughts from the experience and the reflection are linked to the participant's initial knowledge and is performed within the participant. The participants will not do the same generalization, as their knowledge and experience differs, but their shared reflection increases the ability to generalize in the same direction. *Testing* is the last step of the experiential learning cycle. The ideas from the generalization are validated. If the ideas hold they might be included in the participant's knowledge.

By repeating this process the continuous learning of each individual participant in the group will evolve a joint knowledge as they will influence each other during their shared activities throughout the cycles.

2.2 The Study's Utilization of the Four Step Learning Cycle

The study was designed to meet the conditions for experiential learning and the idea that learning is a continuous process. During the training day the participants goes through different levels of learning. To support these levels in the C3Fire environment a training day is performed by a sequence of simulation sessions (Figure 2).

Each of the 5 cycles consist of; a 20 minutes C3Fire simulation trial, 5 minutes of individual questionnaires and then 15 minutes after action review were the whole group is active. The total experiment last for about five hours.

During the simulation trial the participants made experiences thru the team work in the simulated emergency response task. While performing the individual questionnaires the participants had possibility to do personal reflections on the

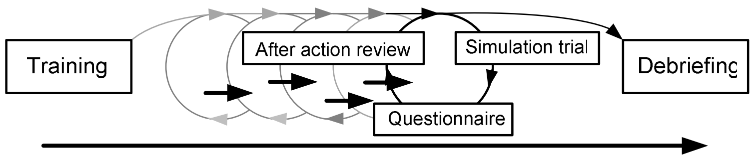


Fig. 2. The experiment procedure with 5 simulation sessions

experience. During the after action review the participants' saw a fast played recording of their latest trial. They performed a shared reflection where they discussed their performance together and made tactical plans for the next session trial. This phase is very intense and takes at least 15 minutes, which can be compared to a simulation trial that takes 20 minutes (Figure 2).

2.3 The Study's Utilization of Learning as a Continuous Process

There are several issues that continuously can be learned during a simulation based team task in C3Fire. Offerman and Spiros (2001) listed eleven common problems team developers deal with regarding team training. Most of which are general for team training and liable also for preparedness training. With the above described cyclic procedure these issues can be learned by the team or the individual in different modes of control; Scrambled, Opportunistic, Tactical (unattended or attended) or strategic. The scrambled mode is defined by no planning, no reflection and actions are ad hoc. In the opportunistic mode behavior is reactive due to lack of understanding and time. The tactical mode is more organized and driven by known rules and procedures and planning is proactive. In the strategic mode actions are not only proactive they also approach higher level goals in an effective manner (Hollnagel & Woods 2005).

Three levels of learning will occur while using the C3Fire environment; individual level, organizational level and task level.

Individual Level: Initially the participants learn to operate the system and the basics of their emergency response task. It is defined by a scrambled control mode. The strength of using a microworld for this phase is that the procedure with repeated simulations assures it to be short. The participants learn to operate the system and understand the task within one trial.

Organizational Level: Next, the group learns to collaborate, they allocate responsibilities and roles and they investigate the systems limits. After this phase, normally 1 or 2 trials, the group has learned the system, understands the task and is ready to act as a team. In this phase the participants should be able to use an opportunistic and in some parts an attended tactical control mode.

Task Level: In the third step the group they can train or learn some specific task. In this phase the participants should be able to uses some type of attended or unattended tactical control mode in some parts of the control.

3 Method

With a series of experiments in a computer simulated crisis environment a total of 108 members from 18 municipal organizations were tested. Each team consisted of six participants, three worked as crisis managers in a command post, three worked as ground chiefs. The command post worked on an operational level, without direct access to the simulation and controlled the simulated world indirectly, through commanding the ground chiefs, who managed three fire brigades each in the simulation (Figure 3 and 4).

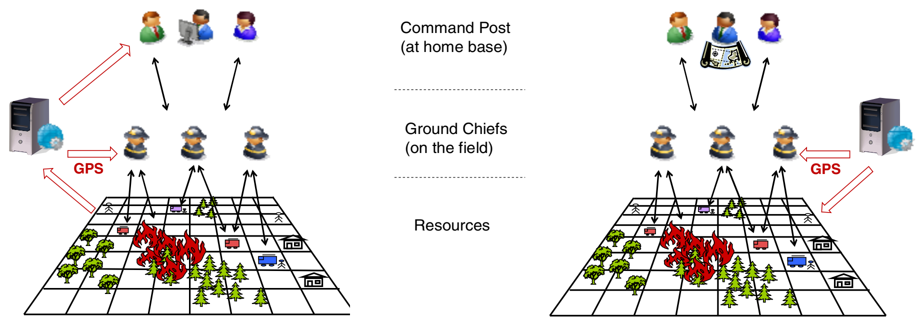


Fig. 3. The GPS condition

Fig. 4. The Paper Map condition

The study had a between-group design with two factors: (a) Teams with professionally homogeneous or heterogeneous command posts, and (b) Decision support in the form of GPS or paper maps. In the professionally homogeneous command posts all three participants was rescue service personnel. These teams are called RSCP, rescue service command post. In professionally heterogeneous command posts the three participants were a mixture of rescue service personnel and other municipal employee groups. These participant groups are called MCP, mixed command post.

4 Learning Effects Visible in the Results

Some of the learning effects identified in the results are related to performance, resource usage, time to first engagement and the participants own view of the perceived training.

4.1 Performance

The main task for the teams is to stop the forest fire and save houses. Normally the participants learn how to solve the task during the simulation trials and perform better at the concluding trial compared to the starting trial. The measure of the success and performance of the team, discussed here, is a measure of the total amount of burned down area at the end of each simulation, BurnedOutArea. Figure 5 shows the average result from each of the five simulation trials. A small amount of BurnedOutArea is preferable to a large.

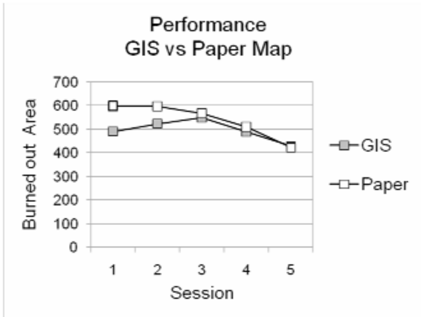


Fig. 5. Amount of burned-out area at the end of each simulation trial

The result shows that the teams with GPS perform better during the initial trials. For the concluding trials the teams with paper maps solve the task equally well as the teams with GPS. This result did not follow the expectations presumed for the study. In a previous study with students as participants the GPS supported teams performed better than the paper map teams in all the five simulation trials (Johansson et al, 2010).

4.2 Team Composition Effects on Team Learning

When the results were divided into the two team compositions, RSCP and MCP, another result appeared. The result show that the *RSCP* teams with GPS perform better in all trials (Figure 6). This result follows the presumed expectations for the study. The results also show that the RSCP teams achieved the expected learning curve during the day.

The result for the *MCP* teams was unexpected (Figure 7). The results show that the teams with GPS do not perform better than the teams with paper maps. The results also show that the teams with GPS have no learning curve. They have the same performance in the last trial as in the first.

The result for GPS setting shows that RSCP and MCP are not equal with regards to performance. The RSCP have significant, t (6) = 4:20, p <.006, less burned out area than the MCP in simulation trial 5. The result for the paper map setting show that RSCP and MCP perform equal in simulation trial 5. The task was not too difficult for any of the teams when a traditional support was used in the command post.

One important task for the teams is to synchronize the activities and use their resources in an effective way. Figure 8 shows the average number of active units. This is an indication of the participants' ability to use their resources. The data shows that all teams have a positive learning curve. It also shows that RSCP teams with GPS put their resources to work to a greater extent than the other team compositions. This is an expected result.

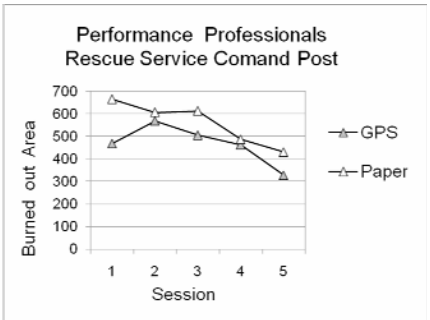


Fig. 6. Performance RSCP

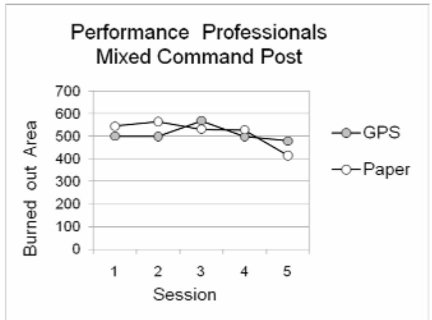


Fig. 7. Performance MCP

4.2 Time to First Engagement

In emergency response the time from alarm to first response is important. After an alarm, there should be two parallel tasks going on. The first is that the command post should create a long-term plan for the response. The second is that the ground chiefs start first response based on their local knowledge about the situation and previous instructions. When the command post have sized up the situation and created a long term plan they should command the fire fighting chiefs to start acting according to plan. The metric “Time to first engagement” (Figure 9) indicates the behavior learned by the ground chiefs, if they act directly or if they wait for order from the command post.

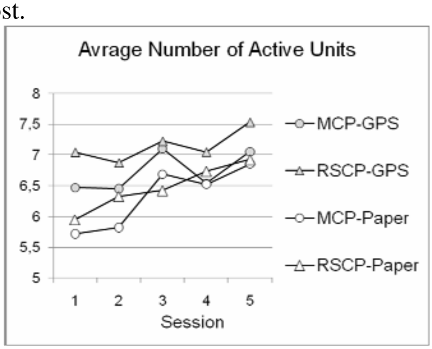


Fig. 8. Average number of active units

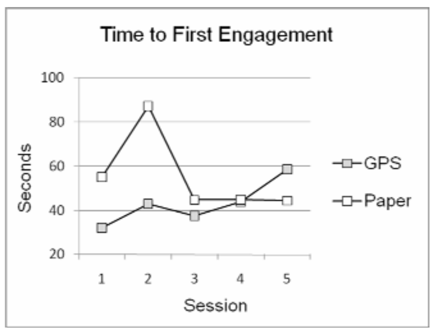


Fig. 9. Average number of seconds between the alarm and the first engagement

The result shows that the ground chiefs in teams that have paper map in the command post, have after the third trial learned a routine for response and they perform the same routine with the same speed every time. The result also shows that the teams that have GPS do not initiate first response according to a predestined routine. They await orders from the command post. This is an indication that the decision making process is altered by the GPS. The short term decisions are in the last trial no longer made by the ground chiefs they are passed over to the command post, and no short term instructions have been developed by the teams.

4.5 Training Assessment

The result on how the participants perceived the training can be seen in Figure 10. On the question about what they thought they had been training on, the participants were given ten topics, over which to distribute a total of 100%. The topics were common goals, common task, functional roles, good communication, common situation awareness, problem solving, decision making, conflict solving, empowerment and managing recourses. The basis for the eleven topics was described by Offermann and Spiros (2001) but somewhat adjusted to meet the simulation based training situation.

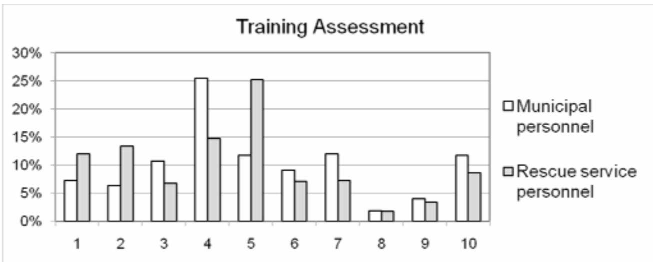


Fig. 10. The percent on each topic the participants think they have been training on

The results show that rescue service and municipal personnel's opinions on the training differs in topic 4, communication and 5, common situation awareness. The municipal personnel strongly thought that they are training on how to communicate in a crisis even. The rescue service personal strongly thought that they are training on how to create common situation awareness in a crisis even.

5 Discussion

The study used experiential learning in the research method. The teams were expected to evolve during the experiment to learn and acquire a working method adapted to the decision support the team had access to. The results in terms of performance, resource usage, time to first engagement, and perceived training gives a picture where the command post with GPS support has more tasks to handle compared to the command post that have paper maps as support, and the work is more evenly distributed between the command post and ground chiefs in the paper map condition.

GPS support offers more information than papers maps to the participants, and the simulated task should reasonably be easier to solve with the help of GPS. With a closer examination of the performance data on the rescue service personnel command post, RSCP, and municipal employee command post, MCP, it appears that this is not the case. Compared with the RSCP and MCP in the paper map condition has RSCP a better result in the GPS condition, but the MCP in GPS score lower, especially in the 5th and most important attempt (Figure 6 och 7). The RSCP and MCP teams in paper map, and the RSCP teams in GPS, all have a positive learning curve with regards to performance. MCP in GPS is different. They have no learning curve. They have no performance development over the five trials.

The problem here is how the groups solve allocation of work and the ability of their command post to handle the work. The teams in GPS and paper map have different strategies to allocate responsibility. In the *paper map condition*, the time between the ground chief detect fire and first response is constant for the simulation trials 3-5. The ground chiefs act in a tactical command mode and in accordance with the procedures developed by the team during their AAR. Discovery of fire is communicated to the command post, and then the direct responsibility for response is with the ground chiefs who act accordingly. The time between fire detection and extinction can be kept constant as c3fire specific tasks, ie moving the units and use the communication tool are at an unattended tactical level. The instruction for fire detection is used in the same way every time; the manner is leaned and generates the same measured time in trial 3-5 (Figure 9).

In the *GPS condition* the allocation of responsibility is different. The time between detection and first extinguished cell is not constant (Figure 9). From the third attempt, where c3fire specific tasks have reached an inattentive tactical level, the time between discovery and response is increased. The explanation is that the instruction that the GPS groups uses is different. The ground chief that detected fire, alarm the command post and awaits orders. The direct, as well as long-term, responsibility for response lies with the command post. The ground chief still has execution at his lot, but wait until the plan is set and reach him via an order. With this the command post in the GPS condition is more stressed than the command post in the paper map condition, and the GPS ground chiefs is relieved of work load. RSCP teams in GPS have the best performance of all subgroups. This means that they can handle the extra stress of the total, direct and long-term, responsibility for planning. MCP teams' performance in GPS has no progress, which means that this allocation of responsibility is unfavourable to them. They have no ability to strategically manage both direct and long-term planning. Their control mode is lowered to an opportunistic level.

The participants' perception of training shows that municipal employees, as opposed to rescue service personnel, strongly feel that it is communication that has been trained during the experiment. One explanation may be that Swedish municipal employees normally are not involved in managing crises, therefore the need of preparedness training. Their communication is normally based on consensus as an ideal. Rescue personnel have from experience knowledge about communication and work processes for managing crises. This in itself means that the municipal employees teams acquire an additional burden and are in the case of GPS not capable to evolve in terms of performance due to cognitive over load.

Otherwise, the participants' perception of the training shows that all 10 proposals for training were trained (Figure 10). This is reasonable as the scenarios in the simulation where designed to be open to team development, no predetermined direction or training area was designated.

Refernces

Artman, H., Waern, Y.: Distributed cognition in an emergency Co-ordination Center. Cognition, Technology & Work 1, 237–246 (1999)

Brehmer, B., Dörner, D.: Experiments With Computer-Simulated Microworlds: Escaping Both the Narrow Straits of the Laboratory and the Deep Blue Sea of the Field Study. Computers in Human Behaviour 9, 171–184 (1993)

Brehmer, B.: Microworlds and the circular relation between people and their environment. Theoretical Issues in Ergonomics Science 6(1), 73–93 (2005); Gestrelius, K.: Simulation and training games: Experiential learning, SAAB Training Systems, AB (1998)

Granlund, H.: Experiential Learning in computer based simulation training - Experiences from research on team decision making. In: proceedings of, International Conference on Information Technology in Education within CSSE, Wuhan, China (2008)

Granlund, R.: Web-based micro-world simulation for emergency management training. In: Future Generation Computer systems (best papers form the conference Websim 1999), vol. 17, pp. 561–572. Elsevier, Amsterdam (2001)

Granlund, R., Granlund, H., Johansson, B., Dahlbäck, N.: The Effect of a Geographical Information System on Communication in Professional Emergency Response Organizations. In: Proceedings of 7th International Conference on Information Systems for Crisis Response and Management, ISCRAM 2010 (2010)

Hollnagel, E., Woods, D.D.: Joint Cognitive Systems. Foundations of Cognitive Systems Engineering. Taylor & Francis, Boca Raton (2005) ISBN 0-8493-2821-7

Hutchins, E.: Cognition in the Wild. MIT Press, Cambridge (1995)

Johansson, B., Trnka, J., Granlund, R., Götmar, A.: The Effect of a Geographical Information System on Performance and Communication of a Command and Control Organization. The International Journal of Human-Computer Interaction, Special issue on Naturalistic Decision Making with Computers 26(2&3), 228–246 (2010)

Kolb, D.A.: Experiential Learning – Experience as a source of learning and development. Prentice Hall, Inc., New Jersey (1984)

Lindgren, I., Smith, K.: Using microworlds to understand cultural influences on distributed collaborative decision making in C2 settings. In: Proceedings of the 11th The International Command and Control Research and Technology Symposium (ICCRTS), Cambridge, UK, Awarded the Willard S. Vaughan, Jr. Best Student Paper Award (2006)

Offerman, L.R., Spiros, R.K.: The Science and Practice of Team Development: Improving the Link. Academy of Management Journal 44(2), 376–392 (2001)

Rasmussen, J.: Skills, Rules, and Knowledge; Signals, Signs, and Symbols, and Other Distinction in Human Performance Models. IEEE Transactions on System, Man, and Cybernetics smc-13(3) (May/June 1983)

Schmidt, K., Bannon, L.: Taking CSCW seriously. Supporting Articulation Work. Computer Supported Cooperative Work 1, 7–47 (1992)

Crisis Management Training: Techniques for Eliciting and Describing Requirements and Early Designs across Different Incident Types

Ebba Thora Hvannberg and Jan Rudinsky

University of Iceland, Dunhaga 5, 107 Reykjavik, Iceland
ebba@hi.is, janr@hi.is

Abstract. Crises occur seldom, but when they occur they have high impact on the enclosing organization and its stakeholders. Examples are plane crashes, train incidents and bomb threats, but the types of crises are virtually endless. We report on research of early phases of the development of a crisis management training simulator, with the goal of understanding different representations and transitions between steps of a development process. The focus of the research study was on how the different representations did align with a given process model and how these representations lent themselves to a consolidation activity. The results were that consolidation across data sources starts early during the understanding phase and that stakeholders like to validate abstract models. The consolidated conceptual models mostly addressed work and strategies. No formal attempt was to consolidate across management and organization structures.

Keywords: Crisis management, software systems, consolidation, process model, representations.

1 Introduction

Fortunately, crises occur seldom, but when they occur they have high impact on the enclosing organization and its stakeholders. Examples are plane crashes, train incidents and bomb threats, but the types of crises are virtually endless [1]. Crisis management is the core of response to such serious accidents. In this context, crisis management is about organizing work, resources and information in a secure and timely manner with the aim of rescuing people and ensuring their best health conditions. The challenges of crisis management are manifold, but include variability of available resources that are sometimes scarce and sometimes plentiful. Examples of those resources are rescue equipment, transportation, medical support and manpower. Additionally, the surroundings and contexts of work, such as in mountains or at sea in different weather conditions, vary greatly. One challenge of the problem domain is that there is a high demand for synchronization and decision making among

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 225–234, 2011.

teams with members in different organizations. Furthermore, time is a critical component, distinguishing between life and death.

To meet these demands of crisis management, organizations install a scheme of activities, command and flow of information. A systematic approach, relating to procedures and plans, and a systemic approach affecting the whole organization, have to be addressed [2]. Training a system of crisis management is typically performed in large exercises where an accident and its response are simulated but in a real environment. In such cases, a hypothetical training scenario is created which participants can act out [3]. Although life training is successful and results in trained responders and commanders, such training is expensive. Therefore, a software system to simulate training is under development.

A simulator to train crisis management has to take into account various aspects of operational and training concerns. It contains multidisciplinary perspectives drawing from psychology, technology, social-political and technological-structural views, including procedures, policies, practices and routines, in addition to the traditional view of technology as a machine or a tool [1].

A crucial part of designing such a simulator for training of crisis management is domain analysis. The objective is to learn about the different actors, their work, their interaction and information flow between them. Typically, different methods are applied to obtain data during domain analysis, such as interviews, empirical observations and surveys [4].

There seems to be a plethora of methods to analyse domains, but there is limited knowledge on how the output of elicitation is abstracted or consolidated, and how the output is formed into design. We know that data is abstracted, but there is a pendulum between using concrete representations such as scenarios to more abstract and formal ones such as task models [5]. It is essential to understand these activities, since in the transition from analysis to design it has to be checked whether all the scenarios of use are accommodated adequately by the design [6].

A domain analysis unavoidably requires consolidation of data across sources, workers, customers and contexts. There are several motivations for such consolidation. First, the aim of consolidation can be to make the models more detailed, by connecting data from e.g. different sources, but it is also to make them more abstract, thus allowing the designer and the developer to see the common factors in the software system. A second motivation is to tailor it to a broader market, resulting in a valuable product for its owner. Third, the recent increased interest in domain specific languages is a motivation to describe domain knowledge in generic models. This can encourage building a common platform for crises management simulation that will allow teams to share lessons learned, throughout their organization, to improve planning [2].

In this paper, we report on research of early phases of the development of a crisis management training simulator, with the goal of understanding different representations and transitions between steps of a development process. The focus of the research study was on how the different representations did align with a given process model and how these representations lent themselves to a consolidation activity.

2 A Process Model

The development team of the training simulator did not use a particular development process model as a reference. For the purpose of a reference process in this research study, we selected a scenario driven process of human-centered interactive system design, described by Benyon [7]. In it, scenarios are suggested and meant to help developers ease into the abstractions by allowing them to write concrete steps, which described tasks, contexts and users roles. In the scenario-driven process, Benyon shows developers how they first can develop stories, scenarios and personas and then write abstract requirements that make up a conceptual model. Once the requirements have been stated, a concrete scenario showing how the system is used can be developed. This can also be visualized with prototypes, mostly of low fidelity. Abstracting again from the visualization is a formalization of the design with use cases, and the final step is to develop an interactive system. The initial phase of the process is carried out to understand the problem domain, i.e. activities, tasks, roles and context. When several of these have been collected into a scenario corpus, it is time to abstract them, via generalization or aggregation, into a conceptual model. This completes the *understanding phase*. For each conceptual model, there are possibly many concrete representations. The responsibility of the *envisioning phase* is to recommend an appropriate concrete design, which is based on requirements and design constraints, such as the technology applied. The final step is the *implementation phase*. This research covered only the understanding and the envisioning phases. It should be noted that the phases are not serial, and can overlap.

3 Understanding Tasks, People and Contexts

Extensive gathering of requirements took place in the initial phases of development for the crisis management training simulator. Three sites were visited in three countries. Two of the sites are airports and their reference scenarios are an aircraft incident and a bomb threat, but the third one is a railway company with a train incident as a reference scenario. The instruments, which have been applied during requirements elicitation, and representations that are output of the process's activities are shown in **Table 1**. Abstract and concrete representations were created during the understanding phase. What is common to them is that they were representations of individual incident types of crisis, such as a bomb threat or an aircraft incident. There is little consolidation across incident types during this initial phase of understanding work. Four work processes were identified, Alert, Rescue, Transport and Triaging of casualties and Security. In addition, there is a separate Communication process describing communication and coordination between the different work processes. For clarity, communication is also included in the work processes.

There were attempts to divide the domain into facets or aspects. Work is separated from training, e.g. in the user requirements and the training requirements. There was an analysis of worker competencies, mainly targeted for training, in the requirements documents. Furthermore, there is an analysis of management and organization, like in

Table 1. Understanding people, tasks, contexts and requirements

Goal	Phase	Elicitation instrument	Representations	Concrete / Abstract
To gather initial information on work and training.	Preliminary	Informal meetings with stakeholders, presentations of stakeholders on work.	Written scenarios describing a series of timed events of work.	Concrete
		Manuals on work. Observations of the working environment.	Time series that show interactions between roles.	Roles are abstracted, times are concrete.
			Mind map of activities of different roles of actors.	Roles and activities are abstracted.
			Contextual Design Work models of individual customers.	Sequences of activities are abstracted, roles too, Physical and Artifact models.
To gather information on work and training and validation of representation s of the previous phase.	Understanding	Observation of live training and table top exercise. Validation of work models with crisis managers.	Contextual Design Work models of individual customers.	Sequences of activities, control and roles are abstracted. Physical and Artifact models.
To gather data on stakeholders' priorities of system's capacities.	Understanding	Questionnaire on training competencies.	Training competencies required (hierarchy).	Competenci es are abstracted.
	Understanding	Questionnaire on user requirements.	User requirements as text.	Abstract
Validation	Understanding	Interviews	Refinement of scenarios and models.	

Table 1. (*Cont.*)

Validate whether the stated training requirements are unclear, incomplete, ambiguous or contradictory.	Understanding	Translate textual scenarios into visual form.	Story boards of scenarios	Concrete scenario per site.

Björner's six facets: intrinsics, support technology, management and organization, rules, regulations and scripts and human behavior [8], and Cognitive Modeling Framework[9]. Emphasis on strategies could only be found in the Contextual design sequence models, which included decisions and decision support. There was little emphasis on information or data modeling, except in the Contextual inquiry artifact model, which had references to data forms used during work. The Contextual design flow model, which would have provided more information on data, was not created due to lack of detailed information. Technology support was analysed in the Physical and Artefact models. These were mostly from the physical world, and not from the world of information technologies, except for communication technology. Because of the importance of communication, including coordination in the domain, analysts decided to include the communication between roles or organizational units in a Communication model and in the individual sequence models, i.e. the Rescue work model, the Alert work model, the Transport work model and the Security work model. Communication is either modeled as a meta-level construct of a sequence model, similar to trigger, intent, decision making, decision support or performance or as a separate sequence.

Some lessons were learned from this first phase. The different representations show that analysts frequently go back and forth between abstract and concrete presentations. Thus, some consolidation across workers and data sources starts even at an early stage. Some stakeholders had difficulty validating the concrete written scenarios, but preferred to look at abstract models. The reason was that they did not trust the scenarios to cover each activity and coordination between them. For them, the abstractions were more accessible and easier to validate, because of the possible different alternatives even within an incident type, e.g. an aircraft incident. The reason may be that the crisis managers that were part of the stakeholder group were used to work with abstract procedures such as parts of plans, organization charts, role specifications etc. A third observation we made was that the representations show various abstractions, e.g. hierarchical structures, mind maps and contextual inquiry work models. This diversity of the representations reflects the developers' experience and practices. All of those are informal in nature. Might it mean that developers have not yet found a formalism which meets their demand?

4 From the Concrete to the Conceptual

Concrete representations, whether they are essential use cases, stories or scenarios, lend themselves to easy readability for users [10]. They are relatively well scoped and describe a series of action in a certain context. However, they rarely describe more difficult constructs such as alternative paths, iterations, parallel tasks, or relationship between data, tasks or roles, except as concrete examples. Besides, they may not lend themselves to be a model after which a system can be developed. Thus, we need abstract representations. Abstraction involves generalization and aggregation over different stories or scenarios. Generalization is about finding common tasks, objects and roles across different scenarios, but aggregation ensures inclusion even if components are only part of some stories and not others. Developers encounter difficulties when starting to abstract from the concrete scenarios as they need to cover wide ground and need to use more complex constructs as mentioned above. This abstraction, or consolidation of representations is described independently in Benyon [7] and Beyer and Holtzblatt [11], but the steps are similar. This phase is an interface between how work is carried out currently and the new system. Whether it is a conceptual scenario or a conceptual work model, all context and individual personas are removed [7]. The question then, how much of the context should be removed during the conceptualization? Are there, for example, technologies that are a part of the work environment which should not be a part of the future system? Are there organizations, e.g. teams, divisions or departments that should be removed from the future system? Will the user behave differently in the new system? If we think that there will be such changes, the particularities belong to the work's context and should be eliminated in the conceptual model. Instead of removing the technologies and support entirely from the domain description, it can be advantageous to separate it clearly as Björner [8] does, since the analyst realizes its purpose in the domain description. An example in the crisis management is triaging. Responders use small colored boards to tick off and count the seriousness of an injury. This task can be described in a user story or a scenario. However, if the developer wants the flexibility to change the technology, i.e. the colored boards, in the future system, it should be excluded in the conceptual scenario. Thus, it is a part of the current context but not the future one. Similarly, Benyon divides the domain into four areas: people, activities, context and technologies.

When the crisis management training simulator was developed, two conceptual models emerged, namely, a consolidated contextual design sequence model of work and topic maps (see **Table 2**). The latter is still under construction. A third one is a generic reference scenario, described with informal flow diagrams. The consolidation of the work models has been quite labor intensive, since the stakeholders are at three sites and the crisis management is different for the three incident types. Still, there are many commonalities, which could form a basis for a domain specific language for crisis management. The consolidation has been carried out on meta-level constructs such as triggers, goals and steps of activities. Also, the Contextual Design work models have meta-level constructs, called decisions and decision support. These were applied during consolidation as an activity step, but sequences were not merged on outcomes of decisions that crisis managers or first responders across incident types need to make. The generic reference scenario models decisions, so clearly it should be a meta-level construct in a conceptual model for crisis management.

The consolidated conceptual models mostly addressed work and strategies and little technology support. There was not a formal attempt to consolidate across management and organization structures. The conceptual model distinguishes commanders from first responders, but does not further divide commanders into levels.

Table 2. Understanding / Conceptual modeling

Goal	Phase	Instrument	Representations	Abstract or Concrete
Consolidation of work models	Conceptual modeling	Consolidation of sequence models	Contextual design Sequence models	Abstract
Consolidation of scenarios	Conceptual modeling	Consolidation of written scenarios	Flow diagram of activities. Management levels, Information flow, Events, Decision points, Actions.	Abstract
Formalizing concepts	Conceptual scenario	Conceptual design	Written text of operational and training concepts.	Mostly Abstract
Formalization of rules in the simulator	Conceptual modeling	Procedures, Organization, Simulation	Topic maps of events in simulator	Abstract

5 From Conceptual Models to System Interactions

Gradually, conceptual scenarios, constituting requirements, can be developed into larger conceptual models of the system with increasing detail. Objects are modeled with their characteristics and relations, tasks are modeled with structures of iterations, alternative paths and parallel threads. Tasks have pre-conditions, post-conditions and other invariants. As soon as the requirements are clear, the conceptual models can describe abstractions of the future system. This phase is still underway. We continue to follow Benyon's development process, where the third phase is envisioning the system through its user interface. The transformation from the conceptual models to system interactions involves adding context of the new system to the models. Some developers will choose to add parts of this context to the abstract representation, i.e. the conceptual model, others will prefer to attempt to envision the new systems in prototypes. In the development of the crisis management training system, low-fidelity prototypes have been created with power point slides. Some of them have had simple interactions but others a series of snapshots (see **Table 3**). As a part of future work, there needs to be an evaluation of how the different incident types require particularities in the concrete implementation.

Table 3. Envisioning and evaluation

Goal	Phase	Instrument	Representations	Abstract/ Concrete
To visualize interactions of parts of scenarios or work model	Envisioning	Develop power point slides based on a conceptual model	Power point slides with simple interactions	Concrete
To compare interaction choices	Envisioning	Develop power point slides based on a conceptual model	Snapshots of different choices of interactions, e.g. input devices	Concrete
To receive feedback on first ideas of interaction	Evaluation	Interview, discussing the previous two representations		

6 Discussion

During the understanding and the envisioning phases, a number of difficulties were encountered. The first one is the difficulty to separate training issues from the operational issues. One of the known objectives of building an information system is to improve processes. It is not only about developing a system to aid with an operation; it is also about improving work [5, 9]. The variation in a training system is that improvement should occur at the training level and not at the operational level. The separation of training from operation on the other hand opens up the opportunity to simplify operations and focus on training objectives. A second difficulty relates to the simulation. Developers are used to build systems which execute tasks on real objects, but do not simulate that execution. Another issue of the simulation is that modern work requires the use of information systems. When training for this work in a simulator, how are these information systems simulated, or integrated into the simulator? The third issue is the concept of a training scenario, which is the input model of the simulator. The training scenario in crisis management training may include a number of things such as the size and type of the event, number of resources initially available etc. Thus, the scoping of the simulation is essential. Finally, a potential difficulty with the development of the crisis management simulator is its evolution. During the understanding phase it became clear that the underlying operation was not a static system, but evolved easily when operational improvements were called for and as new information technologies were adopted and changed work.

7 Conclusion

This paper has attempted to analyse the work carried out by several analysts of a crisis management training simulator. The domain of crisis management is complex and the project is characterized by a project team with different experience and practices. This study showed that simultaneously developers address a number of issues at the macro and the micro level, and they tend to use abstraction to help them. Probably, developers will attempt to create many such abstractions, to suit their needs.

One way of describing work is to divide it into processes, which communicate between them, eliciting triggers of actions, their intent, decisions to be taken, decision support, output of the actions, required performance and breakdowns. This is the approach analysts of the crisis management training simulator have selected with good results. Perhaps it is due to the type of work of crisis management, i.e. it is heavily activity oriented, where coordination, communication, decision making and different roles in a predefined management and organization structure is of utmost importance, but data, as input or output to functions, plays a less of a role. Another approach would have been to focus on functions and objects or data. Probably, these constructs are too constraining, at least in the initial phases, and do not give the developer enough flexibility. It is our conclusion that function or data-oriented specification can only succeed an activity and communication oriented analysis like the one which has been described during the initial understanding phase. Thus, the function or data-oriented specification takes place during the Envisioning phase.

We further conclude that the concrete representations such as storyboard scenarios and low fidelity prototypes have been useful. However, developers need a conceptual model which gives an overview of the domain and ensures that all parts are included. We propose that a domain specific language for crisis management training may be desirable, at least at the conceptual level. There have been some attempts to define a generic model for crisis management [12]. The danger is that such a generic model will too broad and not useful for an innovative system.

Similarly, it may be questioned whether one development process model for all is the most useful one for developers. This paper has used Benyon's process as a reference, and it has proven useful. However, it does not address issues specific to simulators, such as intelligence and realism. It does not address training or scenario based training. It is likely, that such a process would support developers better than a generic one. Already, process models for simulators and game development [13] have been suggested and one for developing training of decision-making in a synthetic environment [14], an area relevant to crisis management training.

References

1. Pearson, C.M., Clair, J.A.: Reframing Crisis Management. The Academy of Management Review 23, 59–76 (1998)
2. Pearson, C.M., Sommer, S.A.: Infusing creativity into crisis management: An essential approach today. Organizational Dynamics 40, 27–33 (2011)
3. Moats, J.B., Chermack, T.J., Dooley, L.M.: Using Scenarios to Develop Crisis Managers: Applications of Scenario Planning and Scenario-Based Training. Advances in Developing Human Resources 10, 397–424 (2008)
4. Zave, P.: Classification of research efforts in requirements engineering. ACM Comput. Surv. 29, 315–321 (1997)
5. Diaper, D.: Task scenarios and thought. Interacting with Computers 14, 629–638 (2002)
6. Stary, C.: Shifting knowledge from analysis to design: requirements for contextual user interface development. Behaviour & Information Technology 21, 425–440 (2002)
7. Benyon, D.: Designing Interactive Systems. Pearson Education, London (2010)

8. Björner, D.: Software Engineering 3 Domains, Requirements, and Software Design, vol. 3. Springer, Heidelberg (1998)
9. Vicente, K.J.: Cognitive Work Analysis. Lawrence Erlbaum associates, Mahwah (1999)
10. Benyon, D., Macaulay, C.: Scenarios and the HCI-SE design problem. Interacting with Computers 14, 397–405 (2002)
11. Beyer, H., Holtzblatt, K.: Contextual Design. Morgan Kaufmann, San Francisco (1998)
12. Kruchten, P., Woo, C., Monu, K., Sotoodeh, M.: A Human-Centered Conceptual Model of Disasters Affecting Critical Infrastructures. In: Proceedings of the 4th International ISCRAM Conference, Delft, The Netherlands (2007)
13. Raybourn, E.M.: Applying simulation experience design methods to creating serious game-based adaptive training systems. Interacting with Computers 19, 206–214 (2007)
14. Jenkins, D.P., Stanton, N.A., Salmon, P.M., Walker, G.H.: A formative approach to developing synthetic environment fidelity requirements for decision-making training. Applied Ergonomic 42(5), 757–769 (2011)

Development of Mobile Evacuation Guides for Travellers and Rescue Personnel

Viveca Jimenez-Mixco[1], Jan Paul Leuteritz[2], Eugenio Gaeta[1], Maria Fernanda Cabrera-Umpierrez[1], María Teresa Arredondo[1], Harald Widlroither[2], and Mary Panou[3]

[1] Life Supporting Technologies – Technical University of Madrid, Madrid, Spain
[2] Fraunhofer IAO / University of Stuttgart IAT, Stuttgart, Germany
[3] Centre for Research and Technology Hellas, Hellenic Institute of Transport, Thessaloniki, Greece
{vjimenez,egaeta,chiqui,mta}@lst.tfo.upm.es,
{Jan-Paul.Leuteritz,Harald.Widlroither}@iao.fraunhofer.de,
mpanou@certh.gr

Abstract. Whenever an emergency situation occurs, a correct response of the rescue teams and an optimal communication and coordination among them is required to assure that all the people needing help are assisted within a short period of time. This paper presents an approach for the development of innovative HMIs within SAVE ME project to save the lives of travellers and rescuers, giving particular emphasis to the most vulnerable people. The emergency response system detects natural and man-made disaster events in public transport terminals/vehicles and critical infrastructures and supports quick and optimal mass evacuation guidance.

Keywords: Self-evacuation, Mobile Device, Emergency, Indoor Navigation.

1 Introduction

The whole world has experienced some type of disaster event within the last years. Not only natural phenomena, like the earthquakes in Haiti and Chile in the past months, but also man-made disasters, like terrorist attacks, have affected wide areas and produced severe personal and material damages.

During an emergency situation, all the involved individuals and rescue teams need to be enabled to react immediately, in order to allow for a quick and secure evacuation of the area. This is especially critical when the event occurs within a closed environment, such as a public transport facility, where travellers are likely to become trapped and/or injured and thus dependent on rapid medical assistance. This implies: 1) that rescuers need precise information on the situation, seamless communication means between themselves and the operations centre and proper guidance to the location of the trapped travellers, and 2) that all travellers, including children, elderly and disabled need appropriate guidance to be able to escape from the affected area [1][2].

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 235–243, 2011.

This paper presents the different steps and methodologies in the development of innovative HMIs for the support of self-rescue and evacuation of travellers, as being applied within the SAVE ME project. The project follows an integrative approach, combining different technologies and thus creating new solutions of how to improve the interaction of the users with this system.

The SAVE ME project is a research and development project, co-funded by the European Commission within the Seventh Framework Programme. SAVE ME aims at integrating state of the art technology not only to detect natural and man-made disaster events in public transport infrastructure but also to actively support life saving activities by facilitating both the self-rescue of the travellers as well as the rescue work done by fire fighters. Particular emphasis is put on the most vulnerable travellers: children, elderly and the disabled.

2 Technical Approaches

2.1 Basic Technology

It is a virtually immutable universal law that when communications are needed the most desperately and urgently, the difficulty of effecting the desired communication increases exponentially. Communication is critical during an emergency and needs to be addressed thoroughly within the disaster-response plan. The communication challenges include reaching people in different locations with different devices quickly and simultaneously; providing the right message (in terms of content, length and format); monitoring delivery and response; and ensuring that the process is initiated and suspended at the right times.

One of the ways the project is progressing beyond state of the art addressing these challenges is through the use of automated-notification technology, which is a very effective method of rapidly distributing information to large numbers of people in a crisis. For that purpose, users will be pre-registered to the system and an application will be preinstalled in the user's device in order to enable the automated communication.

To be effective, the SAVE ME crisis-communication plan provides a high-fault tolerance level, anticipating and overcoming potential obstacles such as power outages and downed phone lines. The objective is to enable continuous operation, reliable and safe through the whole process of events detection, i.e. from sensor detection to the emergency centre and user's devices.

The proposed detection and communication system consists of: 1) a wireless sensor network (WSN) grid, including indoor localization and environmental detection (of fire, flood, temperature, noise, etc.) combined with hybrid localization techniques, to allow detection of the emergency key parameters the positions of the travellers in the emergency area, and their movements; 2) a fault tolerant telecommunication module that allows transmission of these data to the operators centre, even under adverse conditions. It is based on low cost ad-hoc Wi-Fi routers located in vehicles and stations with preinstalled and automatic upgradable emergency software that will become active when an emergency is detected. This specialized network infrastructure is focused on the communication between the decision support system (DSS, the core intelligence of the system) and the end-user components, ensuring the safe transmission of guidance messages. It provides a consistent replication of the evacuation plans and communicates and informs directly the user.

3 Methodology: User Interface Design & Development Issues

3.1 Panic and Stress

One central misconception that we want to tackle within this article is the still widespread belief that in emergency applications, the primary concern of user interface design was to make sure that people will not "panic", run around headlessly and behave silly or even aggressively.

Studies on past emergency situations, such as the evacuation of the World Trade Centre on September 11, have shown that the majority of the people acted with calm [3]. Furthermore, people are actually acting especially pro-social, when confronted with a dangerous situation [4]. This counts especially for groups that have a strong cohesion, such as families and groups of friends. From what is known about such extreme situations, we can assume that between 10% and 15% of the affected people will be resilient enough to behave rationally in a case of emergency. These people can assume leader-roles and thus provoke pro-social and constructive behaviour in the about 70% of the people, who react calm, dazed, and rather passive to threats [5]. Only the remaining 10% to 15% shows either irrational or egoistic behaviour. In fact, often it is rather pro-social than anti-social behaviour that causes an extra threat: In an experimental study, Gershon and colleagues [6] showed that helping others one significant cause of delayed self-rescue.

Stress is an important factor that leads people's actions in an emergency situation. It can be defined as a physiological reaction of the human body to a threat. According to the two-mode model [7], high stress can reduce rational and planned behaviour, and lead to more instinctive behaviour. The degree of stress a person experiences, is assumed to depend on various factors, especially the salience of the danger and the person's belief about his/her ability to behave adequately. Thus, the stress level of an individual is difficult to predict.

Stress is not only a negative mental status. Yerkes and Dodson proved already in 1908 that a medium stress level is necessary for optimum performance. This also applies to emergency situations and as Gershon [6] found people unnecessarily delaying self-rescue in order to complete unimportant tasks (such as collecting personal items before leaving), pondering too long about alternative escape paths and slow walking, we suggest that in some situations, the user interfaces should rather raise than lower the users' stress levels.

However, there are cases in which masses of people show irrational or anti-social behaviour and where others are, e.g. tread to death. These phenomena are widely known under the term of "mass panic" but there are some important influence factors we need to know in order to successfully avoid them. When we talk about panic, we mean a state of very intensive anxiety or fear. However, this fear does not need to be the root of the problem – sometimes, it is just the consequence.

First of all, people change from the generally pro-social emergency behaviour to anti-social behaviour when they assume that they are in a competitive situation and that they need to fight others in order to escape [8]. This can, e.g., be the case when an exit is blocked. Unfortunately, people have a tendency to leave buildings the same way that they entered it [5], which e.g. lead to people being tread to death in a night-club fire in West Warwick on Rhode Island in February 2003.

Another cause of people treading on each other is bad “front-to-back communication”. This was, e.g., the cause of people getting killed at the Love Parade in Duisburg, Germany in 2010. People crowded up in a too tiny space because the people in the back of the crowd did not realize that people at the front were standing in front of a closed door. Once people get jammed up too much, physical powers are unleashed that push people over one another. When caught in such an uncontrollable movement, the feeling of panic is a rather logical reaction.

3.2 Testing Emergency Applications

It is obvious that mobile evacuation applications are difficult to test under realistic conditions, as it is ethically not acceptable to even trick test participants into believing that their lives were in danger. As we have seen, a human being’s capacities of rational thinking are difficult to predict for emergency situations, as there are too many unknown influence factors. Even though we know that people generally react calm and rational, we cannot be sure about what will happen in specific situations.

However, from a practical point of view, we can manipulate testing situations well enough to get an idea about the usability of such systems. Competitiveness can be simulated by granting participants extra-reimbursements if the make it quicker out of the danger zone than the others. Airplane-evacuation experiments are using this kind of scheme [9]. Information overflow or reduced mental processing capabilities can be simulated by assigning the participants extra tasks, such as simple calculations. Finally, social stress can be induced via experimental manipulations, e.g. using Virtual Reality technology. The psychological consequences of socially caused stress (e.g. Kirschbaum et al [10]) are supposed to be comparable to these triggered by other threats.

Within the SAVE ME project, various steps have been undertaken to ensure a high usability of the user interfaces to be created. Apart from a literature analysis, two experiments were performed. One researched how stressed people managed to find their way within a virtual labyrinth, and also their cognitive capacities in the same situation. The other experiment was focused on finding out how people perform in escaping from a labyrinth when they have either mobile guidance only, or collective guidance only, or both of them available. Results are yet to be analyzed and published. The usability of the guidance prototypes is still to be tested, relying on the above mentioned methods.

3.3 The Special Requirements of Emergency Applications

Mobile self-evacuation means have some weak points: First of all, they might be neglected or disregarded as people draw their attention to the catastrophic event happening. Secondly, certain people might not trust the information given there and rely rather on the known escape paths. Finally, the device can fail due to technical problems.

For all these named reasons, it is important to provide mobile rescue guidance in combination with intelligent infrastructure signalling. Apart from redundantly providing information on the escape route, it can also possibly serve to slow people down if an exit is too crowded, thus avoiding bad front-to-back communication and increasing the escape speed of the entire mass of people. Thus, this system is called “collective herding guidance” in SAVE ME.

The next big requirement is accessibility. Due to the fact that the mobile guidance can save the life of a person, it has to be fully accessible. Furthermore, it is especially the group of vulnerable travellers that needs the mobile guidance, as collective guidance means might not serve them. Here, we refer to disabled people, to children and the elderly. In the end, all travellers are expected to profit from accessible interfaces. Multimodal devices with combined voice- and visual output might, e.g., serve all users when vision or hearing are diminished due to loud noise or, e.g., smoke.

Finally, the location also plays an important role. The SAVE ME system will undergo two different pilot tests: One in the NEXUS metro-network of Newcastle upon Tyne (GB) and one at the Gotthard-Straßentunnel (Switzerland). In the scenario of the Gotthard-Tunnel, the biggest challenge is to inform and convince the people travelling in the cars that they need to exit the cars and head towards the nearest shelter room. Experience of the Gotthard Emergency services shows that even if people see a car burn in front of them, they often remain in their cars. In the case of the metro station, the challenge is rather to disperse the people successfully to the various remaining escape routes. An unsolved issue here is that groups, such as families, tend to stay together in emergency situations. Thus, initial escape route calculations by a system that does not know which of the people belong together will not be enacted exactly as planned.

3.4 SAVE ME Guidance Design & Development Principles

The following principles have been phrased in the SAVE ME project in order to guide the design of the mobile evacuation aids:

- The system is “walk-up-and-use”: It needs to work without any learning. Effectiveness and Efficiency are highly important while Satisfaction is not (ISO 9241).
- The situation of use might be extreme, causing limited concentration or slower mental processing. Thus, the user-system-interaction needs to be simple and unambiguous: Reducing interaction steps, options and avoiding menus are of highest priority. The necessary information for the next step must always be available. On the screen, written text is to be avoided whenever possible.
- The accessible user interface (and also the calculation of the optimal route) requires a user profile, naming possible restrictions (such as vision loss).
- Clearly communicate that the proposed escape path is the optimal one.
- The guidance has priority, thus, the application needs to be maintained alive or re-launched automatically after an abortion (e.g. making a telephone call).

- The mobile guidance is provided on the proprietary device, thus the basic design principles of the very device should be used.
- It is not recommendable to give an extra advice that people should help each other. They will do so anyway, possibly even more than they should. Instead, the system should communicate feelings of safety by pointing out that the evacuation goal is realistic and will be reached if complying with the guidance information given.

4 Results: Implementation Issues

4.1 Harmonizing Mass- and Personalized Guidance

Appropriate human interaction in emergency conditions and critical visual, chemical and noisy environments are essential to assure the provision of valuable escape instructions to travellers at need and thus enable a fast and safe evacuation. Human interaction depends on the target person characteristics, such as age, language, mental or physical impairments, that can influence the understanding of the information. Thus, the emergency support strategies depend on parameters such as type of emergency, type of environment, topology of the location, situation criticality, type of device, etc. The decision support system will be in charge of processing the information to provide personalized and group wise routing for the people detected in the area.

The trapped travellers will be informed on the emergency and receive generic guidance by visual and acoustic means (situated displays and panels, speakers) of a collective herding guidance that will transmit group wise evacuation plans and hazard notifications to the travellers. The content of the messages depends on the location of the situated display / speaker. Avoiding confusion due to different notification means at the same position of the site is critical for a proper guidance. For instance, different or uncoordinated messages coming from nearby speakers or guidance coming from displays and speakers may be confusing. Therefore, appropriate mapping between evacuation plans and evacuation guidance provided at different locations of the sites is required.

The most innovative guidance will be provided to those that are equipped with open and functional mobile phones. These users will get personalised guidance (i.e. knowing the person's location and the nearest unblocked exit, a point to point guidance will be provided), which will in addition be adapted to the person's profile, stored at his/her mobile phone (i.e. visual guidance for a deaf person and not acoustic; take into account the existence of stairs between current position and nearest exit for a wheelchair user, thus route him/her to another exit).

On the other hand, rescue units will be guided through their PDAs. They will be safely guided and monitored within the infrastructure and will be given instructions about the path to reach on priority the most endangered and/or most vulnerable citizens.

4.2 Low-Fi and Hi-Fi Guidance

Currently, there are several use cases in two main scenarios available for the provision of self-evacuation guidance through mobile phones. From a usability point-of-view, it seems obvious that the latter, which provides the most exact information, is considered the best one. However, in cases that the optimum cannot be reached due to technical restrictions, little information might be better than none.

- *Full connectivity* to the SAVE ME network, in this scenario the SAVE ME network does not crash after terroristic attack. So, the user has access to the routers, which provide up-to-date information via internet connection. With regard to this scenario we have three use cases due to the technological limitation of the user device
    - *Low-fi guidance:* in this use case the user's mobile device does not have TCP/IP connectivity; thus, static route guidance information is triggered by the SAVE ME network to the user's device via Bluetooth push service.
    - *High-fi dynamic guidance*: in this use case the user mobile device has limited TCP/IP connectivity (e.g. no support of Wi-Fi open library), dynamic browsable web based guidance information is triggered by the SAVE ME network to the user device via Bluetooth push service.
    - *High-fi transparent dynamic guidance*: in this use case the user mobile device has full TCP/IP connectivity (e.g. no support of Wi-Fi open library), dynamic browsable web based application is trigged triggered by the SAVE ME network to the user device via web-based push service.
- *Limited connectivity* to the SAVE ME network, in this scenario the SAVE ME network crashes after terrorist attack and the user has only access to an isolated router, so there is not any updated information available. Regarding this scenario, we have three use cases due to the technological limitation of the user device
    - *Low-fi static guidance:* in this use case the user mobile device has not TCP/IP connectivity; so static route guidance information is triggered by the access router to the mobile device via Bluetooth push service.
    - *High-fi static guidance*: in this use case the user mobile device has limited TCP/IP connectivity (e.g. no support of Wi-Fi open library), static browsable web based guidance information is triggered by the SAVE ME network to the mobile device via Bluetooth push service.
    - *High-fi transparent static guidance*: in this use case the user mobile device has full TCP/IP connectivity (e.g. support of Wi-Fi open library) so static browsable web based guidance information is triggered by the SAVE ME router to the user device via web-based push service.

5 Conclusions

Mass evacuation becomes very challenging during an emergency situation, especially when the event occurs in a closed environment such as transport hubs, road tunnels or metro stations. In order to optimize the life-saving process in emergency situations, tools and more comprehensive models are required to study and predict the user's behaviour. A well-prepared system can significantly reduce the impacts of an attack. Appropriate guidance for a variety of users (travellers, rescue personnel involved in the active mitigation of the event and operators in the control centre of the traffic infrastructure) is essential in these types of situations.

As discussed in the paper, mobile device based guidance, able to provide personalised instructions to the users, has a strong potential to face this issue. The most promising possibilities rely on the integration of different modalities in order to reach the broadest possible range of users, including the most vulnerable ones, for example, combining the mobile rescue guidance with intelligent infrastructure signalling.

The SAVE ME project integrates personalized guidance and mass guidance into a coherent system. The project aims at providing personalized guidance for everyone, independent of the mobile device used and the special user's needs, supporting mass guidance evacuation of public transport vehicles, stations and other critical infrastructures. Its Wireless Sensor Network with sensing, communication, computing and interaction elements and DSS, constitute the basis for its fully integrated and pervasive group guidance solutions.

Acknowledgment. This work has been partially funded by the European Union in the context of the SAVE ME project (SST-2008- 234027), coordinated by the University of Newcastle Upon Tyne. The project started in 1st October 2009, and has a duration of 36 months. The consortium is composed of the following partners: UNEW, CERTH, SIMUDYNE, CNVVF, IES, COAT, GST, MIZAR, USTUTT, UNIMORE and UPM.

References

1. Heath, R.L., Lee, J., Ni, L.: Crisis and Risk Approaches to Emergency Management Planning and Communication: The Role of Similarity and Sensitivity. Journal of Public Relations Research 21(2), 1532-754X (2009)
2. Cabrera, M.F., Arredondo, M.T., Rodriguez, A., et al.: Mobile technologies in the management of disasters: the results of a telemedicine solution. In: Annual Symposium of the American-Medical-Informatics-Association (AMIA 2001), pp. 86–89. Journal of the American Medical Informatics Association, Washington D.C.Source (November 2001)
3. Proulx, G., Fahy, R.F.: Account Analysis of WTC Survivors. In: Proceedings of the 3rd International Symposium on Human Behaviour in Fire, Belfast, UK, September 01-03, pp. 203–214 (2004) (NRCC-47309)
4. Sime, J.D.: Affiliative Behaviour During Escape to Building Exits. Journal of Environmental Psychology 3, 21–41 (1983)
5. Schneider, U., Oswald, M., Lebeda, C.: Evakuierung im Brandfall [Evacuation in the case of fire]. Schloss Schallaburg bei Melk an der Donau - Österreich, FSE Ruhrhofer & Schweitzer OEG, Brandschutz-Fachtagung (February 3, 2004)
6. Gershon, R.R.M.: Preliminary Results of the World Trade Centre Evacuation Study. MMWR 53(35), 815–817 (2004)
7. Carver, C., Johnson, S.L., Joormann, J.: Two-Mode Models of Self-Regulation as a Tool for Conceptualizing Effects of the Serotonin System. Normal Behavior and Diverse Disorders Current Directions. Psychological Science 18, 195–199 (2009)
8. Frey, B.S., Savage, D.A., Torgler, B.: Interaction of natural survival instincts and internalized social norms exploring the Titanic and Lusitania disasters. Proceedings of the National Academy of Sciences of the United States of America 107(11), 4862–4865 (2010)

9. Angelin, J., Blair, P., Carson, N.: Aircraft Evacuation Testing: Research and Technology Issues. Office of Technology Assessment – Congress of the United States (1993)
10. Kirschbaum, C., Pirke, K.M., Hellhammer, D.H.: The Trier Social Stress Test–a tool for investigating psychobiological stress responses in a laboratory setting. Neuropsychobiology 28, 76–81 (1993)
11. ISO 9241, Ergonomics of human-system interaction. International Organization for Standardisation (2008)
12. Mintz, A.: Non-adaptive group behaviour. Journal of Abnormal and Social Psychology 46, 150–159 (1951)
13. Roßnagel, H., Zibuschka, J., Muntermann, J., Scherner, T.: Design of a Mobile Service Platform for Public Events: Improving Visitor Satisfaction and Emergency Management. In: EGOV 2010, Lausanne, Switzerland (2010)
14. SAVE ME Consortium. (SAVE ME) Grant Agreement no 234027 – Annex I - Description of Work. European Commission, Brussels, Belgium (2008)
15. Savidis, A., Antona, M., Stephanidis, C.: A Decision-Making Specification Language for Verifiable User-Interface Adaptation Logic. International Journal of Software Engineering and Knowledge Engineering 15(6), 1063–1094 (2005)
16. Savidis, A., Stephanidis, C.: Unified User Interface Design: Designing Universally Accessible Interactions. International Journal of Interacting with Computers 16(2), 243–270 (2004)
17. Yerkes, R.M., Dodson, J.D.: The relation of strength of stimulus to rapidity of habit-formation. Journal of Comparative Neurology and Psychology 18, 459–482 (1908)

Stakeholder-Driven Business Process Management "An Evaluation of the Suitability, Adequacy and Effectiveness of Quality and Process Management"

Michael Jungen

Münzenberger Straße 6, 64354 Reinheim, Germany
michael.jungen@email.de

Abstract. In recent years, the variety of products and processes in the automotive and supplier industry has increased significantly. This has intensified the discussion about the practical applicability of quality and process management. Based on the interpretation of organizations as living, social, technical and economic structures, this evaluation provides one exemplary solution to the plight of the suitability, adequacy, effectiveness and reputation of quality/process management.

Keywords: Taylorism, two-valued logic, high variety, social anchoring, meta-principles and principles of system preservation, living organizational systems.

1 Introduction

In terms of consistency and depth of corporate quality/process management, the automotive and supplier industry is considered to hold a pole position. However, the dynamic development of customer requirements and hence a significantly increased complexity of products and manufacturing processes (e.g. for electric or hybrid vehicles) can be expected for the near future. This will put the promised benefit and also the organizational effectiveness and efficiency of established organizational solutions such as quality/process management to the test.

2 Initial Situation

It must first be noted that basic standards for quality/process management such as ISO/TS 16949 [1] do not actually require companies to apply stereotypes, but explicitly request system realization to be tailored to the corporate context. Considering the extensive liberty of interpretation the basic standards allow, it is even more astonishing how frequently the following characteristics of management systems can be found in the automotive and supplier industry:

- Major parts of industrial management systems show formally uniform structures far from any operational reality. An easily comprehensible external representation is often given preference over process instructions suitable for internal use.

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 244–252, 2011.

- The standards have nurtured such a large number of interpretation-based requirements that even experts get lost.
- Supplementary methods are often implemented in a technocratic approach and without effect on the realization and control of internal corporate processes.
- Internal and external system audits focus on compliance, thus driving organizational formalism instead of performance and improvement.
- Management reviews actually intended to evaluate suitability, adequacy and effectiveness of the management system are delegated to operational functions and are quite detached from the actual process quality, performance and costs.

etc.

Organizational reality often shows consequences such as:

- When faced with unsatisfactory development projects, customers no longer content themselves with system documents, but set up result-driven "task forces".
- Employees utilize detailed regulations and descriptions provided by the management system to actively delimit or justify their operative action.
- Established instruments for improvement, such as 8D or SIX SIGMA, only have limited impact on the organization (keyword: resistance to optimization)
- Suppliers cause significant problems in supply chains despite extensive development efforts such as product or process releases, audits, etc.
- Uninhibited growth of quality and process management systems that bind resources without showing any visible transparent and attractive use.

etc.

Many companies find themselves in a helpless position: "How could quality and process management lose its innocence like that", and, as a consequence: "how, considering the restrictions of a highly dynamic business environment and the limited resources, can quality and process management be changed so that it ...

- better supports long-term organizational performance instead of impeding it?
- provides an active contribution to improve process quality, performance and costs?
- maintains its continuous and long-term suitability, adequacy and effectiveness?

Any elaboration of the above queries requires a targeted look back at the evolution of modern quality/process management.

3 Origin and Evolution of Quality/Process Management

Early connections to quality/process management can be found in trade associations such as the Hanseatic League or medieval craft guilds. At that time, quality, performance and effort were mostly based on the confidence in the agents' expert knowledge and their social reputation based on their customers' positive experience. According to the common value concept of their customers, honorable tradesmen and craftsmen personally stood in for the quality of the goods they traded or manufactured. If a company violated established quality standards (and hence the principle of good faith), the entrepreneur himself was subject to disgracing public rituals or social exclusion which would actually threaten the economic survival of the company **[2]**.

The history of industrialization and hence that of modern-day quality/process management starts with the period of Enlightenment around 1750. The core issue during this phase of social restructuring was the dissolving of traditional relations between common people, aristocracy and clergy in favor of a new social understanding and self-understanding based on equal rights for individuals and rationality. Philosophically and humanistically, the political movement legitimized itself by the so-called two-valued Aristotelian logic of ancient Greece which was based on free speech without consideration of a counterpart's given context. This circumstance implicitly promoted all disciplines of analytical scientific work, and hence also the representation of complex processes as abstract deterministic models (mechanistic concept of effect). This spirit of the time gave the engineering sciences a major boost which was then followed by the development of modern production processes and the reproducible manufacturing of mass products (known as the Industrial Revolution).

The pioneers of this industrial development were Henry Ford and Frederick W. Taylor (1856-1915) who, for the first time in history, produced automobiles on interconnected assembly lines. This type of manufacturing is based on the principle of the division of labor. Complex procedures are divided into individual steps small enough to be performed by unskilled workers (referred to as Taylorism). Companies thus became less dependent on the resource of highly-skilled workers. Reduced labor costs and continuous rationalization then led to an enormous increase of productivity. However, the limitation of persons to an existence as a mere production resource forced to perform simple work procedures in endless repetition has consequences: Psychologically, such persons often experience the continuous non-use of a major part of their skills as a form of devaluation. Also, the high level of division of labor detaches persons from the work process and the result of their work. In conjunction with the monotony of repetitive processes, attention levels drop, ultimately leading to increased error rates (scrap rates, rework, accidents). Consequently, operators were hence primarily seen as risk factors. Companies counteracted this by professionalizing the training (or rather "drill") based on process descriptions and by installing end-of-line checks that were independent of the value-adding process. The transition from the former principle of trust to quality checks marked the introduction of the basic conditions and functionalities of modern quality/process management. Until the 1970ies, organizational productivity was ensured by a high level of division of labor, the separation of intellectual from operational work, independent quality checks and role assignments as qualified resource for operators. Quality/process management was hence established as the guardian of product and process compliance, ensuring the repeatability of vital production processes and rigorously sanctioning and eliminating any adverse influences on the basis of hierarchical structures.

Driven by individual customer demands, increasing product and process complexity, the call for shorter product life cycles, progressing globalization etc., industrial companies were forced to change from hierarchical-vertical to network-type organizational structures with a high level of participation by the persons involved. Examples of such developments are simultaneous engineering teams or partly autonomous production teams. Among the production resources, operative agents now often assume special positions, acting as intelligent process decision-makers, controllers and operators. In addition to the division of labor, personal resources such as motivation, cooperative skills and confidence in one's actions are gaining importance as drivers of productivity. Accordingly, organizations may now no longer be regarded as purely technical and economic structures. The concept of a living, social, technical and economic system now appears to be the more appropriate description of organizational reality. The formerly prevailing mechanistic concept of men and organizations defined as Taylorism is hence crumbling, with the trustworthiness of agents moving back into focus.

These changes also affect quality/process management. The "end-of-line" checks performed until the end of the 1960ies were gradually amended or replaced by "in-line" self-checks performed by the workers, preventive planning, zero-defect strategies, continuous improvement, business-process management, excellence approaches etc. Despite the progressive erosion of vertical hierarchies as implementation levers in an organization, the traditional doctrines of quality/process management often still suffice to fulfill the values associated with Taylorism. Examples:

a) System compliance ensures the functional integrity of a management system.
b) Organizational deficiencies are traced back to deviations in compliance.
c) Descriptions are effective and guide actions because of their mere existence.
d) Management system auditors are only effective if they are independent.

4 Challenges to Quality and Process Management

Considering the claim "stakeholder-driven process management", the situation stated above imposes a significant challenge on modern quality/process management: <u>to achieve a better understanding of how to bridge the gap between technical orientation and social anchoring/effectiveness within a company</u>.

As already mentioned, the automotive industry expects increases in the variety of products and processes. According to W. R. Ashby's "Only variety absorbs variety" **[3]**, complex systems such as organizations only have <u>one</u> option to respond to the requirement of increased external variety, and that is the internal adjustment of variety, which is a balancing act. (Variety being the level of complexity perceived by the agents. Complexity is reduced e.g. by suppressing, ignoring etc., which in terms of the actual variety equals an illusion of control **[4]**).

Traditionally, such challenges are met by intensifying the formal and directive-oriented quality/process management which defines effectiveness primarily as an issue of clearly defined process regulations, method standards or disciplinary restrictions and hierarchical execution of control.

<u>However, in view and because of the high level of variety</u>, tightening the formal constraints by imposing ever more accurate and detailed process and method requirements seems to be counterproductive in the long term. According to Paul Watzlawick **[5]**, any handling of a complex challenge following the rule of "more of the same" is a so-called 1^{st} order solution. According to that idea, the solution itself inevitably becomes the problem. The German poet Johann Wolfgang von Goethe made the point even clearer when he coined a phrase which could be translated as: "Dirt that we tread isn't hardened, but spread". In the automotive and supplier industry, the first signs of this logic of effect are clearly written on the wall.

Instead of trying to handle increasing variety solely by applying technical and methodical means, the professional utilization of man as newly discovered intelligent resource appears to be the method of choice. Accordingly, a deliberate and systematic increase of the involvement of agents as competent decision-makers, crafters and operators must be pursued to achieve optimal quality, performance and costs.

In view of the requirement for highly flexible thinking and a high number of variants of acting, static process descriptions are increasingly losing their character as guidelines for action. Since individual motivation and social behavior are gaining importance as performance factors in addition to technical expertise, it is becoming less easy to steer agents in the execution of their tasks.

Questions such as "how can motivated agents be reliably connected to processes and methods" or "what makes people highly motivated" and "how can we encourage their cooperative skills" have suddenly moved into the focus. The combination of economic, technical and social aspects as ideas of an adequate "alternative" may be suitable for the continued formalization of management systems. However, the resolution of implicit queries has so far not been a typical core competence of traditionally technology-driven companies.

If processes involving a high number of variants must be realized flexibly, reproducibly and economically and with due consideration of human complexity, the loss of control identified above must by all means be compensated for or balanced.

The complexity of the challenge described above may well be outlined by a technical metaphor: In the mid-1980ies, the combat-aircraft industry was faced with the requirement for increased maneuverability, i.e. the so-called dogfight capability in subsonic flight. This was to be achieved without sacrificing the previously acquired capability of stable supersonic flight. In order to meet the first challenge, aerodynamic lift had to be realized for the so-called unstable flight conditions. To achieve this, it was necessary to abandon the common principle of stable flight, where an aircraft resumes a stable flight position after manipulation of the controls. However, in an unstable flight condition, an aircraft is persistently subject to the risk of an imminent crash. The solution was found in the decentralization of the control logic. The pilot remains in control of the direction of flight. The fine adjustments during unstable flight conditions are made by a computer. This computer identifies early stages of (crash) movements occurring at short intervals and counteracts by immediately actuating the controls before the aircraft gets out of control. This ensures that a condition of a "permanently controlled crash" is maintained at all times, which is perceived externally and internally and fed back as a stable flight condition. When the aircraft accelerates out of a tight curve into supersonic flight, mechanical manipulation moves it out of the unstable flight condition, utilizing the advantages of stable flight.

The metaphor yields three findings: First, the challenge lies in the achievement of "one as well as the other" **[6]**, i.e. achieving one capability without sacrificing the other. Second, any operation in an unstable flight condition requires an intact closed-loop control of the following: check of flight position – execution of adjustments (micro-level) – feedback to pilot – change of flight direction (macro-level) and check of flight position again, etc. Third, the input of adjustment manipulations requires the due consideration of the development of potential critical flight conditions.

When applying this example to our field of interest, it becomes clear that a more technical-formal approach to quality/process management would always be helpful when the need for high operating speeds requires a standardization of the procedure (confer supersonic flight in passage above). With a view to the agents, the greatest challenge is to obtain and focus their attention. If flexible process action is necessary, however, a targeted decentralization of the control logic can be helpful. Put in simpler terms, this means, for instance, that operative team members monitor processes to generate the data they actively use themselves to fine-tune the processes. In addition, a high-frequency, condensed process-performance feedback will be set up for corporate management. They can use it as input for the identification of cross-functional improvement potentials or for the targeted reflection/preparation of a change in direction in their corporate strategy.

To ensure this control loop is carried out adequately without additional risks for the company, the operative agents must be confident in their actions, motivated, cooperative as well as attentive and flexible to change. This requires qualities such as transparency, appreciation and trust within organizations. The most important meta-principle **[7]** for system preservation is the non-denial of facts. Consequently, management must practice leadership behavior that focuses on the present, is visibly inspired by common sense, congruent and consistent. Leadership credibility is of particular importance in this context, because, as a result of context changes on the outside and the resulting variance in business decisions, oscillations between dynamic adjustments in direction and the formation of stable operation standards are likely to occur.

4.1 Summarizing Working Hypotheses

- - Division of labor and formal quality/process management alone cannot compensate for the increasing complexity without bringing about critical consequences.
- - The more complete quality/process-management systems are the less suitable for use they are in terms of a flexible detection /processing of priorities in operative management. Critical gaps are more likely to occur.
- - The principle of "trust" in quality/process management in combination with self-observation from a steering and operative perspective as well as an entrepreneurial assessment of the reasonableness of perceived conditions is on the rise again **[8]**.
- - Strong turbulence on the outside requires transparent, trustworthy leadership behavior on the inside, leaving room for the company's operative actors to develop and maintain their motivation, attentiveness and focus of attention.
- - In times of network-type organization structures, management systems are not effective because of the mere fact that they exist **[9]**. Irrespective of hierarchies, agents must be actively linked to agents and agents to actions **[10]**.
- - If quality/process management is to provide help with the oscillation of adjustments in direction and the formation of standards, it must be designed to support changes between the strategic outside perspective and operative inside perspective.
- - To design management systems with an optimal input/output ratio, an adequate "alternative" approach is required. Every reduction of rules must be balanced, as it carries the risks of reduced attentiveness, uncertainty and overstrain.

- \- The stakeholders of quality/process management in this context are: First of all the organization as a living, socio-technical construct **[11]** that follows the basic principles of system preservation, and second, the individuals (agents) linked to it as fractal parts of the organization, pursuing their respective personal needs **[12]**.

The extended view on quality/process management is referred to as a "more systemic" view in the following [7]. This is supposed to show that neither a claim for completeness is made nor a systematic construct on a higher level is aimed at.

5 A "More Systemic" Quality/Process Management

How exactly must a more systemic view on quality/process management be structured or continuously organized in an ever more complex technological environment? In the following, this will be shown in the form of ***examples*** of basic system-preservation principles [7] and their connection to the agents' requirements structure [12], which will then be translated into organizational implementation methods.

1. **Living organizational systems** strive for the "preservation of the system's existence" [7]. They construct reality by recursively processing rule-formulation **[13]** and by identifying deviations between target and actual and by following rules **[14]**, that is by continuously developing a stable behavior **[15]**. Rule-formulation and rule-following form a purpose-oriented, inseparable complementary pair.

Quality/process management taps this potential, for example by overcoming the formal separation **[16]** from operative performance/cost management (principle of equal membership [7]). As W. E. Deming **[17]** demonstrated as early as in 1950: In the end, quality is productivity, which in turn ensures survival! Quality can hence not be separated from performance and costs and is therefore part of the non-delegable management and leadership responsibilities.

Implemented on the organizational level, this means for example that ...

– ... the responsibility for quality, performance and costs is pooled on management level. Management itself is the company's quality representative.
– … an observation of relevant external environments, such as customers and their changing requirements, is enabled and an effective routine for self-observation and reflection is set up **[18]**.
– … real-time monitoring is employed to supply the organization with significant information on target/actual deviations in quality and process performance.
– … quality is reintegrated into corporate performance/cost controlling (and not vice versa) and its effect on the balance sheets is made more transparent **[19]**.

2. **Living organizational systems** are oriented towards "system growth/propagation [7]. To achieve a best-possible fitness for purpose for management systems, a field of tension must be explored which is best characterized speaking with the words of A. de Saint-Exupéry "Perfection is achieved, not when there is nothing more to add, but when there is nothing left to take away" and A. Einstein "Everything should be made as easy as possible, but not easier."

Quality management taps this potential, for example, when the competition for an even more sophisticated interpretation of standards and even smarter corrective measures in resource-consuming audits is abandoned. Sufficiently complex systems are always either partially incomplete and/or inconsistent **[20]**.

Implemented on the organizational level, this means for example that ...

- … rigid system descriptions are replaced by highly-adaptive and subject-oriented presentations of communication and relation structures **[21]**.
- … the utility value of ever more complex management systems is increased by making them leaner. The leaner organization is to be balanced by, e.g., competence management (keyword: non-knowledge) and network structures.

3. **Living organizational systems** are oriented towards the "development of an immune system" [7]. Based on the principle of priority of the stronger commitment, a company's immunity to crises is strengthened when the commitment of individual system elements to ensure everyone works with and for each other is adequately appreciated and friction loss is hence actively prevented.

Quality/process management taps this potential, for example when the belief in the interconnectedness of effectiveness and compliance, as well as possibly developed fantasies of omnipotence in terms of a central regulation and monopolist form of management systems are abandoned.

Implemented on the organizational level, this means for example that ...

- … so-called preventive management approaches and the illusory safety they convey, which is based on emergency scenarios, is replaced by a culture/structure of attentiveness and crisis responsiveness **[22]**; furthermore a cross-cultural utilization of situation potentials and dealing with the unexpected is trained.
- … a "culture of trust" regarding the handling of mistakes is established [6], as non-compliance can be a sign of organizational vitality. This bridges the gap to the agents' individual requirements structures.

6 Summary and Outlook

As illustrated in the previous passages, the meta-principles and the system-preservation principles can be used to re-think, configure and structure the concrete anchoring of quality/process management in organizations without resorting to hierarchical instruments and formal reinforcement. Here, the personal variety of agents is not considered mainly a risk for process reliability and performance that is to be eliminated, but as a directly usable resource on the road to overall optimal management systems for quality, performance, costs. The key statement is that the individual's commitment, cooperative skills and confidence in his/her actions can be successfully integrated and translated into corporate structures. This will make it possible to realize and operationalize potentials in terms of flexibility and speed required for completing operative and control tasks in the company. This in turn promotes the vital recursion of "Culture follows structure and structure follows culture" – which is stable quality/process management under "unstable flight conditions"!

References

[1] International Automotive Taskforce, DIN ISO/TS 16949:2009, Qualitätsmanagementsysteme. Beuth Verlag (2009)
[2] Warzecha, B.: Problem: Qualitätsmanagement. Verlag für Planung und Org. (2009)
[3] Ross Ashby, W.: An Introduction to Cybernetics. Chapmann & Hall Ltd, London (1956)
[4] Stüttgen, M.: Strategien der Komplexitätsbewältigung in Unternehmen. Haupt Verlag (Auflage 2, 2003)
[5] Watzlawick, P.: Lösungen. Verlag Hans Huber (Auflage 5, 1997)
[6] Ferrari, E.: Wege aus dem Dilemma, Ferrari Media (2010)
[7] Sparrer, I., von Kibéd, M.V.: Klare Sicht im Blindflug. Carl-Auer Verlag (2010)
[8] Osterloh, M., Weibel, A.: Enterprise 2.0. Gabler Verlag (2006)
[9] Koch, M., Richter, A.: Wege aus dem Dilemma, Oldenbourg Verlag, 2. akt. und erw. Auflage (2009)
[10] Simon, F.B.: Einführung in die systemische Organisationstheorie. Carl-Auer Verlag, Heidelberg (2007)
[11] Bammé, A.: Die Neuordnung des Sozialen durch Technologie. Metropolis-Verlag (2007)
[12] Grawe, K.: Neuropsychotherapie. Hogrefe Verlag (2004)
[13] Wittgenstein, L.: Tractatus logico-philosophicus. Dover Publications, Inc., New York (1999)
[14] Wittgenstein, L.: Philosophical Investigations. Wiley-Blackwell (2009)
[15] von Foerster, H.: Sicht und Einsicht. Carl-Auer Verlag (2006)
[16] Brown, G.S.: Laws of Form (Limited Edition). Cognizer Co., Portland (1994)
[17] Edwards Deming, W.: Out of the Crisis. MIT Press, Cambridge (2000)
[18] Wimmer, R.: Führung und Organisation, Revue für postheroisches Mgt., Heft 04 (2009)
[19] Kamiske, G.F.: Rentabel durch TQM. Springer, Heidelberg (1996)
[20] Lau, F.: Die Form der Paradoxie. Carl-Auer Verlag (2006)
[21] Fleischmann, A.: What is S-BPM?, S-BPM ONE (2009)
[22] Weick, K.E., Sutcliffe, K.M.: Managing the Unexpected. John Wiley & Sons, Chichester (2007)

Evaluation of a Mobile AR Tele-Maintenance System

Michael Kleiber and Thomas Alexander

Fraunhofer FKIE,Neuenahrer-Str. 20, 53343 Wachtberg, Germany
{michael.kleiber,thomas.alexander}@fkie.fraunhofer.de
http://www.fkie.fraunhofer.de

Abstract. Despite increased bandwidth in world-wide networks there are still situations in which only limited bandwidth is available for tele-collaboration between a technician maintaining a machine and an expert at a remote location. In this case, tele-cooperation is usually supported by synchronous audio but only asynchronous video exchange. We present an alternative approach for such a collaborative maintenance task. By utilizing techniques and technologies from Augmented Reality (AR) applications our approach can provide a synchronous shared visual context for the collaborators without a direct video link. The effectiveness of our experimental system for a tele-maintenance task was evaluated in a usability study.

Keywords: augmented reality, system design, evaluation.

1 Introduction

Technologies for audiovisual remote cooperation have been available for decades. Despite increasing network bandwidth there are still many situations where only limited data exchange is possible. Yet, powerful mobile devices and data storage allow a technician to access a comprehensive database of support information by means of a portable computer system. It can also provide step-by-step visual instructions. By utilizing AR-technologies additional visual information about an assembly or a maintenance procedure can be integrated into the real scene [1]. However, such systems are limited to standardized, well-known procedural tasks. Non-standard problems often require a remote expert supporting the technician during the maintenance task.

In these situations an interactive cooperation between a remote expert and a local technician is not possible because the available network does not provide sufficient bandwidth for a synchronous transfer of audiovisual information. The lack of visual information requires the technician to describe any visible malfunctions verbally. It also requires the expert to guide and direct the technician by means of verbal descriptions. However, sharing a visual space through the use of video improves communication and interactive cooperation [4]. Our approach to solve the dilemma between the need for visual information exchange and a narrowband network is based on a virtual reconstruction of the maintenance object through the use of AR techniques.

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 253–262, 2011.

A typical tele-maintenance session consists of two interactive phases: Problem analysis and problem solving. In both phases a technician benefits from support by an expert at a remote location. In the problem analysis phase the expert can direct the technician towards identifying potential causes of the problem. During the following problem solving the expert can advise the technician. It is assumed that the support of the expert will increase the performance of the maintenance work. However, usual tele-cooperation methods and systems limit interactivity and, thus, efficiency and effectiveness of the joint maintenance work.

2 Earlier and Related Work

Earlier and related work can mostly be found in the application of AR for industrial purposes. There are also large collaborative research consortiums such as ARVIKA [3], STAR [14] and ARTESAS [5] which investigate the use of AR for almost all aspects of manufacturing. The according research activities describe rather elaborated system concepts which involve multiple cameras, multiple computers and are in general quite complex.

Utilizing AR for maintenance tasks has slightly different requirements. Henderson and Feiner categorize maintenance as consisting of activities involving the inspection, testing, servicing, alignment, installation, removal, assembly, repair, overhaul, or rebuilding of human made systems [7]. In these categories, assembly tasks have received the most attention. These tasks can range from assembling aircraft wire bundles [2] to assembling medical equipment for minimal invasive surgery [13].

However, in most of the related work tele-cooperation is of only minor importance. The AR applications rather resemble an extended electronic handbook which the technician uses without additional help by an expert. The user-friendly creation of AR scenes consisting of a set of maintenance instructions is therefore an important topic in the work of the research consortia cited above as well as in other projects [10].

The problem of network connections with limited bandwidth between the collaborators of a tele-maintenance session has hardly been considered. Although it seems obvious to limit data transfer to a numerical object ID and spatial information, no similar approach describing a tele-maintenance system consisting of an AR system for the technician and a VR system for the expert was found in our literature research. Therefore, the following contribution addresses the description of the concept of tightly interconnected AR/VR systems for tele-maintenance and in the practical evaluation of our prototype implementation.

3 System Concept

The system concept described in this section addresses the problem of providing a shared visual space for a technician and an expert collaborating via a narrowband data connection. It is an alternative to the commonly used video transmission requiring a lot of bandwidth. The expert is provided with an egocentric view of the technician so that instructions and inquiries can be formulated relative to the technicians own view. Instead of transmitting a video stream our system first identifies the machine parts in the view and then transmits the ID of these parts as well as their location and orientation in a machine-specific coordinate system. The expert uses a VR system which reconstructs the 3D view from the technicians point-of-view (Fig. 1). A precondition for this concept to work is that 3D models of the machines and their subparts are available and their real world counterparts are identifiable.

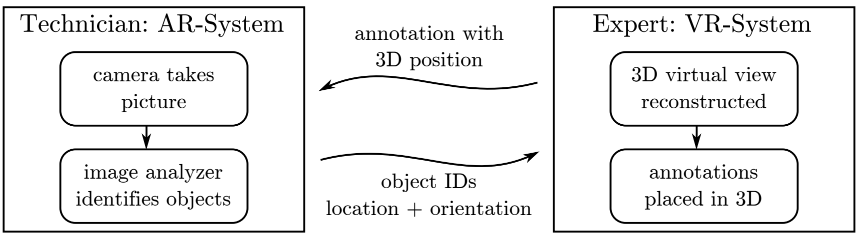


Fig. 1. The basic concept underlying our AR tele-maintenance system

In our case, the expert uses a desktop VR system to view and interact with the virtual 3D view. Besides the virtual view from the technician the expert can examine an interactive model of the maintenance object from an arbitrary viewpoint. This way he can explore the object interactively, e.g. to plan the repair of the machine.

For the AR system of the technician we first proposed using a head-worn display which frees the hands of the technician. However, the weak acceptance of such displays by actual maintenance personal prevented this approach. Instead, an AR system was constructed using a touch screen mounted on a multiple joint arm. The camera of the AR system was mounted on the backside of the screen so that the screen provides a "magic window" to the real world.

4 Implementation

We decided to use the XML-based XMP protocol for the interconnection of the systems of the expert and the technician. The protocol is the basis of many instant messaging services and was designed to be extensible. Similarly to e-mail a user is registered with an XMPP service provider. A user can exchange messages and presence information as well as arbitrary data with other users using any client software which implements the XMP protocol. The protocol is open, standardized and extensible. Since XMPP is text-based and all messages are relayed through a server the transmission of data with real time constraints must be carried out through a separate channel, e.g. audio and video data should be transmitted using the RTP. However, the XMP protocol can be used for signaling and control [12].

The desktop VR system of the expert as well as the mobile AR system of the technician are implemented using the open source OAW VR toolkit [19]. The toolkit is based on the X3D ISO standard [8] and includes many components of realtime 3D applications. However, it does not describe a way to exchange data through a network. There is an attempt by the *X3D Networking* working group to amend the X3D specification to allow such cases. They try to provide networking capabilities at the node level so it is easier for content developers to use networking with no or only limited scripting [18]. However, the proposed amendments, when implemented on top of raw network sockets, do not address issues such as NAT traversal or user identification and authentication. We implemented an X3D node called `XmppSensor` which follows the proposal of the *X3D Networking* working group and is based on an the XMPP extension Jabber-RPC [15]. This extension specifies how remote procedure calls (RPC) can be made through XMPP. We decided to not strictly adhere to the Jabber-RPC standard because it only includes support for basic data types. Instead, we directly encode the X3D types. Thus, the event `vpPos` with type `SFVec3f` and value `(0, 1, 3)` will be encoded as follows:

```
<iq type='set'
  from='remoteUser@xmppServer.local.net/x3d-rpc'
  to='userName@xmppServer.local.net/x3d-rpc'>
  <query xmlns='jabber:iq:rpc'>
    <methodCall>
      <methodName>vpPos</methodName>
      <params>
        <param><value><SFVec3f>0 1 3</SFVec3f></value></param>
      </params>
    </methodCall>
  </query>
</iq>
```

The OAW toolkit can be extended by plugins so that additional functionality besides the core X3D specification can be used. Since the current X3D specification does not contain a method for integrating external sensors like cameras or tracking devices we added a node for the acquisition of live camera images which inherits the interface of the X3D `PixelTexture` node. The camera image is processed by the AR tracking software. It analyzes the image in order to identify and locate maintenance objects. The software also calculates the camera position in relation to the tracked objects. This information is transmitted to the expert and is used in the reconstruction of the technicians view. Since we are using ArToolkitPlus which employs marker tracking to estimate the camera pose, the tracked objects in our case are the markers [17]. We use multiple BCH-markers with ID numbers which can be related to specific maintenance objects or parts. The 8×8 BCH-markers we employ allow 4096 different IDs which is sufficient to distinguish all subparts of a typical maintenance object.

The camera image is also displayed on the technician's touch screen. The technician can select or point at locations in the image using his fingers or a pen. This information is also transmitted to the expert and visualized. Likewise, the expert can indicate a location on the 3D model and optionally annotate it which will be visualized as an augmentation of the live camera image on the touch screen of the technician. Additionally, the expert can select specific steps of available maintenance procedures and transmit them to the technician.

The collaboration between the expert and the technician can be solely based on the virtual view of the scene as it provides an interactive shared visual context. Nonetheless, when a detailed view of a specific part of the maintenance object is required, the expert can request a photo snapshot with high resolution.

5 Evaluation

The developed tele-maintenance system was evaluated for a practical maintenance task in a short usability experiment. This allows insights into the usability of the AR/VR system. In our case, the system was compared to a conventional system. For quantifying overall performance, completeness and correctness of the maintenance work and the time to complete the task were measured and analyzed. We also captured the subjective cognitive workload by a standardized questionnaire (NASA-TLX as in [6]).

Based on the above we formulated the following hypotheses for the comparative evaluation of the develped AR/VR tele-maintenance system ($S_{\mathrm{AR/VR}}$) and a conventional system (S_{conv}):

H_1 The overall time required to complete the maintenance task will be less when using $S_{\mathrm{AR/VR}}$ compared to S_{conv}.

H_2 The expert will need to give less instructions to the technician or ask him less questions when using $S_{\mathrm{AR/VR}}$ compared to S_{conv}.

H_3 When using $S_{\mathrm{AR/VR}}$ the technician will need to ask less questions or give fewer descriptions compared to when using S_{conv}.

H_4 The cognitive strain on the technician is lower when using $S_{\mathrm{AR/VR}}$ compared to S_{conv}.

5.1 Method

Both tele-maintenance systems adhered to the same general conditions: Expert and technician could use a technical handbook, they could communicate using an audio connection with GSM quality and a maximum latency of 0.5 seconds and they could exchange data using a GSM channel (14.4 kBit/s). Because of these conditions, video could not be transmitted and even the transmission of a JPEG compressed medium resolution image took 20 seconds. However, the available data rate is sufficient for transmitting machine part IDs as well as their location and orientation relative to the camera. Both systems were implemented on the same hardware setup.

In the conventional setup the technician was able to take snapshots of the current view and send them to the expert. The expert was able to annotate the received photo and return it to the technician. In the AR/VR setup the current viewpoint position was continuously transmitted to the expert and used to reconstruct the view of the technician. The expert was also able to request a photo snapshot of the current view. Annotating the photo was not possible, but annotations could be placed on the interactive 3D model which were then displayed as augmentations on the touch screen.

5.2 Maintenance Task

The maintenance task was to replace the camshaft of an automobile engine. For this task a considerable part of the engine had to be disassembled following a specific order. The manual skills required for the task were not demanding, but it was rather difficult to complete the task by the technical handbook only. Broken, dirty or missing parts enhanced the difficulty. This was the case with the engine we used because marks required for aligning the cylinders were missing (Fig. 2). For the evaluation the single steps and the technical sketches scattered in the handbook were combined in consecutive order.

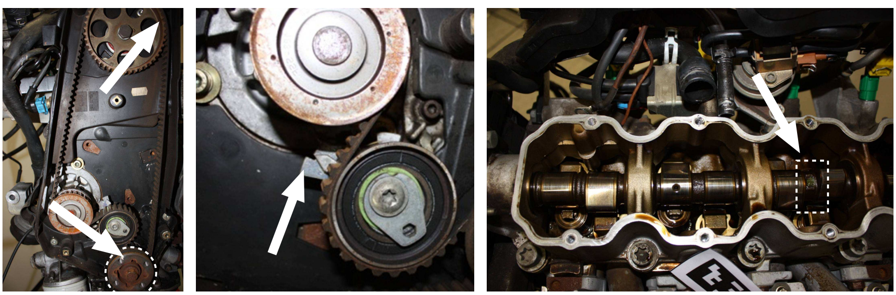


Fig. 2. Before removing the vee-belt the first cylinder has to be brought to the top dead center by aligning the marking on the gear wheel of the camshaft with another marking on the engine block

5.3 Experimental Design and Procedure

We anticipated learning effects in the maintenance task and therefore used a between-subjects experimental design. The two systems served as independent variable. The task completion time was measured as the first dependent variable. We also recorded the audio communication of the expert and the technician for later analysis. The supervisor and instructor of the experiment served as the expert during the tele-maintenance session. The expert was seated in a separate room during the actual experiment so that no means of communication besides through the tele-maintenance system was available. The participants acted as technicians.

We recruited 10 male subjects of similar age (25 – 32 yrs) from the institute's staff who were naive of the experiment's hypotheses. None of the participants had prior experience in automobile maintenance or repair. To prevent further artifacts because of group differences two pretests were performed. To test the mechanical and technical comprehension we used the MTV test by Lienert [11]. The hose-figure test by Stumpf and Fay was used to quantify the spatial abilities of the subjects [16]. The participants were assigned to two groups according to their combined rank in these two pretests.

Prior to the actual maintenance session the participants received written instructions about their general task. This also included an introduction of the functions of the tele-maintenance system. The participants were allowed to inspect the engine thoroughly and the camshaft housing was pointed out. The available tools were ratchets and wrenches of different size, pliers and screwdrivers. The participants wore a protective coverall and gloves during the experiment and used a pen to interact with the touch screen. Prior to the start of the experiment they received the combined repair instructions. The expert opened the tele-maintenance session a few seconds later. The experiment ended when the participants finished unscrewing the camshaft housing. Directly afterwards they rated their subjective workload by the NASA-TLX questionnaire.

5.4 Results

For evaluating the first experimental hypothesis the overall time of the maintenance procedure was analyzed. For H_2 and H_3 the recorded audio communication was analyzed. A talk cycle for the expert was detected when the audio amplitude was louder than -26 dB (with 0 dB counting as the maximum) for at least 0.07 seconds and ended when the amplitude was lower than -26 dB for 0.75 seconds. Since the level of noise at the technicians work place was louder the required amplitude was increased to -22 dB for the recorded audio of the technician (Table 1).

Although the overall maintenance duration was slightly smaller (36.00 minutes versus 37.44 minutes) for the conventional tele-maintenance system the effect is not significant (standard deviation for S_{conv} is ca. 5.8 minutes and for $S_{\text{AR/VR}}$ ca. 6.4 minutes).

The amount of speak cycles of the expert for the two groups are also similar. However, the standard deviation of the number of talk cycles seems to be different. For S_{conv} the standard deviation is 47.58 cycles and for $S_{\text{AR/VR}}$ it is 20.78 cycles. An analysis of the similarity of the variance using the Levene test revealed weak significance (F(1,8)=6.13, p=0.038). The overall talk duration of the expert for the two groups differs by ca. 160 seconds but the standard deviation is almost as high (for S_{conv} sd = 182.71 s, for $S_{\text{AR/VR}}$ sd = 151.77 s). This means there is no statistically significant difference in the talk duration of the expert. The Levene test does not show a significant difference between the two error variances.

Similar to the analysis of the talk durations and cycles of the expert, the analysis for the technician's recorded audio can be done. The talk duration differs by ca. 65 seconds. However, because of the large standard deviation of the data (for S_{conv} sd = ca. 100 s, for $S_{\text{AR/VR}}$ sd = ca. 96 s) this difference is not

Table 1. The results of the experiment

participant	system	duration [min]	talk cycles [num] expert	talk cycles [num] technician	talk duration [s] expert	talk duration [s] technician
5	S_{conv}	44.58	253	258	594,48	200.77
10	S_{conv}	33.25	222	270	469.04	258.98
3	S_{conv}	33.59	162	132	187.81	101.05
7	S_{conv}	38.88	219	281	520.91	344.94
4	S_{conv}	29.69	137	182	223.05	122.17
average	S_{conv}	36.00	198.6	224.6	399.06	205.58
1	$S_{\text{AR/VR}}$	42.57	201	181	722.85	345.93
9	$S_{\text{AR/VR}}$	30.04	189	151	669.75	309.06
8	$S_{\text{AR/VR}}$	34.01	229	230	611.05	360.28
2	$S_{\text{AR/VR}}$	45.42	172	179	362.96	137.39
6	$S_{\text{AR/VR}}$	35.13	198	196	449.59	204.33 (1.04)
average	$S_{\text{AR/VR}}$	37.44	187.4	197.8	563.24	271.4

significant. The number of talk cycles is similar for both systems but just as is the case for the expert the variance of it varies significantly. The standard deviation for S_{conv} is 64.7 cycles and for $S_{\text{AR/VR}}$ it is only 28.8 cycles. According to the Levene test this is a significant difference ($F(1,8) = 6.736$, $p=0.032$).

After they finished the maintenance task the participants answered the NASA-TLX questionaire. There was no significant difference for the two systems.

5.5 Discussion of Results

The usability experiment was performed in order to ensure that the developed system works in a practical setting and to gather first insights about the applicability of the AR/VR tele-maintenance system in contrast to a conventional tele-cooperation system.

The observed overall maintenance time did not differ between the two systems so that hypothesis H_1 must be discarded. The same is true for hypotheses H_2 and H_3 since there is no significant difference in the number of talk cycles and in the length of the talk duration of the expert and the technician. However, the significantly smaller standard deviation of the number of talk cycles for the expert as well as for the technician could mean that when using the AR/VR system individual differences of the subjects have less impact on the results. Yet, when evaluating the performance data the small sample size of the experiment must be kept in mind.

In the subjective rating (NASA-TLX) all but one participant answered the question regarding the achieved performance with "very good" or "perfect". This could be an indication that the task was too easy and that therefore the expert was not really needed to fulfill the task. Furthermore, the participants' motivation was high (according to the NASA-TLX scale but also when asked directly). These factors, combined with the availability of the repair handbook and the technical background of the participants has led to the situation that the expert support was helpful but not necessary.

6 Conclusion and Future Work

The tele-maintenance system is proposed to be useful and applicable for non-standard situations when a predetermined work plan is not available. However, recreating this setting completely for an evaluation in a comparative experiment is not possible. Future evaluations should therefore be specifically designed so that the technician is dependent on the expert.

We want to extend the system to support the acquisition and display of stereoscopic images. When adjusting the stereoscopic display parameters accordingly the expert can view the scene and perceive correct depth [9]. Furthermore, we want to adjust the authoring process of the work steps shown in the AR/VR systems. Currently, manually generated animations are used. This is very time and labor consuming and not possible for all of the different maintenance procedures. We therefore plan to integrate methods to enable the expert to create these animations during the tele-maintenance session.

Acknowledgments. We gratefully acknowledge funding of the project by the German federal office of defense technology and procurement (BWB), Team T-5.2.

References

1. Barfield, W., Baird, K., Shewchuck, J., Ioannou, G.: Applications of wearable computers and augmented reality to manufacturing. In: Fundamentals of Wearable Computers and Augumented Reality, pp. 695–713. Routledge, London (January 2001)
2. Curtis, D., Mizell, D., Gruenbaum, P., Janin, A.: Several devils in the details: making an AR application work in the airplane factory. In: Proceedings of the International Workshop on Augmented Reality: Placing Artificial Objects in Real Scenes: Placing Artificial Objects in Real Scenes, IWAR 1998, pp. 47–60. A. K. Peters, Ltd, Natick (1999)
3. Friedrich, W.: ARVIKA-augmented reality for development, production and service. In: Proceedings of International Symposium on Mixed and Augmented Reality, ISMAR 2002, pp. 3–4 (2002)
4. Fussell, S.R., Kraut, R.E., Siegel, J.: Coordination of communication: Effects of shared visual context on collaborative work. In: Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work, pp. 21–30 (2000)
5. Haberland, U., Brecher, C., Possel-Dlken, F.: Advanced augmented reality-based service technologies for production systems. In: Proceedings of the International Conference on Smart Machining Systems (March 2007)

6. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task load index): Results of empirical and theoretical research. Human Mental Workload 1, 139–183 (1988)
7. Henderson, S., Feiner, S.: Evaluating the benefits of augmented reality for task localization in maintenance of an armored personnel carrier turret. In: 8th IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2009, pp. 135–144 (2009)
8. ISO/IEC 19775-1:2008: Information technology – Computer graphics and image processing – Extensible 3D (X3D) – Part 1: Architecture and base components. ISO, Geneva, Switzerland (2008)
9. Kleiber, M., Winkelholz, C.: Distortion of depth perception in virtual environments using stereoscopic displays: quantitative assessment and corrective measures, vol. 6803. pp. 68030C–11. SPIE, San Jose (February 2008)
10. Knopfle, C., Weidenhausen, J., Chauvigne, L., Stock, I.: Template based authoring for AR based service scenarios. In: Proceedings of Virtual Reality, VR 2005, pp. 237–240. IEEE, Los Alamitos (2005)
11. Lienert, G.A.: Mechanisch-Technischer Verständnistest. Verlag für Psychologie - Hogrefe, Göttingen (1964)
12. Ludwig, S., Beda, J., Saint-Andre, P., McQueen, R., Egan, S., Hildebrand, J.: XEP-0166: Jingle. Draft (December 2009), http://xmpp.org/extensions/xep-0166.html
13. Nilsson, S., Johansson, B.: Fun and usable: augmented reality instructions in a hospital setting. In: Proceedings of the 19th Australasian Conference on Computer-Human Interaction: Entertaining User Interfaces, OZCHI 2007, pp. 123–130. ACM, New York (2007)
14. Raczynski, A., Gussmann, P.: Services and training through augmented reality. In: Proceedings of the 1st European Conference on Visual Media Production (CVMP), pp. 263–271 (March 2004)
15. Saint-Andre, P.: XEP-0009: Jabber-RPC. Final Standard (February 2006), http://xmpp.org/extensions/xep-0009.html
16. Stumpf, H., Fay, E.: Schlauchfiguren. Ein Test zur Beurteilung des räumlichen Vorstellungsvermögens. Hogrefe, Göttingen (1983)
17. Wagner, D., Schmalstieg, D.: Artoolkitplus for pose tracking on mobile devices. In: Proceedings of 12th Computer Vision Winter Workshop, pp. 139–146 (2007)
18. Web3D Consortium: X3D Network Working Group Website (2007), http://www.web3d.org/x3d/workgroups/x3d-networking/
19. Winkelholz, C., Alexander, T., Weiß, M.: Open ActiveWrl: a middleware based software development toolkit for immersive VE systems. In: Proceedings of the Workshop on Virtual Environments, pp. 321–322. ACM, Zurich (2003)

Knowledge in Digital Decision Support System

Erika Matsak[1] and Peeter Lorents[2]

[1] Tallinn University, Narva road 25, 10120, Tallinn, Estonia
[2] Cooperative Cyber Defence Centre of Excellence, Estonia
erika.matsak@tlu.ee, peeter.lorents@mil.ee

Abstract. Digital circuits have been studied, which can support the decision-making process. The circuits allow realizing two types of inference rules: modus ponens and modus tollens. Inference process uses knowledge that comes from a description of situations, and axioms, which comes from the experts.

Keywords: Decision making. Logical deduction. Digital circuits.

1 Introduction

Managing real situations requires operating with growing amounts of information during ever decreasing amounts of time. At the same time, the quality of decisions must not suffer. The quality is ensured by using correct situation descriptions and a correct inference process. The latter, in turn, requires the use of correct inference steps, which give a correct result when used with correct premises. This brings us to the problem: how to support the human decision process with suitable technological solutions that would enable operation with large information amounts during a short amount of time. These suitable technological solutions can nowadays be divided into two broad categories: software and hardware solutions. The credibility of these solutions is based on corresponding theoretical work.

For example, when discussing decision processes, we can refer to situation description [1], event description [2], decision trees [3], statistical methods [4], etc.

In several cases, (custom built) hardware solutions have proven to be faster and more reliable than corresponding software solutions. For example, this has been claimed about neural networks. Based on this, it is interesting to study the use of digital circuits for implementing the correct logic inference steps needed in the decision process. This is the area that we cover in this publication.

2 Description, Formulas, Knowledge - General Overview

A substantial part of situation descriptions is formed by statements that generally contain the following content:

- Some object has or lacks a certain property.
- Some objects have or lack a certain relation.

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 263–271, 2011.

In logic, such statements can be represented by so called atomic formulas or their negations. These formulas can be (in the frame of classical logic) either true or false, depending on whether the describer has correctly understood the state of affairs.

Often the situation descriptions do not contain everything that we need to know or should know.

Example. While describing a disaster area or combat zone, one observer has determined that it is possible to get from location X to location Y. Another observer has determined that it is possible to get from location Y to location Z. If the decision maker

- knows that the binary relation R, which represents the existence of a safe route between two locations, is transitive (meaning $R(X,Y)\&R(Y,Z) \supset R(X,Z)$), and
- can apply the inference step that comes from a rule called *modus ponens* $\dfrac{P \quad P \supset Q}{Q}$,

then he can state that it is possible to get from location X to location Z and he can make the corresponding decision.

We can guess that until that application of the inference step, the decision maker *did not know* that it is possible to get from X to Z. But now he *knows.* More than that, he knows something that actually meets the real conditions in the context of the situation at hand, without having to observe it locally and without having information about this *specific case*. It is a good demonstration of the need for logically correct inferences: it often allows reliably acquiring missing parts of the situation description, even if there is no direct source for the correct information.

When relating knowledge with formulas [5], we should presume that if in some situation some aspects are not as they were reported, then the truth value of the formula with the corresponding knowledge should be 0. However, if everything "checks out", then the truth value should be, for example, 1. (We are using expressions like "should be 0" and "for example, 1", since the truth values *do not have to be numbers*! For example, we can also use the capital first letters of the words *false* and *true*, or even open subsets of some topological space [6]).

In real situations, it is often possible to build a formula based on various pieces of available information, but we cannot verify its correctness. In other words, we may have the formula, but it lacks (for the moment) a truth value of true or false. In such situations it may be useful to use so called tri-valent logic [see, for example, 7, 8, 9], where a third truth value is used (for example, *not determined* or *indeterminable*).

Therefore, we can say that the decision maker may possess knowledge (for example, from observer reports) and corresponding formulas, of which

- some are correct and some are not (since the observer could make a mistake, information transmission could introduce errors, etc.), while
- in case of some formulas it is not clear, whether they are true or false.

In summary, in the decision making process some pieces of knowledge are related to atomic formulas and the truth values of these formulas (and the status showing whether or not a truth value exists). The decision processes are often divided into different parts (for example, because some headquarters has divided decisions based on staff section role). It is therefore useful to "label" or "index" the above mentioned objects, in order to be able to "retrieve" them when needed.

3 Knowledge in a Digital Circuit

In this work we use the basic concepts and results of Lorents [5, 10, 11, 12, 13] to deal with knowledge. Therefore, we call *knowledge* every such ordered pair ⟨A,B⟩, where A is the notation (symbol, sign, etc.) for B and, at the same time, B is the denotation (meaning) of A. In that case we say that A and B have a *notation-denotation relationship* and we represent this as A∫B. The binary relation "∫" has several natural algebraic properties, such as well-foundedness, a specific anti-symmetry, and so called ordinary transitivity [5, 10-13]. The representation of knowledge in the digital circuits discussed below is based on the latter. For the digital circuit, knowledge is an ordered pair where the denotation is some situation description text in the database (some atomic formula or its negation) and the notation is the corresponding ID. Next, the ID is related to a specific block in the digital circuit schema (represented in Figure 2) by the block's address. Therefore, we have the following chain of notations and denotations: (*block address∫ID*), (*ID∫corresponding text or atomic formula in the database*), (*text∫the situation described by the text*). From the transitive property of the relation "∫" we get (*block address∫situation*). This gets us finally the knowledge: ⟨block address, situation⟩. To some extent the above described can be related to the knowledge handling approaches from semiotics [see, for example, 14].

As mentioned above, one way to obtain correct knowledge is the application of the logical deduction apparatus (incl. correct and well-formed inference steps) [see 13, 15]. The technological realization of this has a role for both software and hardware, specifically the digital circuits described below.

4 Software

The implementation of the digital circuits below presumes the existence of suitable software. The software must enable inserting (situation description related) knowledge, the corresponding status and truth values. In addition, it must be possible to insert the axioms (provided by experts) that are required for the implementation of the deduction (inference) that supports the decision making process.

In order to achieve this we can use an SQL database and form three simple tables. The first table, A1, will contain knowledge in text form. Then a unique ID is automatically generated. This table must contain all the knowledge that will be operated with. Thereat, the ID and the corresponding knowledge (or the formula that represents it) are in a notation-denotation relationship (in other words, they form a new piece of knowledge on their own).

The second table, A2, will contain the expert-provided axioms. For example, the experts can claim that *if* there is a route from X to Y *and* at the same time there is a route from Y to Z, *then* there must be a route from X to Z, or shorter: R(X,Y)&R(Y,Z) ⊃

R(X,Z). However, they may not tell us if there is or is not a route *specifically* between X and Z. In this work we only use axioms that are represented by implications (or formulas in the form of P⊃Q). However, in this work we can only form implications based on knowledge that is present in table A1. In order to insert these axioms into table A2 we use the following quartet: ⟨knowledge1 ID, knowledge2 ID, truth value 1, truth value 2⟩. For each inserted axiom a unique ID is generated.

The third table, A3, is filled (at the start of the decision process) by verifying that the information meets reality: is the inserted knowledge correct or not, or is it impossible to verify. If the knowledge is correct, then a 1 is inserted, if not correct, then a 0 is inserted. If verification was unsuccessful, then a NULL is inserted (basically, this "empty value" or NULL represents the third truth value discussed previously.

Since the digital circuits used in this work operate with only two distinct states (0 and 1), we must find a suitable solution for representing the NULL state in "electronic" form. In this work we have used the following solution:

A view A4 is automatically generated, which includes all of table A1 and the corresponding truth values from table A3. A column *status* is then added. If the *TruthValue* variable from table A3 was NULL, then a 0 is inserted in the column *status*. Otherwise (if the corresponding knowledge was found either true or false), the value of 1 is inserted. The column *status* cannot have any unfilled cells (all must have either 0 or 1). This way, we will get from table A4 a full overview of which pieces of knowledge are useable in the decision process, and which ones are not.

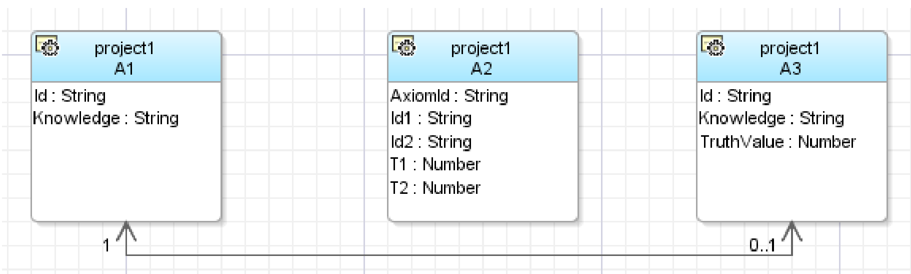


Fig. 1. Structure of tables

5 Hardware

Our long term goal is to create hardware, which

- loads all the data from the tables described above into its memory
- performs the knowledge verification using inference rules (finds the truth values for those knowledge related formulas that do not yet have truth values)
- provides the results to the user.

Next we describe the prototype of the simplified system. We start from the fact that the prototype can only transmit 8 bits at a time. In order to fulfill its "main purpose" (verify the correctness of the knowledge using inference rules), we must "inform" the circuit (using the information in the database) on whether it deals with knowledge from situation descriptions or from an expert-provided axiom; if the truth value of the knowledge is known or unclear; and if it is known, then what is it (true or false?), etc. We represent the corresponding information with sequence of the strings of bits of length eight:

$b_{18}b_{17}b_{16}b_{15}b_{14}b_{13}b_{12}b_{11}$, $b_{28}b_{27}b_{26}b_{25}b_{24}b_{23}b_{22}b_{21}$, … , $b_{n8}b_{n7}b_{n6}b_{n5}b_{n4}b_{n3}b_{n2}b_{n1}$, …

Let us clarify the role and meaning of those bits:

1. Knowledge input is represented by one string of bits (for example, $b_{n8}b_{n7}b_{n6}b_{n5}b_{n4}b_{n3}b_{n2}b_{n1}$).
2. Axiom input is represented by two consecutive strings of bits from our sequence (for example, $b_{n8}b_{n7}b_{n6}b_{n5}b_{n4}b_{n3}b_{n2}b_{n1}$ and $b_{n+18}b_{n+17}b_{n+16}b_{n+15}b_{n+14}b_{n+13}b_{n+12}b_{n+11}$). This is caused by the fact that according to our agreement axioms are only implications from knowledge. Each implication ($P \supset Q$) has two parts: the antecedent (P) and consequent (Q).
3. If $b_{n1}=1$, then it is a case of knowledge input; if $b_{n1}=0$, then it is a case of axiom input. The latter takes place as two consecutive inputs, where the first one is for the antecedent and the second one is for the consequent.
4. Next we look at the value of b_{n1}. If $b_{n1}=1$ (knowledge input), then b_{n2} represents the status of this knowledge: $b_{n2}=1$ means that the knowledge truth value is *known* (0 or 1); $b_{n2}=0$ means that the truth value is not known. However, if $b_{n1}=0$ (axiom input), then $b_{n2}=1$, since we assume that axioms provided by experts are *correct* (therefore it is also not possible to have an unclear truth value for an axiom).
5. Next we look at the values of b_{n1} ja b_{n2}. If $b_{n1}=1$ and $b_{n2}=1$ (the correctness of the inserted knowledge is known, so we have the truth value V (where V=1 or V=0), then $b_{n3}=V$. However, if $b_{n1}=1$ ja $b_{n2}=0$ (correctness of inserted knowledge is not known, so we do not have the truth value V), then $b_{n3}=0$. If $b_{n1}=0$, then (as described previously) $b_{n2}=1$, since it is a case of axiom input. Now (NB!) we must ascertain, whether the first or second component of the axiom is currently being inserted. If it is the first component, then we observe the antecedent P of the corresponding implication $P \supset Q$. If it is the second component, then the consequent Q is observed. The purpose of this observation is to determine if the formula starts with a negation. If it turns out that the formula P is the formula $\neg M$, then $b_{n3}=0$, otherwise (P is not the negation of some other formula) $b_{n3}=1$. In the case of the second input the related string of bits $b_{n+18}b_{n+17}b_{n+16}b_{n+15}b_{n+14}b_{n+13}b_{n+12}b_{n+11}$ is observed as consequent Q. If formula Q is the formula of $\neg N$, then $b_{n+13}=0$, otherwise (Q is not the negation of some other formula) $b_{n+13}=1$.
6. The next four bits ($b_{n7}b_{n6}b_{n5}b_{n4}$) are meant for transmitting the address of the knowledge or axiom in question. For each piece of knowledge we use one address, and for each axiom we use two addresses (for example, 0001 for the antecedent and 0010 for the consequent of the first axiom). The address spaces for both axioms and pieces of knowledge start with 0001. The address increases by one for each new piece of knowledge and by two for each new axiom (using base two numbering).
7. The eighth bit is currently not used.

5.1 Input of Knowledge Related Info (Truth Value and Status)

Depending on the value of the first bit (b_1), the prototype must direct the input either to the knowledge or axiom elements.

First we describe knowledge input. The data is transferred from the database to the circuit bus by *datapath* (see Figure 2). The output Ln1 transmits information about whether the input data represents knowledge or an axiom. The *out* bus may provide additional information after the inference steps, which should be taken into account later (the prototype uses a multiplexer to choose between inputs from buses *Bus1_2* and *out*). Before the inference process the *out* bus does not send any 1 signals (about status or truth value). If the address is not 0001 or if the data does not represent knowledge, then no input can be sent to the block in question (see Figure 2).

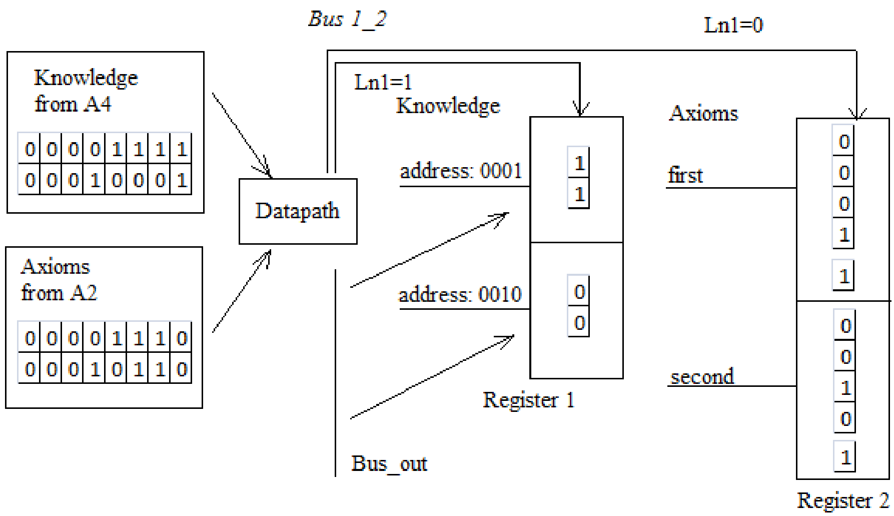


Fig. 2. Data input

5.2 Input of Axiom Related Information

If we wish to insert axiom (implication) related information to addresses 0001 and 0010, then the bus should deliver two bytes in our prototype: 00001110 and 00010110. The implication input block is equipped with a counter. In case of axioms it is important to keep in memory the address of the notation (address) of the antecedent or consequent.

5.3 The Deduction Process in the Digital Circuit

Next we briefly describe the most significant steps of the process that takes place in the circuit.

- The output of knowledge blocks (status and truth value) is sent to the buses.

- If the status is 1 (we know the truth value), then suitable axioms are searched based on the truth value of the knowledge formulas. The axiom A⊃B that fits to formula F is searched as follows: the formula F status (1) indicates that the truth value exists and it is either 0 or 1. If the truth value of formula F is 0, then an axiom A⊃B is considered fitting, if the consequent B is formula F. If the truth value of formula F is 1, then an axiom is A⊃B is considered fitting, if the antecedent A is formula F. Thereat, “just in case”, the truth value (1) of the axiom is checked again (Fig.3).

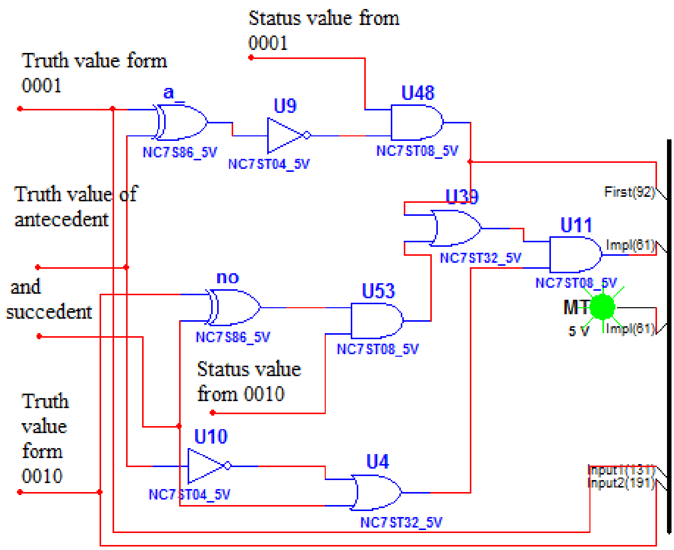


Fig. 3. Axiom search

- The set of inference rules implemented in the prototype is limited to Modus Ponens (MP) and Modus Tollens (MT). Therefore, a choice between the two must be made. If the formula F fits with the axiom A⊃F (truth value of F equals 0, and the truth value of formula ¬F equals 1), then MT is used. If the formula F fits with the axiom F⊃B (truth value of formula F equals 1), then MP is used. The choice of inference rules is made with the help of 1-OF-10 DECODER. In order to activate the part of the circuit that implements the inference rule, a corresponding four bit input signal is generated. The first bit is 1 or 0, depending on whether a fitting axiom was found for formula F. If the first bit is 1, then the next two bits represent the fitting axiom’s antecedent’s and consequent’s status. If the first bit is 0 (no fitting axiom was found) then the next two bits are 0. The fourth bit is always 0.
- The part of the circuit that implements the inference rule MT generates a truth value of 0 for the antecedent of the axiom A⊃F (which corresponds to the truth value of 1 for the formula ¬A). The part of the circuit that implements MP generates the truth value of 1 for the consequent B of F⊃B. Next a signal is generated, which carries information about either formula A (or ¬A) or B: the status of the formula (1, due to the existence of the truth value) and the corresponding generated truth value. This information is sent to the database tables via the *out* bus (see Figure 1).

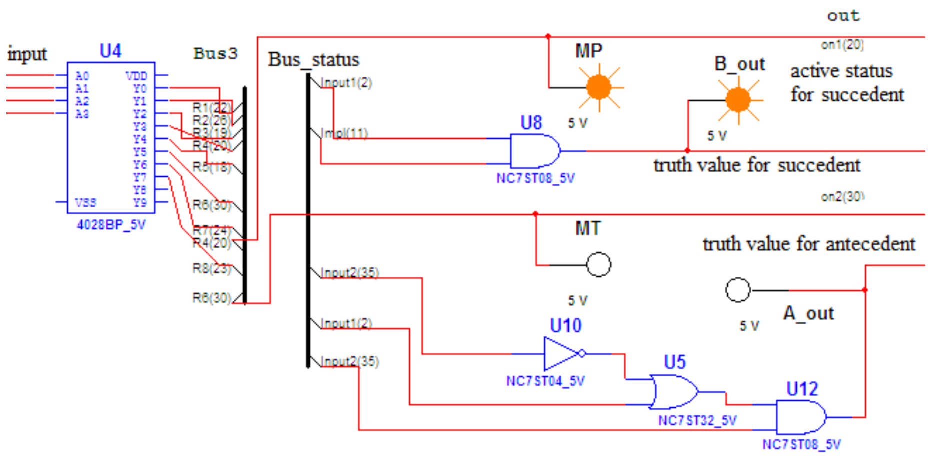


Fig. 4. Inference steps realization

6 Summary

The aim of this work was to study the options of using digital circuits for implementing inference steps that are necessary for the decision process. We described the prototype of one possible digital circuit solution and explained its operating concept. This prototype currently allows the use of only two inference rules (Modus Ponens and Modus Tollens). However, these are some of the most frequently used means for constructing the correct basis for correct decisions. The described prototype has been tested using the simulation software Multisim.

Acknowledgements. The authors would like to thank (NATO) CCD COE scientist Rain Ottis and Tallinn University of Technology Professor Peeter Ellervee for their assistance and support.

References

1. McCarthy, J., Hayes, P.J.: Some philosophical problems from the standpoint of Artificial Intelligence. In: Meltzer, B., Michie, D. (eds.) Machine Intelligence, vol. 4, pp. 463–502. Edinburgh University Press, Edinburgh (1969)
2. Kowalski, R., Sergot, M.: A logic-based calculus of events. New Generation Computing 4, 67–95 (1986)
3. Nguyen, T.D., Ho, T.B., Shimodaira, H.: A visualization tool for interactive learning of large decision trees. In: ICTAI 2000, pp. 28–35 (2000)
4. Renzi, M.F., Vicard, P., Guglielmetti, R., Musella, F.: Probabilistic expert systems for managing information to improve services. The TQM Journal 21(4), 429–442 (2009)
5. Lorents, P.: Knowledge and Logic. In: Proceedings of the International Conference on Artificial Intelligence, IC-AI 2009, July 13-16, 2009, vol. II, pp. 568–570. CSREA Press, Las Vegas (2009)

6. Tarski, A.: Der Aussagenkalkül und die Topologie. Fund. Math. 31, 103–134 (1938)
7. Birkhoff, G., von Neumann, J.: The logic of quantum mechanics. Ann. Math. 37, 823–842 (1936)
8. Kleene, S.C.: Introduction to Metamathematics. Aimed at mathematicians. North Holland, Amsterdam (1952)
9. Goodstein, R.L.: Mathematical Logic. Leicester, England (1957)
10. Lorents, P.: Formalization of Data and Knowledge Based on the Fundamental Notation-denotation Relation. In: Proceedings of International Conference on Artificial Intelligence, vol. III, pp. 1297–1301. CSREA Press, Las Vegas (2001)
11. Lorents, P.: Denotations, Knowledge and Lies. In: Proceedings of the International Conference on Artificial Intelligence, IC-AI 2007, June 14-17, vol. II, pp. 324–329. CSREA Press, Las Vegas (2007)
12. Lorents, P.: Knowledge and Taxonomy of Intellect. In: Proceedings of the International Conference on Artificial Intelligence, IC-AI 2008, July 25-28, vol. II, pp. 484–489. CSREA Press, Las Vegas (2008)
13. Lorents, P.: Knowledge and Information. In: Proceedings of the 2010 International Conference on Artificial Intelligence, pp. 209–215. CSREA Press (2010)
14. Jastroch, N., Marlowe, T.: Knowledge Transfer in Collaborative Knowledge Management: A Semiotic View. In: 4th International Conference on Knowledge Generation, Communication and Management. KGCM, Orlando (2010)
15. Matsak, E., Lorents, P.: Digital solutions for inference rules in decision-supporting systems. In: Fifth International Conference on Digital Information Management. IEEE, Thundar Bay (2010)

Examining the Current State of Group Support Accessibility: A Focus Group Study

John G. Schoeberlein and Yuanqiong (Kathy) Wang

Towson University
7800 York Road
Towson, Maryland 21252-0001 USA
Jschoe4@students.towson.edu and ywangtu@gmail.com

Abstract. Group support applications are widely used in workplace. Unfortunately, persons who are blind often found it difficult to access group support applications, due to the highly graphical nature of the application; this hinders their ability to contribute to the group. As a result, persons who are blind often face problems in gaining and retaining employment. This paper presents preliminary results of a focus group study conducted at the National Federation of Blind (NFB) on accessibility and usability issues of group support applications. How persons who are blind utilize group support applications to support their group tasks, the tasks/steps utilized to complete a group project, the accessibility and usability issues experienced, the reasons for discontinued use of group support applications, the other tools utilized to support group work, the accessibility design considerations, and the accessibility documentation and support needed are discussed.

Keywords: Group support, accessibility, usability, blind.

1 Introduction

According to World Health Organization [19], about 314 million people are visually impaired worldwide, among them 45 million are blind. In today's global economy, collaboration is very important in business [16]. Many organizations utilize group support applications that support communication, coordination and collaboration (such as email, document sharing, scheduling software, conferencing systems, etc) for their collaborative work. Considering the unemployment rate for working age persons who are blind in the united states of 70% to 75% [11], the accessibility problems associated with group support applications become crucial. In order to improve the design of group support applications, the following questions need to be answered: How do persons who are blind work in groups? Are there any differences when compared to a group containing all sighted members? What are the accessibility and usability challenges persons who are blind experience with software applications (specifically groupware or group support application)? What kind of support do persons who are blind require so that they can perform group work?

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 272–281, 2011.

In an initial effort to answer the above research questions, a focus group study was conducted with the help from the National Federation of Blind (NFB). This paper reports the activities of the focus group study conducted and its preliminary result.

2 Related Backgrounds

2.1 Accessibility and Usability Issues for Persons Who Are Blind

Accessibility describes the degree to which a product, device, service, or environment is accessible by as many people as possible. It is often used to focus on people with disabilities and their right of access to entities, often through the use of assistive technology. In order to understand the output from computer systems, persons who are blind are forced to use either tactile displays or sound. A screen reader (e.g. JAWS, Windows-Eyes) is a piece of software that runs in the background to read the screen memory and send any text it finds there to a speech synthesizer [12]. However, the screen reader cannot interpret graphical information on the screen, if the program is not properly programmed.

On the other hand, usability refers to the extent a product (e.g. device or service) can be utilized by specific users to achieve specific goals with effectiveness, efficiency and satisfaction. When persons who are blind are aware of an application's accessibility and usability issues, they are more likely to avoid using that type of application.

2.2 Groupware and Group Support Applications

Groupware applications are the technical applications to assist persons to communicate, coordinate, collaborate and compete [13]. Groupware can be considered a comprehensive tool that includes email, group calendars, as well as tools such as: wikis, blogs, social networking, etc. As the technology evolves, more software applications start to incorporate key features of groupware applications. For instance, a document editor can provide features such as: shared track changes that allow collaborators to co-write papers; or, an email application, which allows the sharing of calendars from multiple users. Even though the above examples are not typical groupware applications, they do incorporate key features of groupware applications that support group work. The researchers refer to these software applications as group support applications.

3 Research Method and Procedure

Focus groups are group interviews. Generally, a focus group is utilized for marketing research purposes while it has been used in social and design research since the 1930's [2]. Focus groups are very useful for the generation of ideas and for discovering problems, challenges, frustrations, likes, and dislikes, among participants especially when the researchers do not have enough information to design a survey. Moreover, focus groups allow for the access of multiple points of view in a short time period (e.g. a single meeting) [2]. The size of the group can vary from 4 to 6 as a mini-group to more than 10 as a full group [5]. Considering the explorative nature of this particular study, the researchers believe a focus group study was an appropriate approach to gathering data regarding group support applications.

As Goebert and Rosenthal [4] noted, "*the primary consideration is who will provide the most insightful information*," related to the topic when conducting focus groups. The participants of this project must meet the following requirements: Persons who are blind, with no residue vision; Persons who work in group settings; Persons who are familiar with screen readers; Persons who utilize computers for their work; and, Persons who have tried to use group support applications and other tools to support his/her collaboration with others. Considering the limited time that each participant can contribute, a mini-group of five participants, from the Access Technology and Education Teams at the NFB, were utilized for this study in order to get more in-depth information [5].

The focus group session was conducted at the National Federation of the Blind (NFB) [10], in Baltimore, Maryland. The focus group session, contained the following steps:

1. Recruit the participants;
2. Prepare demographic, prior-experience questionnaire and meeting guideline;
3. Conduct the focus group session, based on the meeting guideline;
4. Perform content analysis of the audio transcript and the researcher notes, to identify content categories;
5. Validate agreement of the coded content, using Cohen's Kappa [1]; and,
6. Report findings.

Prior to the focus group study, each participant filled out a short questionnaire, regarding demographic information and prior-experiences with group support applications and screen readers. A meeting guideline was designed, containing a list of possible questions for discussion.

In order to be easily accessible for the participants, the focus group session was conducted in a private conference room at the NFB. One researcher served as the moderator, while the other served as a note-taker during the session. Additionally, the focus group session was audio recorded. The focus group started with a brief introduction from the moderator on the purpose of the meeting, a round of brief self-introductions, and followed by a discussion on several topics prepared before the meeting. Follow-up, probing questions were asked to ensure the understanding of the discussion.

Following the focus group study, the researchers' notes and the audio recording of the group discussion were transcribed with the permission from the participants. The audio transcript and detailed notes from the group discussion were coded based on the themes that emerge from the participants' comments. The focus group session's duration was one hour and thirty minutes.

4 Preliminary Results

4.1 Demographics

The participants of the focus group study included five members of the Access Technology and Education Team at the NFB. The five participants consisted of three females and two males. All members are between 20 and 39 years old. All of the participants utilized the computer daily, for business use, and have utilized the computer between 11 and 20 years. All participants utilize the JAWS Screen Reader [3] with two participants having 6 to 10 years experience and two participants having 11 to 20 years experience. The fifth participant did not respond to the number of years experience with the JAWS Screen Reader.

None of the participants reported that they ever used any traditional full fledged groupware applications, so the researchers focused on group support applications that include some groupware features such as: shared calendar, email, track changes, and chat. The participants were asked to identify all the group support applications they have used for group work. As the result, Microsoft Outlook (application version 2008) [9] was identified as the primary group support application utilized most frequently in completing group projects at the NFB. One participant had 1 to 5 years experience, three participants had 6 to 10 years experience and one participant had 11 to 20 years experience using Microsoft Outlook. All participants used Microsoft Outlook daily. When interacting with Microsoft Outlook's interface, the participants utilized the JAWS Screen Reader exclusively.

4.2 Data Analysis of the Focus Group Study Recording

Content analysis was performed on the audio transcript and researchers' notes from the group discussion collected from the focus group study. Two research assistants performed as coders for this analysis. Key points were grouped into categories. Inter-coder reliability [7] was calculated to make sure of the consistency of the coding. Statistical analysis on the agreement between the coders was calculated using SSPS [14], using Cohen's Kappa [1]. The Kappa measurement of agreement sought was 70% or greater, to provide evidence of acceptable agreement and reliability between the coders. For the 86 cases generated from this focus group study, a Cohen's Kappa of 0.757 was achieved, providing evidence of acceptable agreement and reliability between the coders. Seven categories were determined as a result of this analysis:

- Participants identification of group support software utilized;
- The tasks/steps necessary to complete a group project;
- Accessibility and usability issues experienced;
- Group support applications utilized, but discontinued;
- Other tools utilized to support group work;
- Accessibility design considerations; and,
- Suggestions on accessibility documentation and support.

Participants Identification of Group Support Software Utilized. During the focus group study at the NFB, it was discovered that the participants did not utilize any traditionally identified groupware (e.g. Lotus Notes); therefore, our discussion was focused on group support applications that may contain some groupware features. Unsurprisingly, the primary group support application reported was Microsoft Outlook for various activities including: email, journal, calendar, contacts, and tasks.

The Tasks/Steps Necessary to Complete a Group Project. The participants reported that they always start a group project with a face-to-face "brainstorming session" in a conference room. Conferencing calls are normally made in the situations involving group members in another location. During the face-to-face meeting or conference call, a member of the group is normally responsible for note taking. Once the start-off meeting ends, the group's manager sends emails with documents containing the ideas discussed. A project schedule and task list is developed by the manager and sent through Microsoft Outlook, as either an email or calendar items. The group then exchanges emails to discuss further on the project. Weekly or monthly face-to-face meetings or conference calls are conducted throughout the project duration.

Accessibility and Usability Issues Experienced. Based on this initial research, it appears that Microsoft Outlook used in conjunction with the JAWS Screen Reader has both accessibility issues and usability issues.

Most of the participants agreed that the web-based version of Microsoft Outlook is inaccessible and instead the group utilizes the application version of Microsoft Outlook. Therefore, all the comments are focused on the application version (Windows-based) of Microsoft Outlook. The accessibility issues identified for Microsoft Outlook include: inconsistency issues between Microsoft Outlook and the JAWS Screen Reader; reading the wrong message by the JAWS Screen Reader; and, accessibility issues with the Journal, calendar, messages, login screen, and contacts.

The first challenge the participants face is the inconsistency of the accessibility of the various features of Microsoft Outlook, as well as the interactions between the screen reader and the application. For example, the participants commented, "*Some days it (screen reader) works, some days it does not work.* W*hether it works, depends on the individual machine.*" After further discussion, it was realized that the problem may be attributed to the version of the operating system, the version of Microsoft Outlook, or the version of the JAWS screen reader installed on each individual machine.

Another accessibility issue identified by the group is related to the software upgrade and version of the screen reader. For example, participants commented on the problem they experience when trying to switch views in "calendar" feature, "*It worked with JAWS 10. It is not working with JAWS 11*". It appears that upgrading the JAWS Screen Reader to a newer version may cause inconsistency issues with group support applications. In this case, utilizing the daily-view of the calendar was a work-around for the inaccessibility of the monthly-view of the calendar. In some inaccessibility cases, often the solution may be to approach the interface from another direction, such as, utilizing a different view (daily-view versus month-view) to access information.

Since the first step to utilizing Microsoft Outlook is to log into the system, it is necessary that the login screen be accessible. If the login screen is not accessible, the whole system will not be accessible. A participant commented, "*Occasionally, the Login Screen (Microsoft Outlook) will not read correctly (using the JAWS Screen Reader). I have to set Outlook to work offline and then login again.*" The participants proposed a work-around solution to the accessibility issue of the Login Screen, which is to uncheck the work-offline check box within Microsoft Outlook's interface.

The participants ranked the "calendar" feature of Microsoft Outlook as being the most inaccessible feature of Microsoft Outlook. The participants expressed that they did not receive notifications of important events, such as meetings or appointments. As a result of not receiving any notification of a meeting or appointment, they may miss an important deadline. The JAWS Screen Reader may need to be enhanced to respond to events, such as the notifications of meetings or appointments.

The participants also reported that the multiple panels utilized in Microsoft Outlook presents a big challenge for screen readers. The participants had to move from one panel to another panel, to complete certain tasks. One method to work-around the movement between multiple panels was to display on the interface one panel and to hot key directly to other panels. The adjustment of the interface to present one panel of the interface simplified the complexity of the interface and reduced the amount of information presented to the screen reader. The less complex the interface, the better the interaction between the screen reader and that interface.

The Participants identified usability problems of the applications they use during the discussion. The participants ranked the "calendar" feature of Microsoft Outlook as being the most unusable feature of Microsoft Outlook. The participants complained that they were not able to, "*share the calendar.*" Additionally, they commented that appointments were not tracked correctly.

When discussing the "email" feature of Microsoft Outlook, the participants noted that they may not receive a message, or the message may not read correctly. A Participant commented that the messages are not up to date, "*sometimes I have to refresh. The new message is on the screen, but it is still reading the old message.*" The screen reader must properly and consistently read messages in the order they are delivered in order for any group, dependent on a screen reader, to respond and function as a team.

The "contact" feature of Microsoft Outlook presents another challenge for the group participants. They reported that, "*it keeps reading the same contact* (even though I know I have kept more contacts in the system)." The minimum requirement of the screen reader should be to identify an entire list and to be able to parse that entire list in any order.

Additionally, some features such as the reading pane had to be removed from view in order for the JAWS Screen Reader to work in a manner that the participants could understand. Basically, the interface worked better with the JAWS Screen Reader, when the interface was less cluttered.

There are a number of accessibility and usability issues reported by the Access Technology and Education Team, but for the most part they appear to be related to how the screen access software interacts with the graphical interface. Some of the accessibility and usability issues seem to pertain to synchronization issues between the screen reader and the content.

Group Support Applications Utilized, but Discontinued. The focus group identified group support applications attempted for group work and discontinued. The goal was to identify the reasons why applications were discontinued and, if there was a common set of reasons. The Access Technology and Education Team have tried a number of group support applications, such as: WIKI, Survey Monkey, and Microsoft Access. However, they have identified several reasons for their discontinued use including: previous unsuccessful experience; lack of interest; and, additional learning curve.

It appears that the focus group participants are apprehensive to use technologies that they have not tried, due to their previous unsuccessful experience with other software. For example, a participant commented that they are, *"Easily intimidated by the other tools or they are not accessible. When you branch out, you may not get the best accessible tools."* Unlike the widely used discussion forums, such as: group project discussion forums or class discussion forums, utilized among persons without visual impairment, the participants expressed a lack of interest in such applications, "*it does not lend itself to the work environment.*" The extra learning curve was another reason noted for not attempting new applications, *"When you learn to use the computer, you have to learn three things: the operating system, the application and the screen reader software. Sometimes it is just not worth it (to learn new software). Then, there is the extra learning curve."* Anytime persons who are blind attempt to utilize a new group support application, they have to orient the screen reader to the interface of the application and learn the nuances of the application.

Other Tools Utilized to Support Group Work. In addition to Microsoft Outlook, the following tools are utilized occasionally by the Access Technology and Education Team: America Online (AOL) Instant Messaging (IM), for quick messaging; Microsoft Word, for document sharing; WIKI, for document storage and sharing, Twitter and Facebook, for social networking; and, Mobile Phones for, text messaging. As a Team, the Access Technology and Education Team utilize a small number of applications, when working in their group. The key to the limited use of other applications may be due to the accessibility and usability issues experienced, and the difficulty of learning new applications.

Accessibility Design Considerations. During the focus group study session the participants emphasized that the screen access software needs to be consistent when interacting with the application's interface. Considering some of software manufactures developed interfaces that provided voice-over to their application to provide accessibility, a question regarding the participants' preference on voice-over interfaces versus using a screen reader was asked. The voice-over interface application choice was determined to be an unacceptable solution by the focus group participants. The participants commented, "*understanding how to use the self-voicing application in conjunction with the screen access software for other applications would be a hassle*." Even though the screen reader is not perfect, the use of one screen reader for all applications was determined to be easier and the preferred approach for accessibility and usability.

Suggestions on Accessibility Documentation and Support. The support for accessibility features is extremely limited compared to the support provided for other features. A participant noted, "*Normally, to learn about accessibility features, you have to dig down pretty far to know whether an application supports accessibility. When other features of the product tend to be more available.*" And, in most cases, there are few documents on accessibility features that, "*They (accessibility features) are generally a small statement in the back of the documentation.*" These problems may be caused by the reluctance of the design team, "*They really don't think about the user group they are serving. They do accessibility as a matter of course – as a requirement.*" Therefore, "*there is a responsibility of the developers that they provide the same support for accessibility features.*"

The participants also suggested that the support staff of group support applications need to be trained on accessibility issues as well, "*If I go home today and call Apple and say I need help with voiceover, 9-times-out-of-10, they (support staff) will not know what I am talking about. Are accessibility features not important enough to train their support staff on?*"

5 Discussion

Since the focus group participants are blind, this research provides a unique perspective on the group process conducted entirely with blind participants. The participants agreed that the normal group projects follow this general pattern: face-to-face/teleconference group discussion in a conference room setting; and, follow-up discussions and document exchanges utilizing email (MS Outlook). The most frequently utilized features include: the *calendar* for appointment tracking; *email* for information exchange, the *journal* for notes; the *contacts* for contact information; and, the *tasks* for tracking the progress of a project. Although the group process may appear similar to be a standard group process, with meetings and exchanges of email, the difference is that persons who are blind require assistive technologies (e.g. screen readers and note takers) for exchanging ideas.

While conducting the group process, persons who are blind will utilize adaptive technologies such as: screen readers for reading documents and email messages; and, note taking devices to record minutes during meetings. Similarly, Woodfine's [18] research identified the need for special support and assistive tools for persons with dyslexia to interact in synchronous e-learning environments. How can we better integrate these adaptive technologies into the group process for the support of group work? How can researchers and developers create a complete group support solution with accessible email, chat and document exchange?

As more and more applications have made their way to become web services, researchers have been focusing on the accessibility of web interfaces. For instance, Harrison [6] had identified approaches to making web pages accessible but did not report approaches to improving the accessibility of windows-based applications. Unfortunately, windows applications are still utilized in many businesses as tools for group work. This research made the attempt to identify accessibility and usability issues in regard to windows-based applications, specifically Microsoft Outlook. Accessibility and usability issues occurred while reading text messages, or reading the wrong message, accessing items in the calendar, logging in, and accessing the journal and contacts. Since, windows-based applications do not support HTML, Web-based solutions to accessibility do not work for windows-based applications. Many of the features employed in Microsoft Outlook such as: calendars, emails, lists in journals, and contact lists, are common elements among many group support applications. How can researchers and developers improve the accessibility of windows-based applications, which support email, calendars, and lists?

Steep learning curve and the inability to solve the accessibility issues encountered are identified as main reasons why persons who are blind discontinued or would not attempt to utilize new group support applications. This is consistent with the design guidelines generated from previous research [13]. Even though all of the participants of the focus group study were capable users of the operating system, the selected group support applications and screen access software, they still expressed the apprehension of utilizing new group support applications. Given the participants' skill levels and their difficulties with inaccessible group support applications, it can be predicted that less skilled persons who are blind or working in groups with sighted persons may experience more severe challenges when interacting with group support applications.

Takagi's [15] Notes Reader and Watanabe's [17] Blog Reader were two examples of redesigning the web-page interface and including voice-over as feedback. The design suggestion of an accessible self-voicing, or voice-over, group support application was presented as an alternative to utilizing a screen reader. This suggestion was unanimously found to be less acceptable than the use of screen reader software, to access group support applications. Even though Takagi's [15] and Watanabe's [17] voice-over solutions worked in there case, it was determined by this research to be a less-acceptable solution to the use of a screen reader as the interface to group support applications.

The focus group study participants expressed the need for consistency, when discussing design considerations. Similarly, Luk's [8] research with spatial clustering was devoted to the need to organize content and make the interface consistent. For the most part, the focus group participants agreed that the group support application's (Microsoft Outlook) interface was good, but the interaction between the screen reader and the application was the primary cause of their accessibility issues. As different versions of the screen reader software are developed, older versions of the group support applications may no longer be accessible. How can researchers and developers deliver consistent and integrated group support applications that function properly with screen reader software?

Unique to this research is the presentation of the concept of determining the appropriate level of accessibility support and documentation needed by persons who are blind. Often, the participants of the focus group want to use new features of group support applications or even new group support applications, but the support and the documentation on accessibility is limited or non-existent. This lack of support for accessibility causes frustration on the part of persons with disabilities and their discontinued use of the group support application. How can researchers and developers determine the necessary level of support that should be devoted to accessibility and usability?

6 Conclusion

The focus group study participants identified accessibility and usability issues while interacting with group support applications. The limitations of this preliminary focus group study are due to the small number of participants, and the common backgrounds of the participants. The results cannot be generalized. Since the original writing of this research, several field studies and additional focus group studies were conducted.

References

1. Cohen's Kappa,
   http://www.childrens-mercy.org/stats/definitions/kappa.htm
2. Courage, C., Baxter, K.: Understanding Your Users. A Practical guide to user requirements. Methods, Tools, and Techniques. Morgan Kaufmann Publishers, San Francisco (2005)
3. Freedom Scientific JAWS,
   http://www.freedomscientific.com/fs_products/software_jaws.asp
4. Goebert, B., Rosenthal, G.: Beyond Listening: Learning the Secret Language of Focus Groups. John Wiley, New York (2002)
5. Greenbaum, T.L.: The handbook for focus group research, 2nd edn. Sage Publication, Thousand Oaks (1998)
6. Harrison, S.M.: Opening the Eyes of Those Who Can See to the World of Those Who Can't: A Case Study. In: Proc. SIGCSE 2005, St. Louis, Missouri, USA, pp. 22–26 (2005)
7. Kurasaki, K.S.: Intercoder Reliability for Validating Conclusions Drawn from Open-Ended Interview Data. Field Methods 12(3), 179–184 (2000)
8. Luk, R., Yeung, D., Lu, O., Leung, E., Li, S.Y., Leung, F.: Digital Library Access for Chinese Visually Impaired. In: ACM 2000, pp. 244–245 (2000)
9. Microsoft Outlook (2008),
   http://office.microsoft.com/en-us/outlook/default.aspx
10. National Federation of the Blind,
    http://www.nfb.org/nfb/default.asp?SnID=627174770
11. National Federation of the Blind. Assuring opportunities: A 21st century strategy to increase employment of blind Americans,
    http://www.icbv.net/National%20Issues/Opportunities.htm
12. Pitt, I.J., Edwards, A.D.N.: Improving the usability of speech-based interfaces for blind users. In: Proc. of ASSETS 1996, Vancouver, British Columbia, Canada, pp. 124–130 (1996)
13. Shneiderman, B., Plaisant, C.: Designing the User Interface: Strategies for an Effective Human-Computer Interaction, 4th edn., pp. 411–450. Addison Wesley, Boston (2005)
14. SSPS, Inc., http://www.ssps.com
15. Takagi, H., Asakawa, C., Itoh, T.: Non-Visual Groupware Client: Notes Reader. In: Proc. Center on Disability Technology and Persons with Disabilities Conference. California State University, California (2000)
16. University of Washington Department of Ophthalmology. Statistics on Blindness and Blinding Diseases in the United States,
    http://depts.washington.edu/ophthweb/statistics.html
17. Watanabe, M., Okano, A., Asano, Y., Ogawa, K.: VoiceBlog: Universally Designed Voice Browser. International Journal of Human-Computer Interaction 23(1-2), 95–113 (2007)
18. Woodfine, B.P., Nunes, M.B., Wright, D.J.: Text-based synchronous e-learning and dyslexia: Not necessarily the perfect match! Compute. Educ. 50(3), 703–717 (2008)
19. World Health Organization,
    http://www.who.int/mediacentre/factsheets/fs282/en/

BioSEME: A Semantically-Integrated Content Management System for Cancer Research

Ionut Mihai Stancu

Indiana University - School of Informatics, Indianapolis, Indiana, USA
imstancu@umail.iu.edu

Abstract. In this project, we propose a paradigm shift from the use of relational databases and traditional web publishing tools to a Semantically-Integrated Research Content Management System augmented with research collaboration functionality for curation of cancer research data. By using a prototype called BioSEME, we will aggregate cancer biomarker-related PubMed literature data using term-specific data feeds. The data will then be annotated by designated users using the standardized taxonomy and vocabulary derived from known ontologies and cancer-related literature, which is subsequently associated and published through the semantically-enhanced user interface. We believe that in the future such online collaborative-research community and semantically-integrated virtual scientific knowledge environment will help support the cancer research community.

Keywords: BioSEME, Semantically-Integrated Content Management System, Online Collaborative-Research Community, Virtual Scientific Research and Knowledge Environment.

1 Introduction

1.1 Background and Motivation

The *Centers for Disease Control* data shows that one in every four deaths in the United States results from cancer [1]. Multiple deaths might be prevented by early cancer detection and early treatment. Biomarkers are biochemical characteristics of a living organism that can enable early diagnosis of a disease, its targeted molecular therapy, and the monitoring of therapeutic responses. *National Cancer Institute (NCI)* dictionary states that biomarkers can be *"a biological molecule found in blood, other body fluids, or tissues that is a sign of a normal or abnormal process, or of a condition or disease"*[2].

Genomics, proteomics and metabolomics sciences are known to have generated large amounts of heterogeneous candidate cancer biomarkers data. Unfortunately, very few candidate biomarkers have become prognosis and diagnosis tools despite their enormous potential benefits, and even fewer biomarker candidates have been validated and *Food and Drug Administration (FDA)* approved. There are several factors affecting the slow pace of biomarkers discovery and validation. One impediment to protein based biomarker discovery is the ability of available

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 282–291, 2011.

proteomics technology to select low levels of protein biomarkers over the noisy background of general plasma proteins [1]. This shortcoming can result in many useful biomarkers being overlooked during research studies. Nevertheless, many candidate biomarkers have already been discovered; however, they have not been validated through rigorous analysis in large studies with populations from many cancer related diseases. We believe progress in biomarker research can be accelerated as the result of more efficient knowledge-based tools to integrate, mine, and analyze large amounts of heterogeneous biomedical data.

Within the domain of cancer biomarker research, collaboration between researchers is critical to validating biomarkers as early cancer indicators. By sharing study specific information, researchers are able to analyze a biomarker in multiple contexts and identify similarities in the epidemiological characteristics of varying cancer population sets. Collaboration between institutions provides an opportunity to access a larger volume of data than would otherwise be available, but can introduce complexity in to the task of presenting a unified, coherent picture of the research that has been conducted. [3]

By definition the curation process allows for enhancements to the raw research data by means of capturing, annotating, and cataloging of relationships among data elements. This ultimately results in significant improvements to the quality of the research data generated. Also a very important aspect of the curation process is that it requires collaborative agreement based on standards for research results, and also a synchronized effort at maintaining those standards over time. In the absence of the curation process, distributed research data for cancer biomarkers remains isolated in separate repositories making it difficult and sometimes impossible to cross-reference with data from other repositories. This ultimately results in a failure to achieve its full research potential.[3]

1.2 Statement of the Problem

Despite increasing research funding in the biopharmaceutical industry, the biomarker discovery and validation rate has been declining, in part due to both the lack of efficient data integration and curation tools and the lack of domain specific research collaboration tools. Relational databases and limited collaborative applications have been developed to support the discovery and management process in many biomedical applications [4]. The problem is that the integration of heterogeneous data distributed in isolated sources across the Web still poses many challenges, because of the current differences in file formats, domain specific terminology, and communication protocols among distributed information sources. What is also important for biomarker research is the technologies to assist researchers recognize any extraneous or incorrect information sources while drawing attention to key relations hidden in a set of research data. [5]

Advanced Content Management Systems can facilitate data curation and collaboration among scientists, by semantically integrating heterogeneous data resources through the capture of the semantics in the relationship among the information research resources [6]. *The Semantic Web (Web 3.0)* research community has made some progress in coming up with solutions to the semantic integration of heterogeneous data. Also certain efforts are underway to apply Semantic Web technologies on domain specific datasets in biomedical informatics [5]. These efforts are considered important steps for biomedical informatics and cancer biomarker research.

1.3 Purpose of the Project

Our purpose in this project is to develop BioSEME, a prototype of a biomedical research content management system for the discovery and management of cancer biomarkers, using an improved and semantically-integrated network of cancer biomarker literature data and resources from *PubMed*.

Biomarker research depends on heterogeneous information sources that are sometimes questionable in terms of reliability. With the advancements in technology these data sources are becoming increasingly multimedia and large. For this reason in this project, we propose a paradigm shift from the use of relational databases and traditional Web publishing tools to a semantically-integrated research Content Management System augmented with research collaboration functionality for curation of cancer biomarker data to show the potential benefits for improving the discovery, management, and eventual validation of different candidate cancer biomarkers.

For this purpose we will create a case study based on the Plasma Proteome Institute process of identifying cancer biomarker supporting literature data that will provide us with the requirements for the prototype. Also we will use this literature data to assess some feasibility and usability aspects of the BioSEME tool for cancer biomarker data curation. By using the BioSEME prototype, we will aggregate cancer biomarker-related *PubMed* literature data using term-specific data feeds. The data will then be annotated by designated users using the standardized taxonomy and vocabulary derived from known ontologies and cancer-related literature, which is subsequently associated and published through the semantically-enhanced user interface.

We believe that such an online collaborative-research community and semantically-integrated virtual scientific knowledge environment will promote the acceleration of the validation and also sharing of hypotheses and studies supporting cancer biomarker research. Also using the prototype we will be able in the future to help answer the question on how can we semantically structure and integrate cancer biomarker information to improve the decision support process involved in the cancer biomarker validation. As the cancer biomarker research field is such a vast and broad topic, in scope for this project was the development of the BioSEME prototype only.

2 Methodology

2.1 Introduction

The methodology for this project will involve the development of a cancer biomarker literature data curation and annotation case study prototype, called BioSEME, using a semantically integrated research Content Management System augmented with research collaboration functionality. Because the cancer biomarker research field is such a vast and broad topic, in scope for this project was the development of the BioSEME case study prototype only.

This prototype has the potential to evolve in a collaborative research-community facilitating an integrated virtual scientific knowledge environment to support cancer biomarker research. The resources required to complete validation of cancer biomarkers are beyond the capabilities of individual laboratories, investigators, institutions and the scope of this project. However, we believe that by using BioSEME to create an online collaborative research community that can be made available to cancer biomarker researchers and clinical scientists, we can offer insight into how novel web resources can be used to potentially validate promising biomarkers. Using the *Drupal* CMS Taxonomy module, we will explicitly represent the meaning of terms through content types and vocabularies, while mapping the relationships between those terms. The research environment will allow scientists to collaborate on and curate cancer biomarker-related biomedical literature data from *PubMed* using standardized vocabulary derived from cancer biomarker literature. *PubMed* is the largest public biomedical electronic library and archive with over 19 million records of articles in the biomedical literature. The representation of scientific terms and their interrelationships will become the BioSEME cancer biomarker ontology.

Using the *Drupal* Content Construction Kit module, we will create different content types that will be mapped to the BioSEME taxonomy. The multi-tagging capability will allow the content to be modeled in advance to meet the needs and demands of different sets of biomarker data points. BioSEME will store biomarker annotation and collaboration metadata in a semantic format described in the same web format. Data that links to the same concept may be discovered, thus improving the retrieval capabilities of relevant information required for discovery and validation of biomarkers [23].

Research statements in the system will inherit the semantic associations between the cancer biomarker data and the content types relating them in the BioSEME taxonomy. By searching the data space for objects multi-tagged based on the domain specific cancer biomarker taxonomy, related relevant concepts can be uncovered. Also the retrieving data through the semantically linked tags, the researcher can perform very powerful selections [24].

The environment will include semantic-faceted browsing and search capabilities using the Exhibit tool, an open-source Java-based tool developed at the *Massachusetts Institute of Technology (MIT).* The user interface will make use of the integrated data feeds available in RDF/JSON, which will enhance the retrieval and discovery process. By using *Web 3.0* standards and domain specific ontologies online systems can facilitate the creation of distributed infrastructures dedicated to research data sharing and exchange that will ultimately allow for the creation of a an cyberinfrastructure for cancer biomarkers research [25].

2.2 Requirements Analysis

This section presents the prototype requirements of the biomarker repository using a use case scenario. Use cases are the preferred method of defining requirements in the Unified Software Development Process. They are user-centric focusing on the steps that an actual user may perform and making the prototype easier to test, but also it provides metrics for completion and success.

Our use case is designed with a biomarker research curator's workflow in mind derived from the Plasma Proteomes Institute's process in identifying and ranking of the 1261 candidate cancer biomarkers. Also, we will use *PubMed* literature data to assess some feasibility and usability aspects of the BioSEME tool for cancer biomarker data curation.

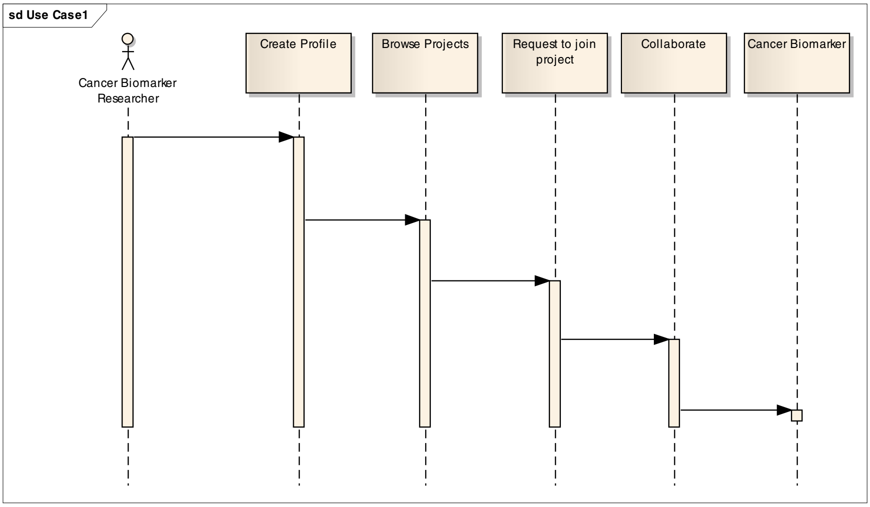


Fig. 1. General Interaction Sequence Diagram

The tool requires a comfortable interface to enter, review and curate data from *PubMed* literature original sources, providing a way to track the progress and completeness of the research for a particular cancer biomarker.

The overall system boundary describes the three main requirements for the BioSEME prototype. The users will have to be provided a way to register in the system. Researchers will be required to build a custom profile that will allow them and the system to track contributions and also the traceability in the identification of research work and content.

The system shall also provide extensive domain specific content contribution and curation functionality. This will allow the cancer biomarker researchers to curate and annotate cancer biomarker literature data. The system will provide semantic-integration of the research content and data using a taxonomy engine. Collaboration features like forums and personal blogs linked to research topics and personal profiles will also facilitate and improve the curation process of candidate cancer biomarkers.

2.3 Architecture Design

For the purpose of creating a prototype to integrate the literature related to cancer biomarkers, we chose to base BioSEME on a *Drupal* CMS. *Drupal* is a very unique and taxonomy driven Content Management System that has a modular architecture that also allows for the creation of semantically- integrated content and collaborative functionality for cancer biomarker literature data curation. Also important is the fact that users with appropriate permissions can extend BioSEME's functionality allowing them to manipulate the data structure on the fly. By choosing *Drupal* CMS, allows us to leverage a large active open-source community of many developers and custom modules for any future developments in the system. For the prototype creation and experimentation, I have selected the latest *Drupal* 6 version, which includes more flexible theming, performance enhancements, and some light support for *Semantic Web Technologies* like RDF.

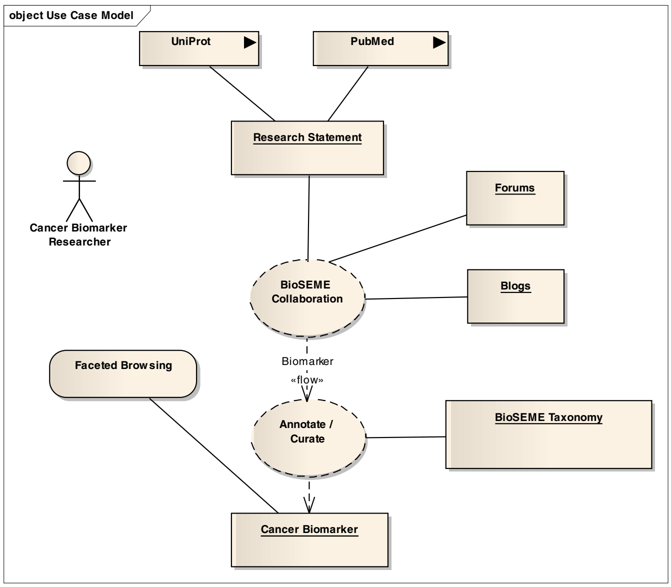


Fig. 2. BioSEME Use Case Diagram

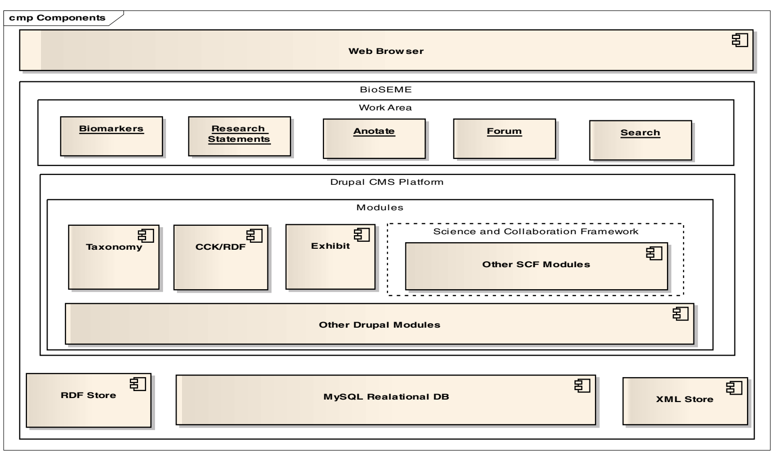


Fig. 3. Component Architecture Diagram

In developing the BioSEME prototype we used the protein cancer biomarker literature data obtained from the Plasma Proteome Institute to drive the case study and the requirements and the architecture of the system. Based on our case study, in Figure 3.3.1 we show how BioSEME will aggregate cancer biomarker related *PubMed* literature enhanced with additional *UniProt* metadata. By using the prototype, we will aggregate cancer biomarker-related *PubMed* literature data using term-specific data feeds. The data will then be annotated by designated contributors using the standardized taxonomy and vocabulary derived from known ontologies and cancer-related literature, which is subsequently associated and published through the semantically-enhanced user interface.

The integration with *PubMed* data will be done using term-specific RSS data feeds. This curated data will then be made public through the faceted browsing user interface as semantically-integrated cancer biomarker research content. The system will be accessible through any Web browser and will have provided a working area where users can search, annotate resources and collaborate.

The *Drupal* platform will be composed of the Science and Collaboration Framework, the powerful Taxonomy module and the Content Construction Kit with the enhanced Resource Description Framework functionality. Also for the user interface and faceted browsing we will be using the Exhibit Module. Other *Drupal* Core modules will be used on demand based on the requirements for the BioSEME prototype. For data store we will be using the *Drupal* system RDF Store, a *MySQL* relational database and for future enhancements we can include and XML data store.

The data architecture revolves around research statements that are extracted by the researcher from *PubMed* cancer biomarker relevant biomedical literature data. Based on our case study, candidate cancer biomarkers can be linked to protein content types

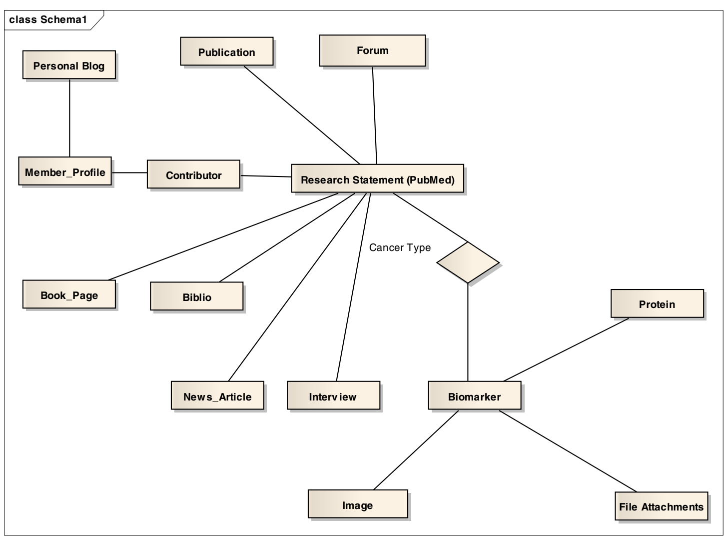


Fig. 4. Content Types Diagram

and can have images and file attachments integrated in the system. Based on different cancer types, biomarkers will be semantically related to different research statements. Also for advanced collaborations, the data will be integrated with its contributor's and collaborators member profiles and also connected to different forum topics and relevant publications. The data will also be connected to public news, interviews, book pages and referenced using the Biblio content type.

3 Results and Discussion

3.1 Results

In this project, we successfully developed BioSEME, a prototype of a semantically integrated research content management system augmented with research collaboration functionality to allow for cancer biomarker PubMed literature data to be curated and collaborated on. We believe that the results show the paradigm shift from the use of relation databases and traditional web publishing content management systems to a semantically integrated research content management system augmented with research collaboration functionality. In scope for this project was the prototype development and loading of a sample set of biomarker data obtained for the *Plasma Proteome Institute*. The prototype creation was based on a case study derived from the development of the candidate cancer biomarker list obtained from *PPI*.

To develop the prototype we utilized the *Drupal* CMS Platform allowing us to build a collaborative environment enhanced by advanced user interfaces and faceted-browsing. In the future, the prototype will allow us to further test and validate how this kind of tool provides support for cancer biomarker research.

3.2 Discussion

The BioSEME prototype is now operational and so far we have setup only one test user with full contribute and system administration privileges. However the system is designed to allow multiple users to register and upon approval, to start using the tools provided for cancer biomarker research.

In order to assess the BioSEME prototype we evaluated a few metrics. The first is an improvement in data visibility and also navigation using advanced and semantically enhanced faceted browsing technology. Based on a literature case study and because it was designed with a biomarker research curator's natural workflow in mind, the prototype offers a comfortable interface to access and review data from *PubMed* literature, providing a way to track the progress and completeness of the research for a particular cancer biomarker. This is a significant improvement over the Microsoft Excel spreadsheet that *PPI* put together at the end of their study. We believe that by using our system the data list can live in perpetuity updated and also allow for additional curation and collaboration to support the cancer biomarker findings and research.

The interface design, along with the integration with *PubMed* external data sources creates a good user experience where users do not have to go outside the system to perform searches on relevant literature. Content indexes and faceted browsing technology allows for the dynamic organization and filtering of data based on predefined system wide taxonomies. Also, cancer researchers can curate data and refine content in a very flexible fashion. This makes the entry and management of new cancer biomarker findings and research content easy to do.

The biomarker curated data in BioSEME system can now be of greater value to a cancer researcher than original data elements found in the *PPI's* protein biomarker list in Microsoft Excel spreadsheet format. Because of the semantic-integration of content, relationships and annotated data, the curator can add value to the biomarker research. The semantic- integration of data enables filters and search queries to return relevant material spanning multiple content types.

Even though not properly tested, the collaboration functionality provided by BioSEME compared to the Microsoft Excel spreadsheet, provides a true environment for data curation and a more dynamic mechanism for updating content in the ever changing cancer biomarker research.

Based on the metrics discussed above we believe that the BioSEME prototype was a success by reaching most of the goals outlined in the initial case study requirements. In the future, BioSEME has the potential to play a critical step in the effort to provide a more comprehensive picture of the complete state of research for a given cancer biomarker.

References

1. Polanski, M.: A List of Candidate Cancer Biomarkers for Targeted Proteomics. Biomarkers Insighs 2, 1–48 (2006)
2. NCI. National Cancer Institute Dictionary (2010) [cited]
3. Hart, A.F.: Enabling Effective Curation of Cancer Biomarker Research Data (2009)
4. Yang, I.S.: IDBD: Infectious Disease Biomarker Database. Nucleic Acids Research (2007)
5. Quan, D.: Improving life sciences information retrieval using semantic web technology. Briefings In Bioinformatics 8(3), 172–182 (2007)
6. Das, S., Girard, L., Green, T., Weitzman, L., Lewis-Bowen, A., Clark, T.: Building biomedical web communities using a semantically aware content management system. Briefing. In: Bioinformatics (2008)
7. Kelly, S. Biomarker Database Prototype [webpage] 2008 [cited], `http://cancer.jpl.nasa.gov/documents/applications/biomarker-database/biomarker-database-prototype`
8. Chen, J.Y.S., Mamidipalli, Huan, T.: HAPPI: an Online Database of Comprehensive Human Annotated and Predicted Protein Interactions BMC Genomics (October 2009)
9. Gupta, R.K.: Challenges in Clinical Trial Specimens Testing Laboratory and Importance of Establishing Specimen Biorepository. Sterling-Hoffman Life Sciences Journal (2007)
10. Nars, S.J., Moses, H.L. (eds.) Cancer Biomarkers: The Promises and Challenges of Improving Detection and Treatment (2007)
11. Olson, S.S.R., Giffin, R.: Accelerating the Development of Biomarkers for Drug Safety: Workshop Summary (2009)
12. Crichton, D.: A Distributed Information Services Architecture to Support Biomarker Discovery in Early Detection of Cancer (2006)
13. Brenner, D.E.: Biomarkers for Cancer Risk, Early Detection, and Prognosis: The Validation Conundrum. Cancer Epidemiol. Biomarkers Prev. (2007)

14. Clark, T., Kinoshita, J.: Alzforum and SWAN: the present and future of scientific web communities. Brief Bioinform 8(3), 163–171 (2007)
15. Ruttenberg, A.: Advancing translational research with the Semantic Web. BMC Bioinformatics (2007)
16. Ruttenberg, A.: Life sciences on the Semantic Web: the Neurocommons and beyond. Briefings in Bioinformatics (2008)
17. Belleau, F., et al.: Bio2RDF: Towards a mashup to build bioinformatics knowledge systems. Journal of Biomedical Informatics 41(5), 706–716 (2008)
18. W3C. A Prototype Knowledge Base for the Life Sciences (2008) [cited], `http://www.w3.org/TR/hcls-kb/`
19. Ceusters, W., Smith, B., Fielding, J.M.: LinkSuite*TM*: Formally Robust Ontology-Based Data and Information Integration. In: Rahm, E. (ed.) DILS 2004. LNCS (LNBI), vol. 2994, pp. 124–139. Springer, Heidelberg (2004)
20. Mooney, S.D., Baenziger, P.H.: Extensible open source content management systems and frameworks: a solution for many needs of a bioinformatics group. Briefings in Bioinformatics (2007)
21. Drupal. Drupal and the New Paradigm for Content Management (2008) [cited], `http://digitalsolutions.ph/couchkamotereviews/newCMS`
22. Stephane Corlosquet, R.D., Clark, T., Axel Polleres, S.D.: Produce and Consume Linked Data with Drupal! (2009)
23. Staab, S.: The semantic web — new ways to present and integrate information 4, 98–103 (2003)
24. Sioutos, N., et al.: NCI Thesaurus: A semantic model integrating cancer-related clinical and molecular information. Journal of Biomedical Informatics 40(1), 30–43 (2007)
25. Tolksdorf, R., Bontas, E.P.: Organizing Knowledge in a Semantic Web for Pathology. In: Weske, M., Liggesmeyer, P. (eds.) NODe 2004. LNCS, vol. 3263, pp. 39–54. Springer, Heidelberg (2004)

Age Dependent Differences in the Usage of a Desktop VR System for Air Force Mission Planning and Preparation

Carsten Winkelholz and Michael Kleiber

Fraunhofer FKIE, Neuenahrer Str. 20,
53343 Wachtberg, Germany
{carsten.winkelholz,michael.kleiber}@fkie.fraunhofer.de

Abstract. This paper presents results on age dependent differences in the usage of a desktop VR system for mission planning and mission preparation in the German air force. The study investigated two tasks: one mission preparation task that demands more fluid cognitive resources like short term memory and pattern recognition; and one mission planning task in which experience was more important to complete the task. In both tasks the usage of a desktop VR system was compared to the usage of traditional media like paper maps and aerial images. In both tasks no significant interaction of age with the used media was found. However, in the mission planning task there was a weak trend that a group of senior pilots used the VR system more intensely than a group of younger pilots. In the mission preparation task pilots had to memorize and recognize objects within a target area. Response latency in identifying memorized objects of the senior pilots was significantly higher than for the two groups of younger pilots. Furthermore, there was a significant interaction of the factor age with the query condition within the tested scenes. The response latency of senior pilots was significantly higher when a false target had been shown in the test condition, rather than the correct target. In contrast the response latency of young pilots was equal for both query conditions.

Keywords: Desktop VR, Human Factors.

1 Introduction

One part of the preparation for training targeting missions of the German air force is the study of the available information about the broader target area and the target objects itself. This includes maps, pictures and intelligence data in general. Most of the information provided to the aircrew is geospatial in nature. Currently, the data used in the preparation of a training targeting mission of the German air force is provided in a so called "electronic target folder" (ETF). The ETF basically consists of a set of interlinked electronic documents. Most of the documents are pictures or 2D maps of the target area. For the actual target and objects in the vicinity additional textual information is usually provided as well. In some cases detailed 3D models have been constructed based on photographs and intelligence data. Exploration and

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 292–300, 2011.

map viewing are two distinct methods for the acquisition of geographical information and they also lead to different cognitive representations [1, 2]. Therefore it is not clear whether for a specific task a presentation of a target area in 3D is more usefull than a 2D-map [3]. A modeled 3D scene depicts more details and could therefore lead the aircrew to focus on unnecessary features. It might therefore be preferable to show only two dimensional aerial images of the target area [4]. The aim of the primary study was to investigate the utility of a stereoscopic 3D-visualization of a target area for mission planning and mission preparation. Three major experiments were conducted for this purpose. One general experiment investigated distortion of depth perception using stereoscopic displays [5]. The second and third experiment were conducted with pilots of the German air force and investigated the utility of a 3D-visualization for a specific task [6, 7]. In a mission planning task pilots were asked to determine a final approach to attack a target. The experiment was designed to gather information on how a 3D-visualization could be used in a mission planning task. In the third experiment pilots used an approximated 3D-model of the target area to memorize the spatial structure of the target area. The focus in this experiment was on whether approximated 3D-models constructed from aerial images improve situational awareness. Both experiments gave evidences that 3D-visualizations are useful for mission planning and mission preparation. No effects were found comparing stereoscopic with monoscopic 3D-visualization. Pilots from three different sites participated in the experiments. On two sites participants were active pilots. On the third site participants were pilots retired from active flight duty, meaning they only complete 40 flight hours per year. Therefore, the latter group was clearly separated from the other two groups by age (younger groups: M=28y (group1) and M=33y (group2); older group: M=46y (group3)) and experience. All pilots in group 3 had completed more than 2000 flight hours each, whereas all pilots in group 1 and 2 had completed less than 2000 flight hours. Therefore we were able to perform a secondary data analysis comparing performance between these two groups. The results are presented in the following sections.

2 Experiments

The augmented workplace for the experiments consisted of a traditional 2D display and additionally of an autostereoscopic 3D display. We used an autostereoscopic display from SeeReal Technologies (Cn 3D Display Technology) which tracks the user and employs vertical lenticulars [8]. The basis of the 3D visualization software was a VR prototyping system [9] which allowed constantly adjusting the stereoscopic projection parameters like location of image plane, stereo base, and camera apex angle [5]. Adapting these parameters continuously ensured comfortable viewing of the stereoscopic images [10]. Navigation and orientation was done using a SpaceMouse with six degrees of freedom. The navigation technique was designed so that the pilots could on the one hand control the viewing parameters of interest and on the other hand always had an adequate stereoscopic presentation (for details see [6]). The standard 2D-PC mouse was only used for the selection and activation of objects in the scene. In the mission planning task the widget on the traditional 2D display provided an additional exocentric viewpoint which supports the pilots in judging their current orientation and heading towards the target on a conventional map (for a review of viewpoints and their implications see [11]).

2.1 Mission Planning

In the experiment for mission planning pilots were asked to plan the attack run for two targets which were made up for the purpose of the experiment only. The experimental setup is shown in Fig. 1. The targets were located in challenging terrain i.e. at the side of an elevation. In order to keep the required preparation and effort to a minimum we asked the pilots to plan only for the direction and heading and disregarded other aspects like weather or possible threats. There was no time limit given, but it usually only took about 10 minutes to decide on a route and draw the route onto the provided map. Group 1 and group 2 used paper maps and the aerial imagery at first. After they had decided on a flight route and drawn the route into the map they had the opportunity to inspect the target and surrounding terrain on the electronic planning system without stereoscopy. They had the opportunity after interactively inspecting the target to change their initial heading and direction. After the monoscopic inspection the view was changed to stereoscopy and the pilots again had the chance to re-evaluate their earlier decisions.

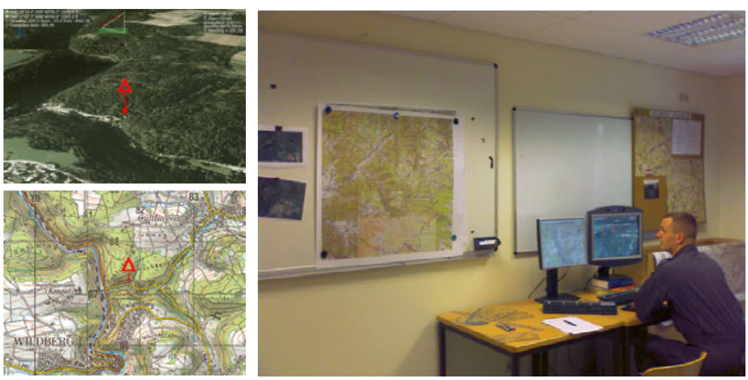


Fig. 1. Experimental setup for the mission planning task

A change of the final heading can occur because of different reasons. Firstly the pilot might have actually discovered new information from the additional 3D view or stereoscopic view. Another reason for a change can be the fact that the pilot is forced to reassess his earlier decision. In order to eliminate the second reason we showed the 3D view to group 3 first and the traditional maps second. One of the advantages of the traditional paper maps is that they allow for a better overview. Although we did provide an electronic overview map on the monoscopic display it does not provide the same easy overview because of the limited display size and resolution. It is therefore not obvious that the pilots will come to a better decision when using the electronic maps. The graph shows clearly that there was almost exclusively a change of heading when going from the papermap to the 3D view. Yet, there was no change after viewing the scene stereoscopically. The control condition i.e. going from the 3D to the 2D view also showed no change of heading. The graph in Fig. 2 shows the percentage instead of the absolute values as there were a different number of pilots for the different conditions. We used a contingency table for the statistical analysis. There is a significant effect for the change of the medium which results from the case of going from the 2D to the 3D view. A direct comparison between 2D → 3D mono and 3D → 2D (2×2 contingency table; $\chi^2 = 10.8$; $p < 0.001$) as well as 2D → 3D mono and 3D mono → 3D stereo (2×2 contingency table; $\chi^2 = 14.9$; $p < 0.001$) show significant effects for each. A comparison between 3D mono → 3D stereo did not show significance (2×2 contingency table; $\chi^2 = 0.527$; $p = 0.47$). Our hypothesis "the 3D view helps the pilots" is supported by the results. However, there is no evidence in the data which supports the hypothesis that a stereoscopic view gives an additional benefit.

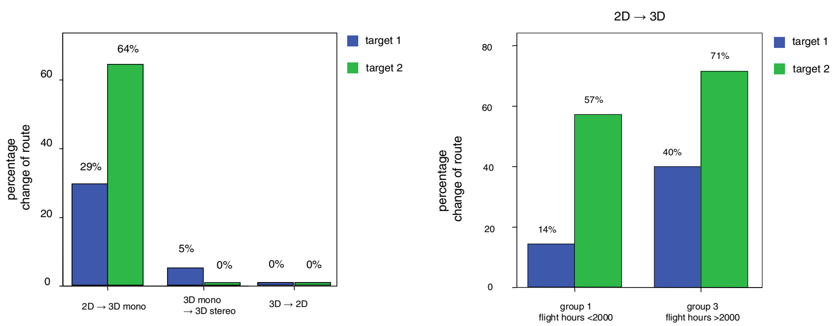


Fig. 2. Results from the mission planning task

The results of the comparison of the two groups of different age are shown in Fig. 2 on the right. The pattern between both groups is very similar. The older pilots seem to change their route more often after having analyzed the terrain in 3D. This seems to be especially true for target 1. A statistical test pooling data for both targets shows no significant effect (2×2 contingency table; $\chi^2 = 0.92$; $p = 0.337$). Also analysing data only for target 1 shows no significant differences between both groups (2×2 contingency table; $\chi^2 = 1.31$; $p = 0.252$). Therefore, this difference can optimistically be interpreted only as a trend. One reason that the more experienced pilots tended to change their flight route more often than the more inexperienced pilots, was in our notion that the senior pilots were much more confident in evaluating the options. Younger pilots seemed to be afraid to make mistakes, and since they had learned to determine the flight route by the paper map they used to keep the flight route they derived from the conventional material they know.

2.2 Mission Preparation

This experiment was motivated by statements of pilots, who claimed that they would prefer the usage of 3D-models only if the level of realism is very high, otherwise they suspected to be irritated during the mission, if the target area looked different. Only in rare cases detailed 3D-models of a target area exist. In general, information in the ETF is created by aerial image analysts by integrating and aggregating information from different sources. In some cases analysts estimate object heights by measuring the lengths of their shadows, comparing them with objects of known size or by using stereoscopic image analysis tools. The aerial image analysts can use the extracted dimensions to create 3D models of the objects at the target site. Although these techniques only allow approximations, integrating these methods into a software application facilitates the quick creation of basic 3D scenes. In effect one always needs to decide between accuracy (details) and cost (time) in the creation of 3D models. Therefore, this experiment focused on the question, whether the usage of approximated 3D-models of buildings facilitates or hinders the identification of targets in the actual mission.

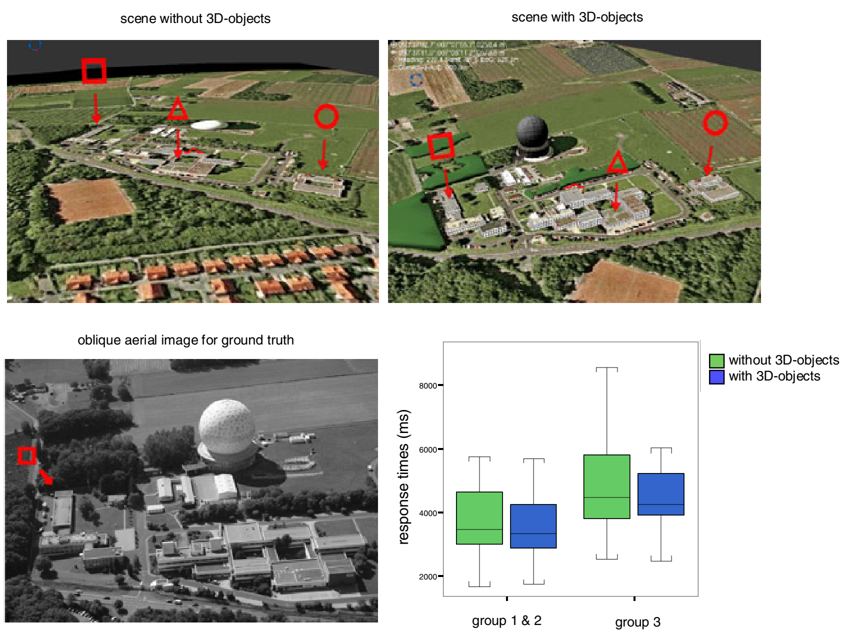


Fig. 3. Top row: Scene with and without 3D-objects target indicators. Bottom left: Oblique photographs with one target indicator as the test probe. Bottom right: Response latencies.

The procedure of the test in the mission preparation task consisted of two parts: first the subjects were shown a target scene within a desktop VR system in which two or three objects were labeled with different target indicators which had to be remembered; secondly the subjects were shown oblique photographs of the target area in which a single object was marked with a target indicator. The subjects' task was to respond as quickly and accurately as possible whether the combination of target indicator and target was present in the scene during the preparation. Two kinds of scenes were presented during the first phase, one set with and one set without approximated 3D-objects (Fig. 3). About two thirds of all shown oblique photographs were false combinations, this means that either a wrong object was indicated or a wrong symbol was used. We randomized the order of the photographs across subjects to prohibit effects of sequence. We recorded response times as well as accuracy of responses. The analysis of the response latencies in respect to the type of scenes (with, without 3D-object) showed that the approximated 3D-objects improved performance in the recognition of the target in the oblique image only by some 100 ms [10].

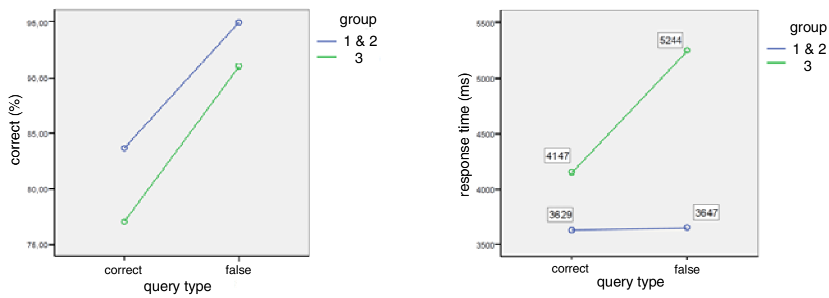


Fig. 4. Results for the performance in the target recognition task

The results of the evaluation showed very large differences in error rate and reaction times between users and also between the different 3D scenes. Nonetheless, the results are distributed normally, so that we were able to conduct analyses of variance (ANOVA). For the analysis of the correctness we differentiated between true and false oblique scenes. This means there were four possible combinations of answer and scene: true/false answer for true/false scene. The rate of correctness therefore includes the percentage of true images which were marked as true, as well as the percentage of false images which were marked as false. The reaction time was measured from the moment the oblique image was shown till the point the user clicked a mouse button to indicate his answer. We used seven different scenes and overall recorded about 1500 answers. The scenes differed in complexity and also in the number of available oblique photographs. Of the 1500 images shown, 1000 were false images and 500 were true images. Half of the answers belonged to scenes which included the approximate 3D objects. We first compared the three groups of pilots to assess whether the groups can be considered to be homogeneous. An analysis of variance with the factors target model type (2), query condition (2), scene group (2) and pilot group (3) concerning the correctness of the answers given did only show weak significance for the correctness ($F_{2,88}$=3.26 with $p<0.05$). A pair wise comparison between the two groups showed only weak significance between two groups. However, an analysis concerning the response time showed strong significance ($F_{2,88}$=9.25 with $p<0.001$). A pair wise comparison supported this and indicates that the group of older pilots (M_r=4.7s, SE_r=0.2) was significantly slower as either of the other groups (M_1=3.5s, SE_1=0.2; M_2=3.7s, SE_2=0.2). This finding was expected. However, the most interesting result was found investigating the interaction of the factors query type with the factor pilot group. Overall the response times were significantly slower ($F_{2,95}$=18.77; p<0.001), if the target shown was not present within the targets that had to be memorized. As is illustrated in Fig. 4, this effect mainly occurs in the group of the senior pilots, whereas in the group of the young pilots the response times were nearly equal for both query types. The interaction is significant ($F_{1,96}$= 5.00; p<0.05).

3 Conclusions

The most interesting finding was the interaction of the factor age with the factor query condition for the response latency in the target recognition task. The fact that the overall performance in the memory tasks starts to decrease by the age of 30 is well known. Therefore this finding was expected. The observed interaction suggests that the strategies to retrieve a representation of a building from memory and compare it with the building actually shown, differs for young and senior pilots. In general the target recognition task can be compared with the memory scanning task from Sternberg [12, 13]. In the Sternberg task participants have to learn a list of items. After retaining these items in memory a single item is presented and participants are asked to answer as fast as possible whether this single item was part of the list. The typical result is that response times increase with the number of items in the memory set, but that response times do not depend on whether the test item was part of it, say whether the correct answer was “yes” or “no”. Sternberg proposed a serial exhaustive scanning model in which participants compare the test item with all items in the list, regardless where in the memory list the test item occurs. A self-terminating scanning model would be more intuitive in which participants would stop scanning the memory set, when a match was found. But in a self-terminated scanning model one would expect that response latency would be higher if the test item was within the list, because on average participants could give a response after scanning half of the items in the memory set. However, the interpretation of the serial exhaustive model should not be taken literately. In modern theories of cognition like ACT-R [14], the behaviour of the serial exhaustive model can be modelled in a way that a participant tries to associate the test probe with the memory set. Response latency is determined within ACT-R by sub-symbolic parameters that consider the length of the memory list, but the access to the items in the lists assumed to be parallel and outside of consciousness cognition. The participant does not retrieve one item from the list step by step, but tries to use the test item and recall it directly from memory with a link to the learned list as an additional retrieval key. If this attempt fails he/she assumes that the item was not in the list. This strategy is fundamentally different from recalling the items of the memory step by step and comparing each item with the test item shown on the display. This serial recall process can also be described within ACT-R but the corresponding model would perform several memory retrieval requests whereas the recall of an association only needs one single memory retrieval request. However, the advantage of the serial recall is that the participant is able to reflect about the validity of his answer, whereas if the association of the item with the learned list fails it is not clear whether the item was not in the list or the list in memory was incomplete.

Transferring theses models on the mission preparation task suggests that senior pilots recall each target from the learning phase and compare it with the target in the test phase, whereas young pilots might use the target shown in the test phase and try to recall an association to a target from the learning phase. Referring to the previous discussion this would mean that the senior pilots might have reflected on the confidence level of their answer. The knowledge about the confidence level of the retrieval was not rewarded within the design of the experiment. The more reflective behaviour is consistent with the image of senior people and possibly compensates the general degradation of memory. In many real life situations also the answer “I don’t know” might prevent you to take an action that might be a mistake. However, in this scenario of targeting both strategies of recall would not lead to the attack of a false target. In some situations this behaviour is preferred, to avoid collateral damage, but in a situation, where a target causes damage if it is not identified, it might be vitally important to reflect about a false recall from memory. Therefore, the observed pattern in the response latency might depict a more deliberate procedure of the senior pilots, rather than their unconfidence. The result of the mission planning task showed that the senior pilots were self confident within the experiments. By trend senior pilots revised their opinion more often than the young pilots if new information from the 3D-visualization was available.

Acknowledgements. The research was carried out under a grant from the Federal Ministry of Defence (BMVg).

References

1. Tversky, B.: Remembering Spaces. Handbook of memory, pp. 363–378. Oxford University Press, New York (2000)
2. Mark, D.M., Freksa, C., Hirtle, S.C., Lloyd, R., Tversky, B.: Cognitive models of geographical space. International Journal of Geographical Information Science 13(8), 747–774 (1999)
3. John, M.S., Smallman, H.S., Bank, T.E., Cowen, M.B.: Tactical Routing Using Two-Dimensional and Three-Dimensional Views of Terrain. In: Proceedings of Human Factors and Ergonomics Society Annual Meeting, vol. 45, pp. 1409–1413 (2001)
4. Shah, P., Freedman, E.G., Vekiri, I.: The Comprehension of Quantitative Information in Graphical Displays. In: The Cambridge Handbook of Visuospatial Thinking, pp. 426–476. Cambridge University Press, Cambridge (2005)
5. Kleiber, M., Winkelholz, C.: Distortion of depth perception in virtual environments using stereoscopic displays: quantitative assessment and corrective measures. In: Proceedings of the Conference Stereoscopic Displays and Applications XIX, pp. 68030C (2008)
6. Kleiber, M., Winkelholz, C.: Case study: using a stereoscopic display for mission planning. In: Proceedings of the Conference Stereoscopic Displays and Applications XX, pp. 723704 (2009)
7. Kleiber, M., Winkelholz, C.: Usability of approximated 3D target objects within a desktop VR system for mission preparation of air force pilots. In: 3rd International Conference on Applied Human Factors and Ergonomics (2010)
8. Schwerdtner, A., Heidrich, H.: Dresden 3D display (D4D). In: Proceedings of the Conference Stereoscopic Displays and Virtual Reality Systems V, pp. 203–210 (1998)

9. Winkelholz, C., Alexander, T., Weiß, M.: Open ActiveWrl: a middleware based software development toolkit for immersive VE systems. In: Proceedings of the Workshop on Virtual Environments, pp. 321–322. ACM-Press, New York (2003)
10. Lambooij, M.: Stereoscopic displays and visual comfort: a review. SPIE Newsroom (2007), http://spie.org/x14604.xml
11. Wickens, C.D., Vincow, M., Yeh, M.: Design Applications of Visual Spatial Thinking. In: The Cambridge Handbook of Visuospatial Thinking, pp. 383–425. Cambridge University Press, Cambridge (2005)
12. Sternberg, S.: High-speed scanning in human memory. Science 153, 652–654 (1966)
13. Sternberg, S.: Memory scanning: New findings and current controversies. Quarterly Journal of Experimental Psychology 27, 1–32 (1975)
14. Anderson, J.R., Bothell, D., Byrne, M.D., Douglass, S., Lebiere, C., Qin, Y.: An integrated theory of the mind. Psychological Review 111(4), 1036–1060 (2004)

A Concept for User-Centered Development of Accessible User Interfaces for Industrial Automation Systems and Web Applications

Farzan Yazdi[1], Helmut Vieritz[2], Nasser Jazdi[1], Daniel Schilberg[2], Peter Göhner[1], and Sabina Jeschke[2]

[1] Institute of Industrial Automation and Software Engineering, University of Stuttgart, Pfaffenwaldring 47, D-70569 Stuttgart, Germany
{farzan.yazdi,nasser.jazdi, peter.goehner}@ias.uni-stuttgart.de

[2] Institute of Information Management in Mechanical Engineering, RWTH Aachen University, Dennewartstraße 27, D-52068 Aachen, Germany
{helmut.vieritz,daniel.schilberg, sabina.jeschke}@ima-zlw-ifu.rwth-aachen.de

Abstract. The importance of industrial automation systems and Web applications, often in combination, has been a growing area during the past decades. They are becoming an inseparable part of our lives. Hence they must be accessible to all users. Often certain user groups are being neglected in the development of such systems. Therefore, a systematic concept is required to support the development of such systems. In this paper, we will present a concept, which is proposed in the context of an ongoing research project, addressing the accessibility problem of the user interface. Both, industrial automation systems and Web applications share similar accessibility requirements. Hence, our concept addresses both systems, while discussing the suited methods to effectively assess accessibility requirements.

Keywords: Accessibility, User-centered Design, User interface modeling, Model-driven development.

1 Introduction

To overcome the problem of accessibility one can start with the development process or try to provide a run-time solution. Our concept focuses on the development process. Many critical decisions upon system structure are being already made in the early stages of development [1]. Therefore, an early assessment is very reasonable.

Web-based applications are increasingly gaining importance in our lives for a wider range of users, including those with disabilities. Hence, they must be accessible to all their users. To overcome such problems, a wide range of accessibility guidelines, e.g. WCAG 2.0, for Web applications have been created. Yet, they do not address the development process; a systematic approach for developing accessible Web applications is missing.

Similarly, industrial automation systems have become an inseparable part of our lives, e.g. ticket vending machines. With the phrase *industrial automation systems* we

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 301–310, 2011.

refer to product automation or technical devices [2]. The big competition and saturated market has forced the producers to try to gain a bigger market fraction by embedding more new functionalities into their industrial automation systems. Hence, usage of such systems is getting more complex. However, often certain user groups are being neglected in development of such systems. Lack of a systematic approach for directing the development and the diversity of industrial automation systems make assessing accessibility to them cumbersome [3, 4]. The existing solutions are very dependent on the underlying technologies. Hence, a generic development solution is still missing.

In the upcoming sections we will provide some basic methods, which are used in the proposed development concept for accessible *user interfaces* (UIs). Section 2 contains the state of the art. In section 3 we will describe our concept and in section 4 a conclusion will bring this paper to the end.

2 State of the Art

User-centered Development (UCD) is a possible approach to embed accessibility into the development process. It involves the end users in the development process. Here the development is done in several iterations, while in each iteration developers are being provided with the feedbacks of the users on using the under development system. The iterative nature of UCD offers parallel treatment of system design, definition of functionalities and UI design. Hence, an appropriate architecture pattern is required to support this parallelization. The basic idea is to separate UI layout and control from system core functionalities. This may be realized by different alternatives, including the separation of front- and backend, a client-server concept, the Model-View-Controller (MVC) architecture pattern, or three-tier architecture.

Furthermore, Model-driven Development (MDD) is among the important methods used in combination with UCD. Typically, UML-based methods are being employed by the developers for the purpose of modeling. However, modeling of accessibility requirements of UI acquires a strong focus on user activities, which is not trivial using standard UML. Furthermore, it must provide a possibility of modeling UI elements abstract enough to be independent of the underlying technology. The existing solutions for modeling UI can be grouped into two categories: 1) practical approaches, e.g. *User Interface Markup Language* (UIML) [5], *User Interface Description Language* (UIDL) [6], or *Extensible Application Markup Language* (XAML) [7] and 2) analytical approaches, e.g. *User Interface Extensible Markup Language* (UsiXML) [8] or the *Unified Modeling Language for Interactive Applications* (UMLi) [9].

Practical approaches are relied on platform-independent development of UIs, using certain development tools. Recent implementations such as Microsoft Silverlight [10] or Adobe Flex [11] provide the separation of UI layout from UI behavior as well as from system core functionalities. Whereas, analytical approaches are more abstract and include a meta-model for all UI aspects, which describes UI structure, components, behavior, navigation, domain, and resources. However, the analytical approaches are still under research and there is lack of tools for actual development. The existing analytical approaches for the modeling of UIs, e.g. useML, UMLi [9], *Task Modeling Language* (TaskML) and *Dialog Modeling Language* (DiaMODL) integrate the requirements of the modeling process including analysis.

Few modeling solutions address complete and systematic integration of accessibility in model-driven design and more importantly in the development process [3, 12].

3 A UCD-Based Concept to Develop Accessible UIs

Our concept combines UCD and MDD, since UCD is focused on user's requirements and modeling allows describing particular aspects and views on the UI as navigation, state or behavior. The combination of UCD and MDD allows a clear analysis and declaration of accessibility onto UI design. Later on, the Abstract Presentation Model (APM) provides a semantic description of UI navigation, views, structure, elements, their roles, states and behavior. Finally, existing development tools can be used to implement the UI. Moreover, our concept considers that MVC architecture pattern for modeling is used; the separation of presentation (view) and functionality (controller) allows focusing on the UI with little interference to the system logic.

The concept addresses UI accessibility problems of industrial automation systems and Web applications, where often some user groups have difficulties interacting with the system, e.g. the ticket vending machines, which are not applicable for those with visual impairment or cell phones. Furthermore, it covers both industrial automation systems and Web applications by defining a generic approach which is applicable to both fields and limiting the target system specifications to guidelines and recommendations.

Industrial automation systems and Web applications may have different accessibility requirements, as far as the technology is related. However, both have accessibility requirements, which rely on similar principles. Based on this principle, the concept tries to provide the developers with guidelines on what factors needs to be considered. For this purpose, a repository of static helps has been included in the concept. It contains e.g. accessibility guidelines. In the following sections different aspects of the proposed concept will be described.

3.1 Accessibility

Accessibility in UI design means free perception and control of all relevant information for all users, also for those with disabilities [13]. Here, the information refers to those being needed by the user to follow his intended workflow constituted by his activities, actions and their relations. Despite the importance of accessibility, it is still a big challenge integrating it to the industrial automation systems, due to numerous reasons. Some reasons are:

- Variety of the requirements of different user groups
- Limiting UI technologies e.g. cell phones with small GUIs
- Difference between developer's mindset with that of the user

The existing accessibility solutions are focused on accessibility guidelines, e.g. on the Web Content Accessibility Guidelines (WCAG) [14]. They mostly aim at run-time behavior of the system, i.e. how the system should work but not how it should be developed. Additionally, they are specifically created for Web applications. As long as human interaction with UI is concerned, industrial automation systems have similar accessibility requirements as Web applications. Therefore, the guidelines can be adopted for industrial automation systems and UCD is a promising approach for the UI development of industrial automation systems, since both systems require similar accessibility requirements, regarding their UI. Moreover, industrial automation systems and Web applications are being often used in combination. Projects like W3C Mobile Web Best Practices 1.0 [15] have used such guidelines for industrial automation systems, e.g. to overcome the accessibility problems of the elderly using cell phones. In that example more than 70% of the guidelines could be employed for the industrial automation system [15]. Hence, our concept uses WCAG as the basis for addressing the critical points regarding the accessibility.

3.2 Multimodality

Multimodal interaction offers users to use various interaction channels in interfacing with a system. It uses different senses of human being, e.g. hearing or sight, for input and output of data. A simple example for a multimodal interaction can occur in a navigation system in a car, which can be interfaced via touchpad or voice input. The goal of a multimodal interaction is to use *optimally* the users' capabilities, by considering extra senses of the human users in the human-computer interaction (HCI) [16]. In comparison to traditional interfaces (single-modal) multimodal interaction allows users to communicate more naturally and interface with complex information with more freedom of expression [17]. In [16] multimodality is considered as a combination of interface styles, which can be used to improve the individual components in isolation so that the combination works better. However, this has to be differentiated from instantiating several copies of a UI in systems like in Mac systems. The proposed concept uses the approach used in [16] for parameterization of the interaction channel.

3.3 Methodology

The proposed concept consists of two parts; the *Generic Parallel Process* (GPP) and the *Accessibility-supportive Repository* (ASR) (see Fig. 1). GPP is a supportive process running in parallel to the typical development process. It helps developers to identify critical spots that are relevant to accessibility, during the analysis and design phase, as these two phases shape the basis of the under development system. GPP, then, provides the developers with tools and guidelines to take actions and assess the identified accessibility requirements. The identification and assessment of accessibility requirements are being supported by ASR, which is a predefined repository containing reusable components such as accessibility guidelines, design patterns and norms, which are generally necessary for an accessible UI.

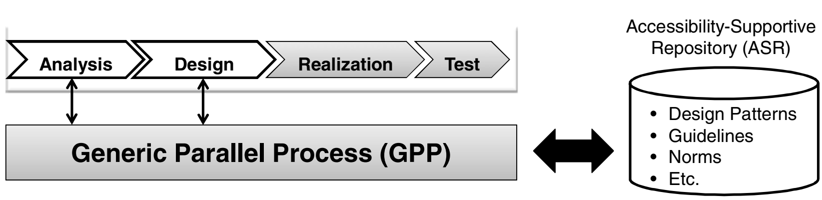


Fig. 1. An overview on the proposed concept

3.4 Generic Parallel Process in Analysis Phase

In the analysis phase the requirements of the system are being collected. While in the traditional approach, a great focus on functional requirements is given, GPP tries to help developers identify accessibility requirements (none functional requirements). Besides methods like *appreciative inquiry* or *structure laying technique*, GPP investigates sight, hearing, motoric capabilities and knowledge (the 6^{th} sense) of a user in the interaction with the systems [16]. The aforementioned aspects can be affected in two ways: (1.) by user impediments, e.g. visual impairment, and/or (2.) external constraints, e.g. voice input in a noisy environment. Considering these aspects with the two affecting sources, the accessibility requirements can be analyzed. For example, Table 1 shows interaction aspects and constraints for a coffee dispenser and its process of maintenance.

Table 1. Speculations of user impediments and his/her external constraints in the example of coffee dispenser

	Sight	**Hearing**	**Motoric**	**Knowledge**
User impediments	Very good	Very good	Very good	Technician
external constraints	High light intensity	Very Noisy Environment	Back and front side of the machine must be accessed. Access at the same time needed.	n. a.

Out of the information provided in Table 1, following accessibility requirements can be concluded.

1. Visual interaction with the system is possible.
2. Technical staff functionalities are required.
3. High contrast of the UI elements is required.
4. Voice input and output is not possible for the user.
5. Multimodal interaction is needed.

Another important step is to model with their correlations and conflicts. This will help identifying requirements that need to be eliminated or might be added to the requirements set. Currently, the concept assumes that this step of analyzing requirements is to be done by project managers or similar authorities.

For the purpose of modeling requirements, the *requirements graph* method presented in [1] is used. This method is very suitable for the purpose of modeling none-functional requirements [1]. It arranges requirements as graph nodes, connected to each other by labeled arcs. The labels are “hurt”, “help”, “make”, or “brake”, with “Hurt” referring to a conflict, “Make” to dependency, “Help” to recommendation, and “Hurt” to discouraging relation between the requirement nodes. The intensity of each relation can be defined using an intensity value from the set *I={--, -, 0, +, ++}*, where “--” is the strongest negative and “++” is the most positive intensity value [1].

3.5 Generic Parallel Process in Design Phase

In the design phase, the concept supports the developers to design the UI in parallel to the design of the system logic. It helps identifying critical points in the interaction that affect the accessibility. In this phase a scenario-based approach is being employed for designing the structure of the system. Once the scenarios are defined, all cases of the scenario are being modeled. The proposed concept employs a UML-based approach, since it is universally used, hence it can be easily used in parallel to designing the system logic. Furthermore, it is easily extendable using UML profiles. Focusing on the *architecture pattern*, *workflow*, and *relevant information* on the end users, three aspects of the interaction with UI is being modeled: task modeling, dialogue modeling, and presentation modeling. Despite existing methods, which are very much syntax based, the concept uses guidelines aimed at the aforementioned criteria for modeling the three aspects of interaction. The syntactical complications of modeling UI are being removed semantically. Fig. 2 depicts the important guidelines for each criterion, used in the concept.

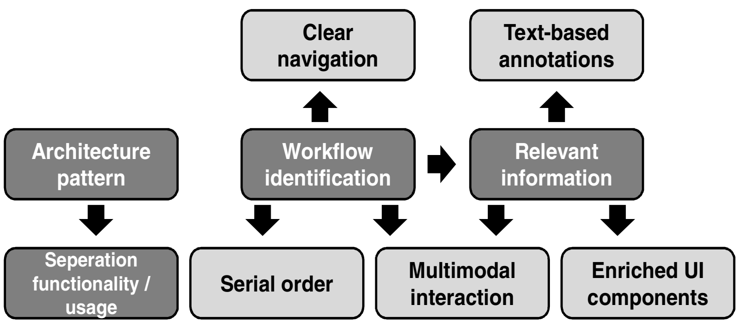


Fig. 2. The important guidelines for each criterion

As it can be seen in Fig. 2, the knowledge about user's workflow is necessary to determine the relevant information on the end user, since it is only being considered in the usage context. The relevant information is being stored using a text-based annotation such as XML. The separation of data model/core-functionality and UI-presentation/control enables multimodal concepts for usage and different layouts for particular input/output devices. Using a well-suited architecture pattern, e.g. MVC, this separation can be easily achieved. Furthermore, deployment of multimodal interactions is tightly dependent on knowing the workflow and the relevant information. Design of interface components in a serial order – corresponding to user's workflow – to support Assistive Technology (AT) such as screen readers and structuring the interface navigation distinctly and avoiding deep hierarchical structures, in order to support sitemaps and breadcrumbs for instance are further guidelines used to model UI. In Table 2 the sequence of UML diagrams, which should be implemented in each of the three aspects of interaction with UI are listed. The diagrams in each row should be considered together, while modeling.

Table 2. Overview of the UML diagrams in the GPP (AD – UML activity diagram, UD – use case diagram, CD – class diagram, SC – state chart)

Task Modeling Diagrams	Dialog Modeling Diagrams	Presentation Modeling Diagrams
Use case – UD		
Activities – AD	Modes, macro-navigation – AD	Views, main & utility navigation, site-map – CD
Actions – AD	Interactors, Roles of UI components – CD	UI elements, attributes & behavior – CD
Work objects (UI data) – AD		
Workflow – AD	Micro-navigation (Sequences, alternatives), UI behavior – SC	Order of UI elements – CD

3.6 Accessibility-Supportive Repository

ASR feeds the GPP with design patterns, guidelines, norms and standards, which have been collected. Based on existing patterns, the concept includes a pattern structure as it can be seen in Fig. 3. The design pattern contains information on the problem, usage context and solution, illustration, reference to other patterns or code/design idioms. Using the patterns developers can design each step more easily. The design patterns are based on a selection/adoption of WCAG 2.0 accessibility guidelines [14]. The patterns can be categorized into four categories: perceivable, operable, understandable and robust; each addressing a certain aspect of accessibility [14]. Fig. 3 illustrates the various categories of the accessibility patterns and their connection to diagrams.

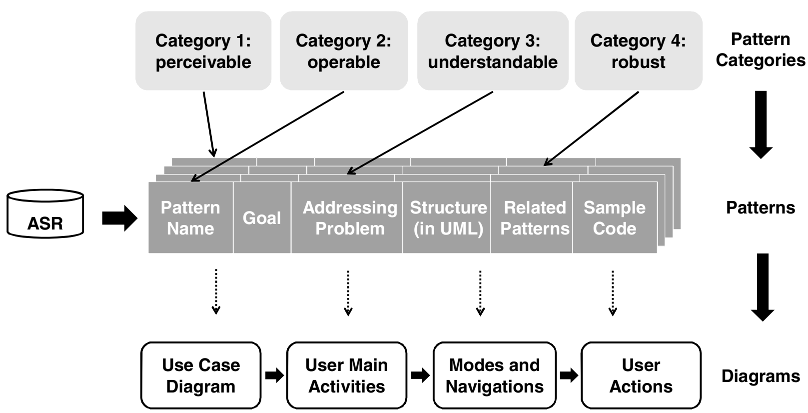


Fig. 3. Various categories of the accessibility patterns, their instances and derivation of design diagrams

The UI of a coffee dispenser as a case study shows examples of the GPP analysis and design phase. Fig. 4 shows the first step in analysis – the identification of user modes. Based on the use cases, main activities are detected. In the example of Fig. 4 three main activities “Change product”, “View product”, and “Search product” are detected. Correspondingly, the user workflow is grouped by modes of usage.

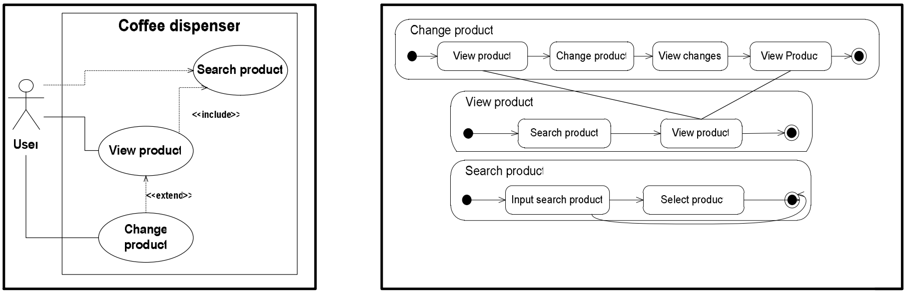


Fig. 4. Use cases and user main activities

Fig. 5 shows the main navigation network of modes drawn in dashed boxes. The relation of each node describes the possibilities of moving from one mode to another one. Here the technological description is not of interest. The selection of a drink, in this example, might be a simple read through the buttons on a coffee dispenser or a sound module reading out the available drinks.

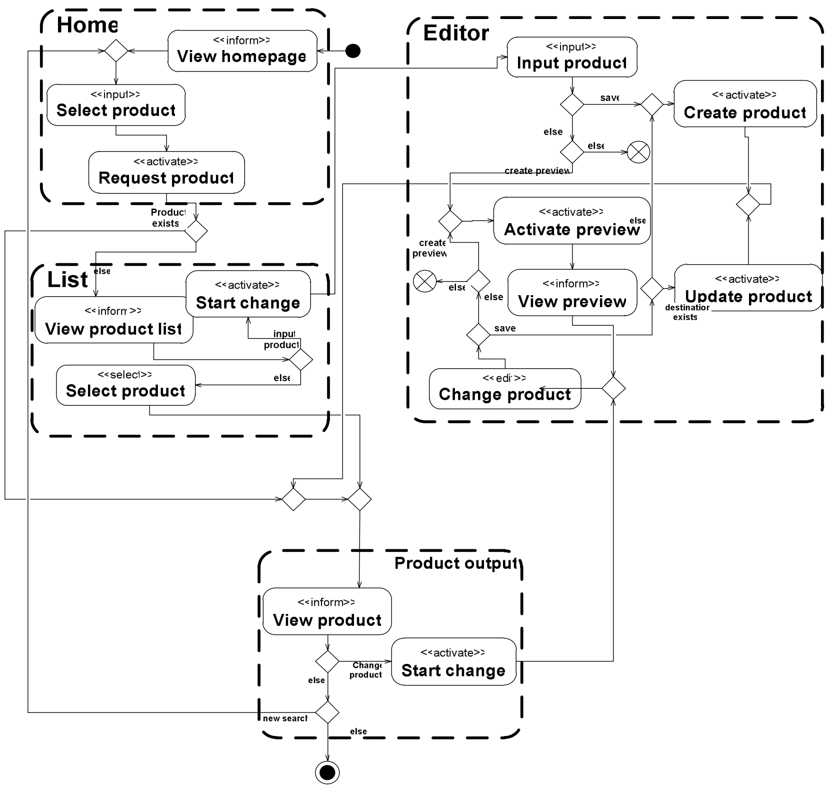


Fig. 5. UI main navigation modes and user's actions within each mode in the example of coffee dispenser

Furthermore, each navigation mode contains several actions. For instance, once the user has decided upon selecting his drink, he might move to the corresponding navigation page. Then he can take several steps, i.e. actually selecting his desired drink, reviewing his selection or canceling it. Here, each mode itself is being refined by internal actions. The actions are based on atomic action categorization described in [18]. The actions are defined as five stereotypes *select*, *activate*, *inform*, *input*, and *edit*. In the concept the two steps of defining mode navigation and user actions are being done using separate diagrams. In this paper, for more compactness we have shown both in one diagram, in Fig. 5.

4 Conclusion and Outlook

The proposed concept is based on the survey made on the state of the art and consists of a set of suitable methods. It provides a parallel process to the development process, while considering the end users.

We have discussed the potential of UCD and MDE in maintaining accessibility in developing UIs. Having focused on the early development stages, i.e. analysis and design phase, we have investigated suitable methods and technologies, which can assist developers by designing an accessible UI for industrial automation systems and Web applications.

The future work of our project is focused on an automated identification and analysis of the requirements, a systematic bridge between analysis and design phase, and an automated transformation of the detected diagrams.

Acknowledgements. We highly appreciate Deutsche Forschungsgemeinschaft (DFG) for funding this project.

References

1. Cysneiros, L.M., Werneck, V., Kushniruk, A.: Reusable Knowledge for Satisficing Usability Requirements. In: 13th IEEE International Conference on Requirements Engineering, pp. 463–464 (2005)
2. Lauber, R., Göhner, P.: Prozessautomatisierung I, 3rd edn. Springer, Heidelberg (1999)
3. Nieminen, M.: Information Support for User-oriented Development Organisation, Dissertation for the degree of Doctor of Science in Technology, Helsinki University of Technology Department of Computer Science and Engineering (2004)
4. Göhner, P., et al.: Integrated Accessibility Models of User Interfaces for IT and Automation Systems. In: Proceedings of the 21st International Conference on Computer Applications in Industry and Engineering (2008)
5. UIML community. User Interface Markup Language (UIML) (2009), `http://www.uiml.org/`
6. UIDL community. User Interface Description Language (UIDL) (2008), `http://www.uidl.net/`
7. Microsoft Corp. Extensible Application Markup Language (XAML), `http://msdn.microsoft.com/en-us/library/ms747122.aspx`

8. UsiXML community. User Interface Extensible Markup Language (UsiXML) (2010), http://www.itea.defimedia.be/
9. Da Silva, P.P.: Object Modelling of Interactive Systems: The UMLi Approach. University of Manchester (2002)
10. MacDonald, M.: Silverlight 2 Visual Essentials. Apress (2008)
11. Tapper, J., Boles, M., Labriola, M., Talbot, J.: Adobe FLEX 3. Addison-Wesley, Reading (2008)
12. Jeschke, S., Pfeiffer, O., Vieritz, H.: Developing Accessible Applications with User- Centered Architecture. In: ICIS 2008, Seventh IEEE/ACIS International Conference on Computer and Information Science, pp. 684–689. IEEE Computer Society, Los Alamitos (2008)
13. ISO13407:1999, Human-centred design processes for interactive systems (1999)
14. World Wide Web Consortium (W3C). Web Content Accessibility Guidelines 2.0 (2008), http://www.w3.org/TR/WCAG20/
15. World Wide Web Consortium (W3C). Mobile Web Best Practices 1.0, http://www.w3.org/TR/mobile-bp/
16. Stiedl, T.: Multimodal interaction with industrial automation systems. Dissertation at IAS, University of Stuttgart (2009)
17. Taib, R., Ruiz, N.: Integrating semantics into multimodal interaction patterns. In: Popescu-Belis, A., Renals, S., Bourlard, H. (eds.) MLMI 2007. LNCS, vol. 4892, pp. 96–107. Springer, Heidelberg (2008)
18. Reuther, A.: useML – Systematische Entwicklung von Maschinenbediensystemen mit XML, Universität Kaiserslautern (2003)

Part IV

Well-Being, Health and Rehabilitation Applications

Forms of Interaction in Virtual Space: Applications to Psychotherapy and Counselling

Shagun Chawla and Nigel Foreman

Psychology Department, Middlesex University, UK
{s.chawla,n.foreman}@mdx.ac.uk

Abstract. Electronic virtual communication is likely to become increasingly used and relied upon in the future, particularly because en-face interactions between client/patients and professionals will become less convenient and health systems overloaded. However, this raises interesting issues about the form of communication which is best suited for the imparting of authoritative advise to clients/patients, how to ensure that advice is treated seriously and followed, the types of condition and situation that most lend themselves to virtual (as opposed to en-face) therapeutic interactions, and how behaviours in virtual situations varies from en-face reality. The present paper discusses some of these issues and data will be presented from preliminary studies aimed at addressing the above issues.

Keywords: Virtual, en-face, psychotherapy, counseling, avatars, second- life, personality, youth.

1 Introduction

Many professionals believe that the applications of virtual and distant interactions will become increasingly used in the future, as health systems become overloaded, and more sophisticated intelligent software becomes available. The benefits would be substantial and would include the preservation of anonymity (if required), particularly for disorders that clients/patients find embarrassing to discuss en-face. Some groups of individuals might be especially likely to use virtual interaction in preference to en-face interactions; for example teenagers, who are usually very familiar with electronic media and are used to obtaining information on-line, may benefit since this is a client group that is often hard to access by conventional means [1]. Any such group that tends to rely especially heavily upon informal advice from peers [see 1] is likely to be drawn to the use of virtual interactions in a similar way, and thus it is important to ensure that the information that is imparted and received is accurate, also that the advice is regarded as authoritative so that it is followed, and clients/patients feel comfortable about returning to follow-up interactions. Clearly, in countries where distances are significant, on-line therapies and advice becomes especially important, and it is interesting that early advice help-lines were pioneered in Australia, for example.

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 313–321, 2011.

The availability of high speed broadband has meant that en-face interactions can also be used for interaction, supplementing the familiar e-mail and electronic messaging services. However, not all users of Skype report feeling comfortable when interacting, particularly with unfamiliar people. The anonymity of less personal interactive media can potentially be an advantage. Moreover, depending on the particular medium, it is possible to revisit conversations and maintain records of interactions with some media but not (easily) others. Clients/patients will often report that they feel reassured by being able to return to conversations and revisit interactions with those providing advice.

1.1 What Areas of Concern Might Be Addressed?

Clearly, the above also relates in important ways to the concept of universal access. The widespread use of a range of media is likely to encompass a much broader social group than would be expected from the use of any single medium. Of course, this may depend upon the topic concerned. Therapeutic or advisory sessions might cover a range of issues, but those featuring commonly to date include marriage and relationship counselling advice, deficient social skills, sexual health and fertility, depression and suicidal tendencies, anxiety and associated neurotic disorders, financial worries and associated family tensions, smoking and smoking cessation, stress counselling including examination stress, parenting issues, gender identity, among others. Clearly, the approach to the design of counselling sessions and advisory interactions must be tailored to the issue concerned, and most likely to the individual, taking into account their circumstances and personality. Since the availability of on-line counselling and therapy is a relatively new phenomenon, such issues have not been addressed, so that a one size fits all approach tends to have been adopted. Barak, Klein and Proudfoot [2] comment that the field [variously termed web-based therapy, e-therapy, cybertherapy, eHealth, e-Interventions, computer-mediated interventions, on-line therapy and on-line counselling] has “suffered from a lack of clarity and consistency... inconsistent, diffuse, incoherent and sometimes even perplexing” (pp.4/5). The availability of suitable software and interactive media has been regarded as the primary goal, and only now might it be possible to use software to collect data on such things as whether individuals in virtual interactions speak as much as in en-face situations, whether they feel able to raise sensitive issues, how content they feel with the interactions, whether they take advice as seriously when obtained from an on-line medium, whether they are as likely to follow advice as when prescribed by a professional en-face, and how these issues relate to the type of issue being discussed. For disorders that involve any form of depersonalisation, for example, it might be positively disadvantageous to deal with a client/patient in a virtual medium. On the other hand, a patient suffering from a neurotic fear might be encouraged to use software and pre-recoded guidance, to take therapy at their own pace, gradually exposing themselves to stimuli that approximate more closely the stimulus that they most feat while desensitising themselves in a way that maintains their comfort and dignity. The availability of software that can be easily authored to suit client needs can facilitate such developments.

Sessions can be organised in a variety of ways, depending on the resources available, the suitability of media, and the degree of interaction required. It is not the intention of this project, in the first instance, to deal with areas that are highly ethically sensitive or where any shortcoming in the advice provided might pose potential risks to clients. There is a wide range of issues that have been addressed via on-line services, which might be addressed using web-based educational interventions, self-guided web-based therapeutic interactions, or human-supported web-based therapeutic interactions [2]. At one end of the scale, en-face advice and direct personal interactions can be indispensable, particularly in an area such as PTSD, or depression and suicide-prevention counselling. In many clinical situations, therapist may find it impossible to discuss a sensitive issue with a client without having feedback from facial expressions, hesitations and other features of conversations that are hard to obtain from virtual media. At the opposite end of the scale, it is possible that many concerns of clients/patients can be effectively met via the use of information alone – frequently asked questions (FAQs) can be accumulated over time and new clients may find that their concerns are so similar to those of many previous enquirers that the provision of personalised advice is unnecessary. Between these extremes is, for example, interaction via avatar representations. Here, the individual is able to interact in a relatively personal way but their identity is disguised and they may feel more comfortable in raising sensitive issues within this medium. An example might be a virtual consulting room in a Second Life environment. The advantage here is that the advisor/therapist is aware that they are dealing with a particular client and advice can still be tailored to the needs of that individual; clearly, however, in that case there is no saving on advisor/therapist time, since their presence is always required in the interaction. Group sessions involving interacting avatar representations in Second Life might hold possibilities, of course. As Barak et al. [2] have pointed out, feedback is a crucial issue, and the total amount provided to a patient can vary from minutes to hours, and "Immediacy of response is dependent upon which communication modality is being employed... e-mails and forum postings generally provide delayed feedback, whereas chat room/instant messaging sessions, Skype/telephone calls. Webcam and face-to-face meetings all provide immediate feedback; yet these different communication modalities also vary in the degree of direct human contact". Clearly this needs to be addressed and quantified for the different areas for which web-based advice might be sought.

2 Features of Computer- Mediated Communication

Computer- mediated communication (CMC) especially the internet, has become one of the key areas for psychological research today due to a variety of reasons. Firstly, increasing number of computer and internet users all over the world. Secondly, a high volume of interpersonal communication takes place over the internet [3] implying that there is significant amount of social behaviour worthy of exploration [4]. And finally, a few features of CMC such as visual anonymity and limited channel have been implicated in a variety of online behaviours of individuals in the form of self-disclosure and in many cases heightened private self awareness.

Recent studies have shown that there is a higher level of self- disclosure due in CMC than en- face due to visual anonymity and heighted private self- awareness [5, 6, 4, 7, 8, 9]. Even medical patients tend to report more symptoms and undesirable behaviours when interviewed by a computer as appose to en- face [10]. Moreover, in a study conducted by Ferriter [11] on ‘computer aided interviewing and psychiatric social history’, it was demonstrated that patients spoke more truthfully and openly during interviews conducted using CMC compared to en- face. Other studies have also shown that individuals are more inclined to higher levels of self- disclosure on electronic surveys as compared to paper based surveys. Participants gave out more information about socially undesirable behaviours while taking psychological measures on electronic survey systems [12]. Such studies lend evidence to the fact that “technology can gather information of greater quantity and higher quality than clinician- administered assessments” [13].

3 Therapy or Games?

The interactive session does not need to be in the form of one to one dissemination of information and advice; a self-help client-centred approach can be adopted via the use of a computer gaming format that allows clients to direct their own therapy [1]. Working in Dublin, Ireland, Coyle and colleagues used Personal Investigator (PI), which employs a detective narrative, in which the teenager hunts for solutions to personal problems. There is collaboration between client and therapist but the emphasis is on clients’ “[setting] their own therapeutic goals, recognise their own strengths and values, identify people in their lives who can help them, teach new coping strategies and focus on their future not the past”. The approach appears to have been successful with a small group of clients having anxiety and behaviour problems, attempted suicide and social skills difficulties.

Recently another study using a game format has been conducted where participants are allowed to explore their own identities via a multi- player online role playing game. The players have to create an “embodied representation of themselves” [14]. After creation, the characters have to travel the virtual world gaining knowledge, skills and defeating monsters. The results of the study showed that the players on an average rated their representation as having more favourable attribute than themselves and this tendency was more prominent in players with lower psychological well-being [14]. The authors have further discussed that the discrepancy between the actual- self and ideal- self may be reduced through the creation and enacting of their ideal self online. These findings can have a great impact in terms of using virtual therapy with individuals having low self- esteem, depression or addiction.

4 Online Self- Representation: Avatars and Second- Life

Today, a number of virtual spaces online allow users to communicate and interact by creating an avatar that is a digital representation of themselves. Over the years, a variety of 3-D virtual worlds have developed such as Second- Life, There and Active Worlds which are mainly based on avatar interactions. Such virtual worlds create strong feelings of social presence among the users thereby, increasing the “feeling of togetherness” [15]. This may be a contributing factor to a number of therapeutic applications that that have evolved on Second- Life. For example, Brigadoon- a private island created to provide a virtual space for individuals with Asperger's Syndrome. The aim is to provide the users with a platform to enhance their social skills by interacting with other people experiencing the same problems. Similarly Live2Give, another second life island, is targeted at individuals affected by cerebral palsy. Recently, a new island, Eureka has been developed for addiction prevention and treatment by assisting people in getting greater self- insight and improving their lifestyles [15].

Research in the area has led to several interesting findings on avatar behavior in virtual space. In a recent study, it was found that users generally tended to adapt their avatars to reflect their own appearance as a result of which they experienced heightened private self- awareness [9]. In another study, it was demonstrated that users revealed more personal information about themselves when interacting with avatars as oppose to real humans [16]. Hence, one might conclude that avatars may lead to the enrichment of interpersonal communication thereby, strengthening the interpersonal bond among users. This may be a key factor in the case of a therapeutic setting, especially while therapists engage in rapport building with their respective clients.

5 Activity-Passivity

An issue that frequently arises in the context of 3-D interactive activities, where a participant interacts with a 3-D environment, is the issue of activity-passivity. It is widely assumed that active engagement with an environment leads to better learning (usually, spatial learning) than passive observation of the exploration of another individual. However, that has not always been found to apply. Whether activity is beneficial is arguably due to several factors, including the familiarity of the user with the medium – usually computer-related experience [17]. It is important to take into consideration the familiarity of users of therapeutic interactive environments with the medium, to ensure that they can fully attend to the information being provided rather than being distracted by their operation of input devices and coping with interface requirements. However, passive engagement could be beneficial in some instances. In the case of an individual who is hesitant about seeking advice and unsure of what questions to raise with their virtual therapist, it might be reassuring for that individual to observe a recorded interaction between two virtual agents, especially if there are frequently asked questions and frequently raised issues that could be incorporated into the interaction, that would have the effect of answering their own uncertainties without their having to divulge their identity or interact directly. Once again, this issue could potentially load on to the concept of universal access. Often the concept of universal access is applied to situations in which individuals might be prevented or discouraged from engagement as a result of physical or cognitive disabilities, although it should be recognised that access may be restricted for reasons such as gender and personality. Adolescents, in particular, are apparently discouraged from engagement with therapists because they feel that no-one can help with their difficulties, or that the problem, they feel, is too personal to divulge, or that they should handle the problem on their own [18]. In some cases it might be a helpful start to a therapeutic process if such individuals can be reassured that their problem is more widespread than they think and that others are wrestling with the same issues that confront them.

Ethically, it is, of course, essential that steps are taken to ensure that individuals who need alternative forms of therapeutic intervention are provided with those interventions or with the routes to accessing them, and that the use of 3-D interactive software does not delay their obtaining en face assistance when it is needed.

6 Individual Differences among Potential User Groups

The one size fits all approach can ignore important individual differences, such as gender differences in computer use and familiarity which might differentially affect an individual's comfort in using computer based media and interface familiarity can affect the level of spare cognitive capacity that an individual can bring to bear on the therapeutic interaction, if some users could be so engaged with use of the medium that this detracts from their engagement with the information gathering and interaction processes. Second, personality differences (introversion, openness) can influence the degree to which a computer-mediated interaction is attractive to the client. In fact, one of the factors that might indicate the level of self- disclosure online may be the personality type of an individual. According to Suler [8], individuals with histrionic personality types may be very open and emotionally expressive whereas compulsive individuals maybe more restrained and controlling. So far, research in this area has shown that individuals with a more trusting personality in terms of a secure attachment style and generalised feelings of trust tend to engage in more disclosure type related behaviours [5]. However, it will be extremely useful to carry out research with a focus on personality types in order to predict the circumstances under which different individuals will be more predisposed to greater self-disclosure. One may also wonder whether people with avoidant personality styles may benefit from various anonymous online environments or if the different CMC formats such as chat forums, web – sites, emails and online groups, may be detrimental to people with dissociative disorders by leading to fragmentation of different online personalities [19].

Other factors that may determine the 'suitability' of any individual for online therapy [19] could be, firstly, the individual's knowledge about the psychological aspects of online communication such as lifestyle in cyberspace, online groups, relationships and social other activities. Secondly, an individual's reading and writing skills: Some people might prefer one over the other and this might have a significant impact on text based CMC. Thirdly, cross cultural differences which might be magnified during online therapy, especially if the client and therapist are from different countries. Fourthly, an individual's medical condition or certain disabilities or medications that might affect his or her motivation to use online therapy. Few individuals might prefer CMC to en-face communication as they might have social phobia or might want to conceal their physical appearance [19]. Finally, an individual's attitude towards online therapy may also affect the efficiency of intervention employed by the counsellor or therapist. It is envisaged that brief evaluative exercises could be used to decide, in particular instances, to which medium an individual would be best suited.

7 Primary Objectives of the Project

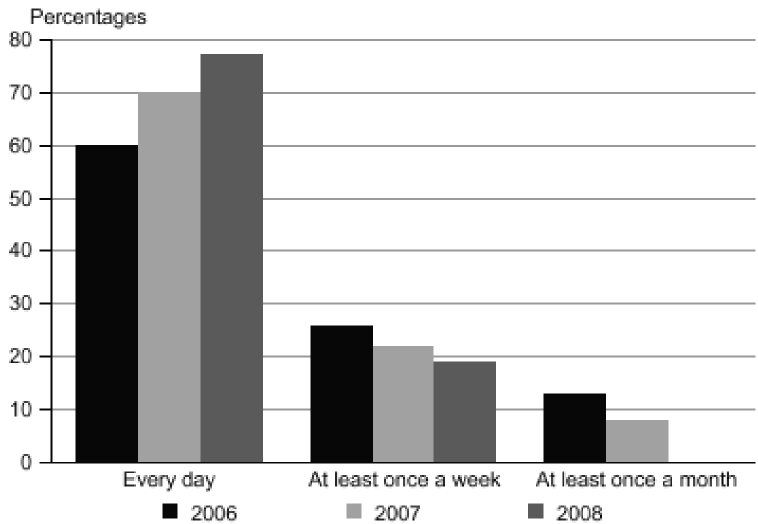
1. To use suitable questionnaires to identify those issues for which individuals report that they would be most, or least, likely to prefer to access on-line information and support rather than conventional consultations.
2. To identify the best means of conveying information associated with particular disorders or concerns, comparing media such as e-mail, Skype, or avatar-based interactions.
3. To assess the degree to which individuals feel comfortable in using a particular medium, relating this to demographic, gender and personality measures.
4. To assess whether the medium employed preserves essential features of the conventional consultation situation; whether clients feel as confident in posing questions or challenging the advice provided.
5. To assess whether clients who participate in web-based consultations take advice seriously and the extent to which they are likely to follow advice delivered by various media.
6. To identify the degree to which the above issues are specific to the area concerned, comparing among the various issues for which web-based advice is most likely to be sought.

7.1 Target Population: Younger People

For the purposes of this project an effort will be made to understand the online advice and counselling needs of the younger population (between 16 and 25 years of age) for the prevention of mental health difficulties. The main motive for choosing this particular group is because it has been observed that young people are generally reluctant to seek help due to a variety of reasons such as feeling that their problem is personal to them, or no one will understand them, and hence no one may be able to help and also due to concerns regarding confidentiality [1, 20] due to which counsellors and therapists often find it difficult to engage young people in the therapeutic process [18]. Therefore, young people who are faced with mental health issues or other personal problems are more likely to turn to informal sources such as their peer groups and even the internet in order to obtain more information and seek help [21]; using the internet anonymously and in an autonomous manner [20]. Hence, as Coyle and colleagues [22] have rightly said "it is incumbent upon mental health services to find ever more creative ways to engage young people and in ways that reflect their own cultures and forms" (p. 27).

Also, it is important to note that young people use the internet regularly for various activities like communication, social networking, information seeking and much more. In the United Kingdom, nearly 77% of young adults aged 16- 24 years reported using the internet everyday in 2008 [23]. This percentage is very likely to have increased today.

Due to the increase in online activities of the younger population today, the Internet becomes a strategic tool to provide this generation a means for online mental health promotion through early intervention and prevention [20].



(Source: Office for National Statistics, 2008)

Fig.1. Frequency of Internet use: by recent users aged 16 to 24, UK

References

1. Coyle, D., Matthews, M., Sharry, J., Nisbet, A., Doherty: Personal Investigator: A Therapeutic 3D Game for Adolescent Psychotherapy. Journal of interactive Technology and Smart Education 2(2), 73–88 (2005)
2. Barak, A., Klein, B., Proudfoot, J.G.: Defining internet-supported therapeutic interventions. Annals of Behavioral Medicine 38, 4–17 (2009)
3. Kraut, R., Mukhopadhyay, T., Szczypula, J., Kiesler, S., Scherlis, B.: Information and Communication: Alternative Uses of Internet in Households. Information Systems Research 10, 255–263 (2000)
4. Joinson, A.N.: Self- disclosure in Comuter- mediated Communication: The Role of Self-awareness and Visual Anonymity. European Journal of Social Psychology 31, 177–192 (2001)
5. Frye, N.E., Dornisch, M.M.: When is Trust not Enough? The Role of Perceived Privacy of Communication Tools in Comfort with Self- disclosure? Computers in Human Behavior 26, 1120–1127 (2010)
6. Goss, S., Anthony, K.: Developments in the Use of Technology. British Journal of Guidance and Counselling 37(3), 223–230 (2010)
7. Rochlen, A.B., Zack, J.S., Speyer, C.: Online Therapy: Review of Relevant Definitions, Debates, and Current Empirical Support. Journal of Clinical Psychology 60(3), 269–283 (2004)
8. Suler, J.R.: The Online Disinhibition Effect. Cyberpsychology & Behavior 7(3), 321–326 (2004)
9. Vasalou, A., Joinson, A.N., Pitt, J.: Constructing my Online Self: Avatars that Increase Self- focused Attention. In: Proceedings of CHI 2007, Online Representation of Self, San Jose, USA (2007)

10. Greist, J.H., Klein, M.H., VanCura, L.J.: A computer Interview by Psychiatric Patient Target Symptoms. Archives of General Psychiatry 29, 247–253 (1973)
11. Ferriter, M.: Computer Aided Interviewing and the Psychiatric Social History. Social Work and Social Sciences Review 4, 255–263 (1993)
12. Kiesler, S., Sproull, L.S.: Response Effects in the Electronic Survey. Public Opinion Quarterly 50, 402–413 (1986)
13. Newman, M.G.: Technology in Psychotherapy: An Introduction. Journal of Clinical Psychology/In Session 60(2), 141–145 (2004)
14. Bessiere, K., Seay, A.F., Kiesler, S.: The Ideal Elf: Identity Exploration in the World of Warcraft. CyberPsychology & Behavior 10(4), 530–535 (2007)
15. Gorini, A., Gaggioli, A., Vigna, C., Riva, G.: A Second Life for eHealth: Prospects for the Use of 3-D Virtual Worlds in Clinical Psychology. Journal of Medical Internet Research 10(3), 1–11 (2008)
16. Kang, S.H., Gratch, J.: The Effect of Avatar Realism of Virtual Humans on Self-Disclosure in Anonymous Social Interactions. In: CHI 2010: Work in Progress, Atlanta, USA (2010)
17. Sandamas, G., Foreman, N., Coulson, M.: Input device training reinstates active benefit in children when exploring virtual environments. Spatial Cognition and Computation 9, 96–108 (2009)
18. Offer, D., Howard, K.I., Schonert, K.A., Ostrov, E.: To Whom do Adolescents Turn to Help? Differences between Disturbed and Non- disturbed Adolescents. Journal of American Academy of Child and Adolescent Psychiatry 30(4), 623–630 (1991)
19. Suler, J.R.: Assessing a Person's Suitability for Online Therapy: The ISMHO Clinical Case Study Group. CyberPsychology & Behavior 4(6), 675–679 (2001)
20. Burns, J., Ellis, L.A., Mackenzie, A., Reicher, J.S.: Reach Out: Online Mental Health Promotion for young People. Counselling, Psychotherapy and Health 5(1), 171–186 (2009)
21. Gould, M.S., Munfakh, J.L., Lubell, K., Klienman, M., Parker, S.: Seeking Help from the Internet During Adolescence. Journal of the American Academy of Child and Adolescent Psychiatry 4(10), 1182–1189 (2002)
22. Coyle, D., Sharry, J., Nisbet, A., Matthews, M.: Virtual Perspectives: Developing a Therapeutic 3D Virtual Environment for Adolescents. Eisteach: A Quarterly. Journal of Counselling & Psychotherapy 2(25), 27–31 (2003)
23. Office for National statistics. Internet Access: 2008 Households and Individuals (2008), `http://www.statistics.gov.uk/pdfdir/iahi0808.pdf` (accessed December 7, 2010)

Control of Powered Prosthetic Hand Using Multidimensional Ultrasound Signals: A Pilot Study

Xin Chen, Siping Chen, and Guo Dan

Department of Biomedical Engineering, School of Medicine, Shenzhen University
Shenzhen Key Lab of Biomedical Engineering
Shenzhen 518060, China

Abstract. Various sophisticated signal processing techniques have been developed for EMG control strategy. However, some inherent properties of the signal prevent it from providing a natural control of powered prosthetic hand. This paper reported on a pilot study of an alternative, multidimensional ultrasound signals which can collect the architecture change of muscle during contraction. We designed a system to collect and analyze multi-channel A-mode ultrasound, joint angle, and surface EMG simultaneously. Using this system, we investigated the feasibility of controlling powered prosthesis by multidimensional ultrasound signals. In the experiment, the subjects were instructed to perform the wrist extension and flexion movement. Multi-channel ultrasound signals, collected from the forearm muscle, were used to estimate the wrist angle and then to control the power prosthesis. The results suggest that multidimensional ultrasound signals, based on further improvement, have great feasibility to be an alternative method to control prosthesis.

Keywords: Ultrasonography, Prosthesis, Multidimensional Ultrasound, Muscle.

1 Introduction

Powered prostheses have been used for decades to provide an artificial extension for amputees. A typical powered prosthesis is comprised of mechanical and electrical components, capable of extracting features or patterns from the electrophysiological signal to drive the mechanical actuators.

Electromyography (EMG) signal has been widely used for controlling prosthetic devices [1]. Various signal processing techniques have been proposed for EMG control strategy [2]. All the efforts aim to improve the flexibility of EMG control along with the reduction of control complexity for users. However, there still exist some inherent limitations of EMG control which are difficult to overcome. For example, it is difficult to provide a natural control of the prosthesis with multiple degrees of freedom (DoF) based on multi-channel EMG signals. Most of the current commercially available hand prostheses still use two-channel EMG inputs to provide one or two DoF(s) [3]. Brain activity is another kind of electrophysiological signal that can be potentially used for the prosthetic control. This control approach is often termed as brain-machine interface (BMI), human-machine interface (HMI), and neuroprosthesis [4] Various brain signals have been intensively investigated such as

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 322–327, 2011.

EEG, MEG, ECoG, and cortical neural signal [5]. However, there are still many conceptual and technological obstacles before developing neuroprosthetic devices for clinical applications. In addition to the electrophysiological signal, mechanical signals generated by muscle contraction have also been used for prosthetic control, including mechanomyography (MMG) signal [6] and myokinemetric (MK) signal [7].

Due to its ability to reflect the architecture of muscle and tissue, ultrasonography has been used to estimate muscle dimensional change [8]. Zheng et al reported the continuous monitoring of muscle activity by real-time ultrasound imaging and first suggested its potentials for the prosthetic control [9]. They also developed an A-mode ultrasound system to investigate feasibility of the new control method [10]. Based on this system, the performance of real-time ultrasound control for powered prosthesis with normal subjects was investigated [11].

The aim of this study was to extend ultrasound signal from one dimension to multi-dimension and investigate the potential for controlling prosthetic hand. A system equipped with multi-channel A-mode ultrasound was applied to detect dimensional changes of forearm antagonistic muscles. The relationships between 2-D ultrasound signals and wrist angle were quantitatively studied. The results were used to assess the potential of multidimensional ultrasound signals as a noninvasive method for prosthetic hand control.

2 Methods

2.1 Hardware Setup

As shown in Fig. 1, two ultrasound transducers (model V129; GE Panametrics, Inc., West Chester, OH, USA) were used to measure the dimensional change of antagonistic muscles. The ultrasound signals were transmitted, received and digitized by an 8-channel ultrasound phase array system (PCIAD850, PHA8; US Ultratek Inc, CA, USA). The pulse repetition frequency was 100 Hz and the sampling frequency was 50 MHz. An electronic goniometer (TSD130; BIOPAC System Inc., CA, USA) was attached in the middle of the posterior hand to measure the wrist angle during wrist extension. The collected angle signal was digitized by the NI DAQ card (NI-DAQ 6024E; National Instruments Corporation, Austin, TX, USA) with sampling rate of 2.0 KHz. A prosthetic hand (MH22; Shanghai Kesheng Prostheses Co., Shanghai, China) was controlled to open and close by the analog pulse outputted from the same DAQ card. The signal acquisition, synchronization, and display, as well as prosthetic control tasks were achieved using a custom-designed software developed in Labview.

2.2 Experiment Procedure

Two healthy subjects volunteered to participate in this study. In the experiment, each subject was seated in a comfortable chair with his forearm resting on the table. The ultrasound transducers were attached to the belly of extensor carpi radialis and flexor carpi radialis respectively. Each subject was asked to perform wrist extension for two cycles with guided moving speed. Six repeated trials were performed, with a rest of 5 minutes between two adjacent trials. The angle of the wrist was detected by the goniometer simultaneously. Both ultrasound and angle signals were digitized and stored in the computer for further processing.

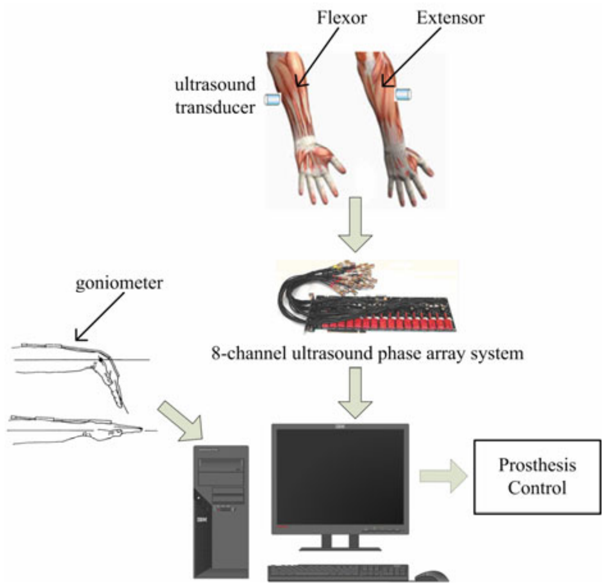


Fig. 1. The illustration of the hardware setup. Two ultrasound transducers were applied to measure the dimensional change of antagonistic muscles. The electronic goniometer was attached to the forearm to detect the wrist angle.

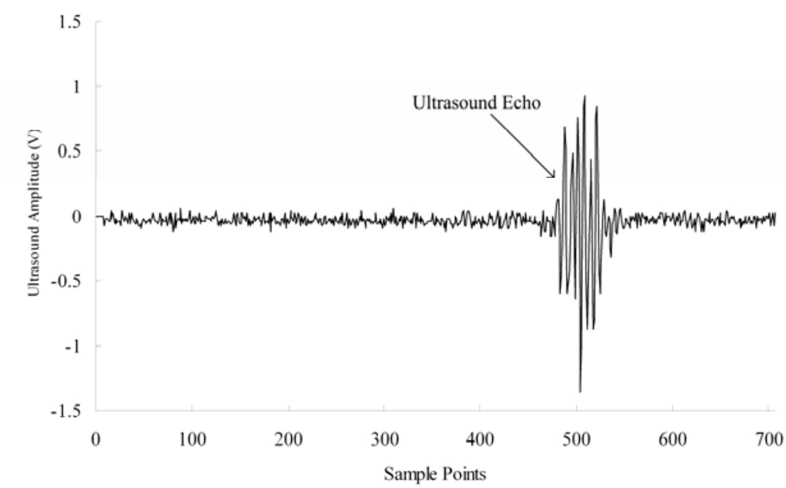


Fig. 2. A typical frame of one channel A-mode ultrasound signal. The position of the ultrasound echo reflects the muscle deformation.

2.3 Signal Processing

All the signals were processed and analyzed in Matlab program. Two-channel A-mode ultrasound signals were captured respectively frame by frame according to pulse repetition frequency. A typical ultrasound frame of one channel was shown in Fig. 2.

The A-mode ultrasound echo reflected from the muscle-bone interfaces was manually selected in the first frame as reference. To estimate the echo shift, which reflected the dimensional change of specific muscle, we utilized the cross-correlation algorithm to search the segment of signal most similar to the reference echo signal in each frame. The equation used to calculate the normalized one-dimensional cross-correlation coefficient was as follow:

$$R_{xy} = \frac{\sum_{i=0}^{N-1}[x(i)-\bar{X}][y(i)-\bar{Y}]}{\sqrt{\sum_{i=0}^{N-1}[x(i)-\bar{X}]^2 \sum_{j=0}^{N-1}[y(j)-\bar{Y}]^2}} \tag{1}$$

where $x(i)$ is the reference signal, $y(i)$is the selected signal, and $\bar{X}$ and $\bar{Y}$ are the means respectively. During the search process of each frame, the segment of signal with maximal cross-correlation coefficient was considered as the echo signal. Two echo shift series, $S_1(n)$, $S_2(n)$, were extracted from two ultrasound channels. The dimensional changes of extensor and flexor muscles in wrist extension movement were represented by these echo shift series.

The wrist angle signal was segmented as 32-point epochs. The middle of each epoch was aligned in time with the corresponding ultrasound frame according to the timestamp; hence the epochs were synchronized with the echo shift sequence in time domain. The root mean square (RMS) values $Z(n)$ were calculated for each epoch.

2.4 Data Analysis

Multivariable linear regression was conducted to model the relationship between forearm movement and dimensional changes of extensor and flexor muscles. The regression equation is as follow:

$$Z(n) = \beta_1 S_1(n) + \beta_2 S_2(n) + \varepsilon(n) \tag{2}$$

where β_1, β_2 are the regression coefficients, ε is the random noise.

Ordinary Least Squares (OLS) method was applied to estimate the coefficients in Eq.2. This OLS estimator minimizes the sum of squared residuals, and leads to a closed-form expression for the estimated value of β_1 and β_2. For each subject, the data sets of the first three trials were used for estimation and the data sets of other trials were used to test the estimation performance which was quantified using the normalized root mean square error (RMSE) by the following equation, where $\hat{Z}(n)$ is the estimated angle.

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{N}}\sum_{n=1}^{N}(Z(n)-\hat{Z}(n))^2} \tag{3}$$

2.5 Prosthesis Control

The estimated angle signal was then applied to investigate its potential in prosthesis control. It was used to open and close the prosthesis through the DAQ card. The amplitude of the signal linearly correlated with the opening position of the prosthesis. The linear algorithm helped the subjects to control the prosthetic hand according to his wrist angle with a few conscious efforts.

3 Preliminary Results

For each subject, the data sets of the first three trials were applied for the multivariable linear regression. The relationship between the wrist angle measured by the goniometer and the corresponding estimated angle in the other trials was further investigated. Fig. 3 illustrates a typical data set of the measured and estimated angles. The overall mean value of the normalized RMSE was 12.5%. In the prosthetic hand control experiment, the hand could smoothly open and close based on the linear control. Further study is being conducted to quantitatively evaluate the control performance.

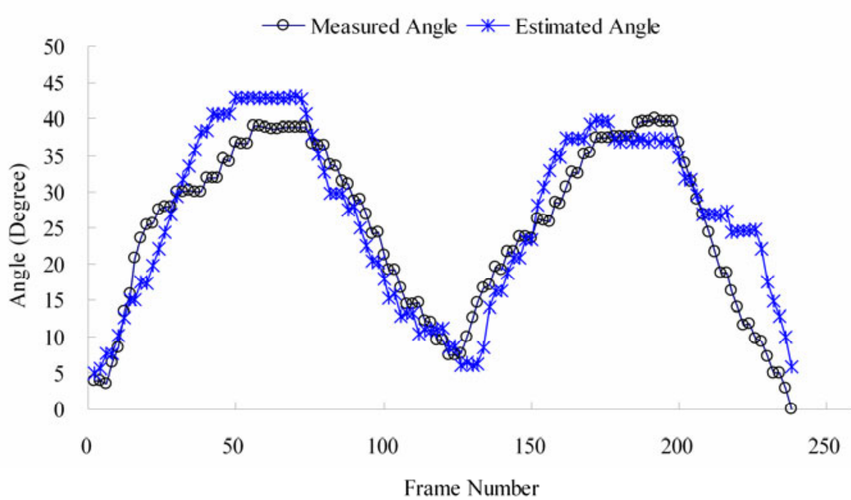


Fig. 3. Measured and estimated wrist angles calculated from the 2-channel ultrasound data in a typical trial

4 Discussion

In this study, the feasibility of using multidimensional ultrasound signal to control powered prosthesis was investigated. Based on the previous work that the morphological changes of forearm muscles extracted from one dimensional ultrasound can be used to control powered prosthesis, we custom-designed a system to synchronously collect multi-channel ultrasound, joint angle, and EMG signals of forearm muscles. Multivariable linear regression was used to estimate the wrist angle during flexion-extension movement. The estimated angle was further applied to control the opening-closure function of the prosthetic hand and the control performance was qualitative evaluated. The preliminary results indicated that the prosthetic hand could be well controlled by the proposed method. However further studies are required to quantitatively investigate the control performance and compare it with the traditional EMG control.

Acknowledgments. This work was supported by the National Natural Science Foundation of China (Grant No. 81000637), the State Key Program of National Natural Science of China (Grant No. 61031003), the National Basic Research Program 973 from Ministry of Science and Technology of China (Grant No. 2010CB732600).

References

1. Muzumdar, A.: Powered upper limb prostheses: control, implementation and clinical application. Springer, Berlin (2004)
2. Oskoei, M.A., Hu, H.S.: Myoelectric control systems-A survey. Biomedical Signal Processing and Control 2, 275–294 (2007)
3. Parker, P., Englehart, K., Hudgins, B.: Myoelectric signal processing for control of powered limb prostheses. J. Electromyogr. Kinesiol. 16, 541–548 (2006)
4. Nicolelis, M.A.L.: Actions from thoughts. Nature 409, 403–407 (2001)
5. Lebedev, M.A., Nicolelis, M.A.L.: Brain-machine interfaces: past, present and future. Trends Neurosci. 29, 536–546 (2006)
6. Silva, J., Heim, W., Chau, T.: A self-contained, mechanomyography-driven externally powered prosthesis. Arch. Phys. Med. Rehabil. 86, 2066–2070 (2005)
7. Heath, G.H.: Control of proportional grasping using a myokinemetric signal. Technology and Disability 15, 73–83 (2003)
8. Bianchi, S., Martinoli, C.: Ultrasound of the Musculoskeletal System. Springer, Berlin (2007)
9. Zheng, Y.P., Chan, M.M.F., Shi, J., Chen, X., Huang, Q.H.: Sonomyography: Monitoring morphological changes of forearm muscles in actions with the feasibility for the control of powered prosthesis. Med. Eng. Phys. 28, 405–415 (2006)
10. Guo, J.Y., Zheng, Y.P., Huang, Q.H., Chen, X., He, J.F., Lai-Wa Chan, H.: Performances of one-Dimensional Sonomyography and Surface Electromyography in Tracking Guided Patterns of Wrist Extension. Ultrasound Med. Biol. 35, 894–902 (2009)
11. Chen, X., Zheng, Y.P., Guo, J.Y., Shi, J.: Sonomyography (SMG) Control for Powered Prosthesis Hand: A Study with Normal Subjects. Ultrasound Med. Biol. 36, 1076–1088 (2010)

A Top-k Analysis Using Multi-level Association Rule Mining for Autism Treatments

Kelley M. Engle and Roy Rada

Department of Information Systems,
University of Maryland Baltimore County (UMBC),
Baltimore, MD 21250 USA
{kelley.engle,rada}@umbc.edu

Abstract. Association rule mining is based on associations of attribute values in a database. To facilitate finding meaningful rules, we segment the database by a categorization of database records based on a taxonomy on one of the attribute value sets. To test the value of this approach we have applied it to a large database about treatment impacts on autistic children. The segmented analyses lead to interestingly, different results from the analyses done without segmentation.

Keywords: association rule mining, autism, data mining.

1 Introduction

Medical data analysis is challenging due to the inherently complex and critical nature of the data and the far reaching implications of the analysis. In this paper we propose a novel approach to combine domain knowledge with an association rule mining technique to generate association rules. Our application domain is treatments for autism.

1.1 Data Mining Rules

Association rules identify collections of data attributes that are statistically related in the underlying data. An association rule is of the form X => Y where X and Y are disjoint conjunctions of attribute-value pairs. To illustrate, imagine a set of attributes {milk,bread,butter,beer}. The data comes from a grocery where each customer is represented by a conjunction of bought or not bought values for each of that attributes. If customers who buy bread always also buy butter, then one association rule could be 'bread is bought' => 'butter is bought'.

The confidence of the rule is the conditional probability of Y given X, Pr(Y|X), and the support of the rule is the prior probability of X and Y, Pr(X and Y) where probability is the observed frequency in the data. The traditional association rule mining problem is:

- Given a database of transactions, a minimal confidence threshold and a minimal support threshold,

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 328–334, 2011.

- find all association rules whose confidence and support are above the corresponding thresholds.

Rules that meet the minimum support and confidence thresholds may be further analyzed for their correlation values. The correlation for {X, Y} is computed as the support value for that pair, divided by the square root of the product of the support values for X and Y. Rules that have a correlation value that is greater than some minimum correlation value will be retained.

Often in association rule mining it can be difficult to generate rules due to the sparsity of the data at the atomic level [1]. Multi-level association rule mining alleviates this problem by segmenting the data according to a user-defined hierarchy. Such segmenting will be performed in this research.

1.2 Autism

Autism is a neurodevelopmental disorder that appears during childhood and generally follows a steady course without remission. The characteristic triad of symptoms involves impairments in social interaction; impairments in communication; and restricted interests and repetitive behavior. Numerous disorders overlap in their symptomatology with those of autism, and the terminology and diagnosis in this area is often clouded but incidence can be broadly stated to be several children per thousand children.

For autistic children, no single treatment is best, and many, different approaches are often tried with each child. On average, an autistic child is on seven different treatments at any given point in time [2]. The efficacy of the majority of these treatments is not supported by clinical trials with few exceptions. In terms of available pharmacological agents, only risperidone and Ritalin have replicable results for the autistic population [3].

The Interactive Autism Network (IAN) web project is designed to serve as a national registry for autistic children and their families. Its purpose is two-fold: (1) to connect researchers to large, diverse subject bank and; (2) to collect data for secondary analysis and release the de-identified data to researchers. The IAN web project has registered thousands of families. There are more than 500 treatments in the IAN dataset. These treatments include medications, vitamin and mineral supplements, diets, behavioral/educational interventions as well as alternative treatments designed to treat the autistic child holistically. It would not be feasible or prudent to conduct clinical trials on such a large number of treatments. Data mining, for this particular scenario, would be helpful in reducing the solution space to determine the best treatments. The efficacy for this smaller group of treatments could then be properly confirmed in a clinical trial. Multiple analyses would be advantageous in pruning the solution space in order to overcome any limitations in one particular method.

1.3 Autism and Data Mining

The vast majority of autism research is conducted using "traditional" analysis methods such as statistics. There are few studies that have incorporated data mining methods for the autism domain. Because the current knowledge regarding autism is limited, borrowing methods from data mining could be a valuable aid in theory-building. In the following paragraphs, the few studies that have been conducted using methods from the data mining area are presented.

Self-organizing maps were used to model attention shift impairment and familiarity preference, both hallmarks of autistic behavior [4]. This work concluded that the familiarity preference exhibited by autistic patients may be in fact the primary cause of other autistic characteristics such as shift impairment.

Heart rate patterns were compared with behavioral problems such as self-injury and aggression which are common in the autistic population [5]. The analysis was conducted using LERS (Learning from Examples using Rough Sets) that generates classification rules from the presented data. The learning mechanism in LERS is similar to the bucket brigade algorithm [6, 7]. The analyses indicate that both high and very high heart rates are associated with problem behaviors. Further research avenues were discussed and included physiological data with behavioral treatment plans. The analyses of EEG data to support research in epileptology and autism was also presented [8].

An expert system was developed as a screener for developmental disorders including autism [9]. The purpose of the system was to enable clinicians in countries where specialists, such as developmental pediatricians, are few in number, diagnose autism and other developmental disorders in children at an earlier age. Text mining was applied to biomedical literature on autism in order to "identify potential contributions to a better understanding of autism [10]." The results of the text mining allowed for the construction of an ontology which identified "rare relations" in the autism data. This research led to the development of RaJoLink [11, 12], a literature mining method for "uncovering hidden relations from a set of articles in a given domain."

Genetic and environmental factors were examined using CFA (Combinatorial Fusion Analysis) and association rule mining to determine associations between autism prevalence and the exposure to mercury and lead during critical stages of a child development [13]. The results of these analyses indicate an association between an increase of mercury concentration in the environment with an increase in autism prevalence three years later.

2 Method

We use an association rule mining tool called Weka (www.weka.org) to implement a novel approach to rule mining. Our approach utilizes an external treatment ontology and multi-level association rules. We discuss detailed experimental results demonstrating the efficacy of our framework. Utilizing a host of measures including: (1) support, (2) confidence, (3) lift, (4) leverage, (5) conviction and, (6) the chi-square test, we were able to prune the original rules generated to include only those with strong positive correlation.

Our approach has two components:

- Association rule mining conducted on the entire treatment dataset and;
- Multi-level association rule mining based on a dataset segmented according to the treatment ontology.

Furthermore, the multi-level association rule mining was conducted at two sub-levels:

- First an analysis based on only categories and not the individual treatments was conducted in order to find the categories most associated with improvement; and
- Second an analysis for each category was conducted which included the treatments in order to find the top-performing treatments.

In addition to the more commonly used support and confidence, we incorporated lift, leverage and conviction measures in our analyses in order to detect negative correlation in the rules. Additionally, we conducted a separate chi-square analysis in order to confirm the correlation as well as determine the strength of that correlation. Table 1 explains the interpretation of these measures.

Table 1. Explanation of measures

Measure	Interpretation
Support	Minimum support was varied from .1 to .01 in order to pick up less commonly used treatments in the IAN dataset.
Confidence	Confidence was set very low .1 in order to pick up as many rules as possible.
Lift	< 1 - Negative correlation = 1 - Independent > 1 - Positive correlation The higher this value, the more likely that the rule is not just a random occurrence, but because of some relationship between them.
Leverage	< 0 - Negative correlation = 0 - Independent > 0 - Positive correlation
Conviction	< 1 - Negative correlation = 1 - Independent > 1 - Positive correlation Unlike lift, this measure is not symmetric which means it has no upper bound.
Chi-Square	The number itself denotes the strength of the correlation whereas a second heading is necessary to relate the direction of the correlation.

Finally we take all the rules generated using our ensemble methods and perform a top-K analysis on them to find the strong association rules. We applied this framework on the IAN data to identify strong association rules between treatments for autism and the outcomes of the treatments with promising results.

We collected data from IAN on 3,283 children In total, there were 14,351 treatment records with corresponding efficacy ratings.

There were three attributes used for this analysis: (1) Treatment name; (2) Treatment efficacy and; (3) Taxonomy. The treatment efficacy is rated by the parent using a 9-point Likert scale with four ratings for worsening, one for no change and four for improvement. During the course of our analysis, we found that consolidating the Likert scale was advantageous to the analysis particularly to garner enough support/confidence to produce meaningful rules.

The ontology used was borrowed from the IAN website where a number of categories were presented to the users (parents of autistic children) as they would search for and enter treatment information. The categories for the taxonomy include: (1) Complementary and Alternative; (2) Educational and Behavioral; (3) Medications; (4) Physiological; (5) Special Diets and; (6) Vitamins and Supplements and (7) Top Ten.

3 Results

The top three treatments from the original dataset were not necessarily surprising since they are widely available for most autistic children in various settings (home, community, and school). The results of the statistical measures including chi-square emphatically confirm that Applied Behavioral Analysis and Speech and Language Therapy, and Occupational Therapy are indeed positively correlated with improvement.

When analyzing by ontology – three categories seem to have the strongest positive correlation with improvement – these include Educational and Behavioral, Top Ten, and Medications. Top Ten is simply a category IAN used to compile the most commonly used treatments. Intuitively, associating the Top Ten category with improvement would seem judicious since most parents would probably not use treatments in such high numbers unless some improvement was noted.

In the Complementary and Alternative segment, 5 treatments had been associated with improvement. After making a segment specific analysis, only Prayer is considered a strong rule with a lift value of 1.19, conviction of 2.55 and a very strong positive correlation. The rest of the treatments are considered weak rules with lift values below 0 (negative correlation), leverage values at 0 or below and conviction values below 1 (see Table 2).

Table 2. Complementary and Alternative

Treatment	Confidence	Lift	Leverage	Conviction	X^2	Correlation
Prayer	*0.92*	*1.19*	*0.02*	*2.55*	*268.71*	*Positive*
Weighted Blanket	*0.74*	0.96	-0.01	0.88	1.18	Negative
Chelation	*0.74*	0.96	0	0.84	0.5	Independent
Music Therapy	*0.73*	0.95	-0.01	0.99	0.5	Independent
Equine Therapy	*0.72*	0.93	-0.01	0.8	2.1	Negative

Doing a segment only analysis of medications also provided interesting results. Clonidine was confirmed as a positively correlated rule. Risperidone, on the other hand, had a lift measure of .99 and conviction of .95 which suggests a negative correlation and a chi-square value indicating independence. This is quite surprising since risperidone is one medication for autism that has shown efficacy in clinical trials [14].

The following criterion was used to include treatments in the top ‘10’:

- Confidence > .70
- Two of the three statistical measures (lift, leverage and conviction) had to indicate

positive correlation. For lift and conviction this meant for any value above 1 and for leverage any value above 0.
- A positive correlation for the chi-square test.

This analysis of the top-10 does not indicate treatment efficacy, since that can only be determined in a properly controlled clinical study. However, out of the 500 treatments present in the IAN dataset, this top-10 list indicates those treatments most like to be effectual and good candidates for clinical study.

4 Conclusion

Multi-level association rule mining is an effective tool in allowing domain knowledge to guide model development. Through the use of a treatment ontology, we were able to extract more meaningful rules from the IAN dataset. Subsequent analyses of measures such as lift, leverage and a separate chi-square test pruned the list of treatments to include only those that are positively correlated. As a result, a top-k list of treatments for autism was presented.

Without segmentation, the only treatments shown to be effective were ABA (Applied Behaviorial Analysis), Speech and Language therapy and Occupational therapy. These three therapies are one of the most common treatments for autism and therefore had the support and confidence necessary to surface as rules due to the large number of autistic children receiving them. Segmenting the dataset via the treatment ontology provided an opportunity to examine alternate treatments that are not as widely used among the autistic population. Determining absolute efficacy of these treatments is not feasible using association rule mining. The results reported only indicate treatments that are good candidates for clinical studies.

The methodology presented could be extended to other medical domains that contain well-established ontologies. The inclusion of such ontologies can serve as a powerful aid in guiding the association rule mining in generating more significant rules. The inclusion of the other measures to determine correlation are an important part of the pruning process. Medical domains where a fundamental understanding of a particular disorder is limited and a rich dataset of patient data is available would benefit from this 'theory-building' approach.

References

1. Han, J., Kamber, M.: Data Mining: Concepts and Techniques, 2nd edn. Morgan Kaufman, Amsterdam (2006)
2. Green, V.A., et al.: Internet survey of treatments used by parents of children with autism. Research in Developmental Disabilities 27(1), 70–84 (2006)
3. Parikh, M.S., Kolevzon, A., Hollander, E.: Psychopharmacology of aggression in children and adolescents with autism: a critical review of efficacy and tolerability. Journal of Child and Adolescent Psychopharmacology 18(2), 157–178 (2008)
4. Paplinski, A.P., Gustafsson, L.: An attempt in modelling early intervention in autism using neural networks. In: 2004 IEEE International Joint Conference on Neural Networks, 2002. IEEE, Piscataway (2004)

5. Freeman, R.L., Grzymala-Busse, J.W., Harvey, M.: Functional behavioral assessment using the LERS data mining system - Strategies for understanding complex physiological and behavioral patterns. Journal of Intelligent Information Systems 21(2), 173–181 (2003)
6. Booker, L.B., Goldberg, D.E., Holland, J.F.: Classifier systems and genetic algorithms. In: Carbonell, J.G. (ed.) Machine Learning Paradigms and Methods, pp. 236–282. Elsevier North-Holland, Inc., New York (1990)
7. Holland, J.H., Holyoak, K.J., Nisbett, R.E.: Induction processes of infererence, learning, and discovery. The MIT Press, Cambridge (1986)
8. Hrncir, Z., Komarek, V.: Analyses of EEG recordings. Neural Network World 14(1), 21–25 (2004)
9. Veeraraghavan, S., Srinivasan, K., Latifi, S.: Exploration of autism using expert systems. In: 4th International Conference on Information Technology New Generations. IEEE Comput. Soc., Los Alamitos (2007)
10. Petric, I., Urbancic, T., Cestnik, B.: Discovering hidden knowledge from biomedical literature. Informatica Slovenia 31(1), 15–20 (2007)
11. Urbancic, T., Cestnik, B., Macedoni-Luksic, M.: Literature mining method RaJoLink for uncovering relations between biomedical concepts. Journal of Biomedical Informatics 42(2), 219–227 (2009)
12. Urbancic, T., et al.: Literature mining: Towards better understanding of autism. In: Bellazzi, R., Abu-Hanna, A., Hunter, J. (eds.) AIME 2007. LNCS (LNAI), vol. 4594, pp. 217–226. Springer, Heidelberg (2007)
13. Schweikert, C., et al.: Analysis of autism prevalence and neurotoxins using combinatorial fusion and association rule mining. In: 9th IEEE International Conference on Bioinformatics and BioEngineering (BIBE), p. 400–404 (2009)
14. Aman, M.G., Arnold, L.E., Ramadan, Y.: Randomized, controlled, crossover trial of methylphenidate in pervasive developmental disorders with hyperactivity. Archives of General Psychiatry 62(11), 1266–1274 (2005)

Pregnancy Test for the Vision-Impaired Users

Tereza Hyková, Adam J. Sporka, Jan Vystrčil,
Martin Klíma, and Pavel Slavík

Department of Computer Graphics and Interaction
Czech Technical University in Prague, Faculty of Electrical Engineering
Karlovo nám. 13, 12135 Praha 2, Czech Republic
{hykovter,sporkaa,vystrjan,xklima,slavik}@fel.cvut.cz

Abstract. The vision-impaired people manage to cover most of their everyday needs with the help of technological equipment. Some needs, however, still require special attention: For example, commonly used pregnancy tests provide results that may be interpreted only visually. The users of these tests who are vision-impaired are dependent on a help of a sighted person which is a drastic invasion of users' privacy. This paper proposes a solution to this problem: A design of a prototype of a system allowing the vision-impaired users to interpret the results of a pregnancy test using a PC with scanner is presented and evaluated.

1 Introduction

A partial or total loss of sight severely limits the amount of information that a vision-impaired person may receive through a visual modality. This has a significant impact on a number of everyday activities, including reading printed text, navigation in the environment, or using medication.

This paper presents a project whose aim is to enable visually impaired women interpret the result of a pregnancy test in privacy, without any (direct) help of other people. A prototype of a system has been developed and subjected to a usability study. Our solution is based on the machine-performed interpretation of the visible patterns produced by commonly available pregnancy tests.

The suggested solutions were either the use of a scanner and a desktop application, or the use of a camera integrated in a mobile phone. We examined the advantages and disadvantages of both solutions.

The article begins with a short description of the background, followed by the report of the user research we performed. Then we describe the creation of the first prototype as well as its testing with the target group.

2 Background

Current state of the assistive technology allows the vision-impaired users access information by means of modality conversion [3]. Information presented in textual form is accessible by screen reader software with output via text-to-speech synthesis or on a Braille line. However, a number of activities remain in which the vision-impaired

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 335–344, 2011.

users rely on the help of the sighted people, such as navigation in unknown environment. Relying on others is not a significant problem unless some deeply personal and privacy-sensitive activity is involved, such as taking a pregnancy test.

Many types of pregnancy tests are available. They differ in various technical parameters such as their reliability or usage. They mostly use simple patterns (lines, plus or minus signs) to show the result. More advanced tests are based on digital processing and its direct interpretation (words “pregnant” or “not pregnant”) [1]. None of them, however, offers a non-visual interpretation, such as sound or haptic output.

As a 2008 April Fools’ Day joke, the ThinkGeek e-shop presented a pregnancy test with a USB connector [2], allegedly able to report the results via a personal computer. If it were true (and good practices of the accessibility design were employed), it would be an optimal solution to this problem.

3 Process of the Prototype Design

Our aim was to design a prototype of a desired application using User Centered Design (UCD) methodology [4], where the users, their understanding of the task, and their requirements are the most important aspects of the design. The user research is therefore one of the most important tools in the UCD.

Based on the user research we summarize the functions that our application should provide and describe how these functions would be used in a set of use cases and scenarios. Before an application is developed, one or several iterations of user interface prototyping are carried out. A prototype covers various aspects of how the user interacts with the application. The fidelity of the prototype is increased in each iteration. Low-fidelity prototypes are simple mock-ups, which can be easily changed or completely replaced if they do not prove to be well designed. More advanced prototypes (high fidelity prototype) are created later during the process. These can be so detailed that their look and feel could be easily confused with an actual, fully functioning application.

Before moving on to next prototype design level, each prototype needs to be tested by the users. The testing provides useful information about which parts of the prototype are well accepted by the users and which need to be changed. Using these results, the corrected prototype can be designed and the whole process is reiterated, until users’ requirements are met.

In this paper we describe a user research we performed, the scenario of use we developed, and the design and a test of the first prototype. This resulted in a corrected version of the prototype, ready for further process.

4 User Research

The purpose of the user research was to gain insights into the current situation and position of the vision-impaired users regarding their experience with the pregnancy tests as well as their experience with the computing technology. We interviewed three people with a total loss of sight. All of them lived in the greater Prague area, Czech Republic.

Interviewee A. The first interviewee is a woman in her forties. She has not used a pregnancy test her whole life until recently, when she was provided with one at the doctor's. She believes that many blind women are not even aware of their existence and deal with the consequences when it is already late. She also thinks that there is currently no solution for visually impaired people available – the only option is to visit a doctor or get a help from acquaintances who can see.

The Interviewee A did some personal research and found out that the manuals how to use the specific tests are available on the internet. She would prefer that these manuals are available through the application, including the countdown timer. She does not directly refuse the option of anonymously sending the image of the test to a random person who can see to ask for the interpretation, but does not trust the provided anonymity.

The Interviewee A is very comfortable with using a scanner and a laptop with a screen reading program. These devices are very common to her and she believes that it is an essential equipment of every visually impaired person. She controls the laptop using keyboard only. She does not think that the scanner is capable of a high enough resolution that is needed to interpret a small test stripe. She also fears that it would be difficult to keep the scanner clean when scanning the test.

A mobile phone is very common equipment to her as well – she uses a completely ordinary phone with the Mobile Speak[1] software installed. However, she does not use the integrated camera at all because she is not capable of targeting a specific object. It seems that she can hardly imagine the capabilities of a camera or guess even an approximate composition of the taken image. She is willing to try to use it for this purpose though.

The combination of a camera on a mobile phone and a mobile application would be much preferred by the Interviewee A. The main reasons are the good portability of a phone and the distrust in scanner's capabilities.

Interviewee B. The Interviewee B is a middle aged man with a total loss of sight who has no experience with pregnancy test from a blind person's view. However, he believes that a special solution would be very useful.

He is highly skilled in controlling both mobile phone and PC with scanner; he even visited a special course where visually impaired people are taught these skills. However, he refuses to give any advices himself and leaves this matter exclusively to the experts in this field. He considers these devices as an absolute must for all blind people. Both devices communicate with him using reading software; he controls the mobile phone as an every other phone and laptop solely with keyboard.

The Interviewee B absolutely refuses using a camera and considers it a completely useless device for the blind people. On the other hand he highly supports the use of the scanner and the PC, mainly because of the fact that they are so widely available and used among the target group. He states that it is much simpler to put something on a scanner desk than target a desired object with a camera.

Interviewee C. The Interviewee C is a middle aged woman. She has a husband and a son, both sighted. She however feels that she as well as the other women would

[1] http://www.codefactory.es/en/

welcome the chance to be capable of interpreting the pregnancy test on their own. So far she has been dealing with the problem using a help of another person and claims it is the only option. She does not know what types of tests there are and how they are interpreted because she never actually needed to know. She finds the countdown timer a useless tool because she will use her mobile phone for that anyway, but a small database of tests manuals seems like a good idea to her.

The idea of sending the image anonymously for interpretation to another person seems rather ridiculous to her. She claims that the community in the Czech Republic is too small to keep it anonymous and thinks some women who are trying to get pregnant and use the test often would be embarrassed.

The Interviewee C owns a PC and a scanner, knows how to use them, but does not actually use them much anymore – she has no need for it since e-books became widely available. She has little understanding what the output of a scanner is. She understands that the output is directly a text that can be read by the screen reader. She uses a keyboard and a screen reader or Braille display to interact with the PC.

She thinks that not all visually impaired people should be expected to own these devices, because they are allowed to ask for compensation devices only once per 5 years.

She uses a mobile phone often, but does not use the integrated camera. She tried taking pictures a few times but never managed to target what she wanted. She claims that some blind people use it, but they only make abstract photographs and need other person to describe what is in it. Therefore she does not think the option of taking a picture of a taken test is an acceptable idea.

5 Summary

A pregnancy test for the visually impaired women would be a very welcome tool, as there is no solution of this problem yet. We did not find any unexpected problems in the user research that would not be obvious from the task description – we can stay focused on an application for getting the interpretation of visual information.

The manipulation with a standard PC, a scanner and a screen reader is a very common thing to most of the visually impaired people, however all of the interviewees agree that they do not and probably will not use the camera or a camera-phone, mainly because of their inability to target a desired object and choose an appropriate distance.

All the interviewees control the mobile phone and PC by keys and have reading software installed. Some use a Braille display.

No other options than scanning or taking a picture with a camera was suggested by any of the interviewees and they mostly prefer scanning. The idea of sending an anonymous image for interpretation raised concerns of privacy.

6 Requirements and Recommendations

All the interviewees require only the basic functionality, i.e. a simple program that will interpret the test result for them – whether it is positive or negative. They do not refuse the other functions such as countdown timer or most common tests database and their manuals, but they hardly require them.

From the suggested options we use a scanner rather than a camera because of the technical limitations. The picture taken by visually impaired people will probably not have a high enough quality for a machine processing and therefore the application would not provide a reliable interpretation. Also, the option of sending the image anonymously for interpretation is unacceptable for various reasons given above.

7 Proposed Functionality

Based on the user research we summarized the functions that our application should provide and describe how they should be used in the following use cases and scenarios:

1. *Browse the database of tests*
   - *Find what types of tests the application supports.* Before the user wants to take a pregnancy test, she needs to know what types of tests the application supports. To do this, she opens the application database and "loops" through available test names.
   - *Find a user manual for a desired type of the test.* Once the user purchases a specific test, she needs to know how to use it. She opens the application database, selects the type she wants and has the detail open. Detail shows all the information needed for performing the test correctly.
2. *Countdown timer*
   - *Let the application count down the desired amount of time.* Once the test is taken, it is common that the user needs to wait a specific number of minutes before she can see the correct result. It is also common that the results should not be interpreted after a given number of minutes.
     1. The user selects the countdown time option, chooses the number of minutes for the first and optionally the number of minutes for the second alarm. Once the alarm rings, she can stop it.
     2. Another option to use the countdown time feature is to start it right from a test detail screen. In this case, the numbers of minutes is automatically set according to the user manual in the database.
3. *Interpret a test*
   - *Open a scanned test in the application.* The user scans the test in another scanning program she is used to and saves it as an image. Then, in our application, she selects the option to open a file and chooses this image. The application will report whether the image was successfully open.
   - *Scan an image right inside the application.* The user selects the option to scan an image. The settings will be automatically set to common values with highest resolution possible. The user puts the test in the scanner and selects the start scanning option. The application will report whether the image was successfully scanned.

- *Define used test type.* The user needs to define the type of test she used, so that the application knows how to interpret the image data. She selects it from the database the way already described.
- *Interpret a test.* With image and test type defined, the user can order the interpretation to start. If the processing is successful, the application announces the result of the test (positive or negative) and the probability of correct answer (for example in the first month of pregnancy, only a very light stripe can be seen, which indicates a positive result). If it is unsuccessful, an error and its cause are reported. This can happen for example if the defined test type is not found in the image or if the check indicator (e. g. stripe that should always appear when taking the test) does not appear.

7.1 Prototype Design and Study

Since we are creating an application for people with significant vision impairments, we do not design a visual user interface. A Wizard-of-Oz protocol was used in this study: The experimenter observes the user's input and simulates the corresponding response.

The behavior of the application can be described by a state machine, where each node describes a certain state and has a label. Each edge is named according to what key it is connected and specifies the state that is reached upon pressing this key.

Although the state machine is what the application actually works like, it is not very comfortable and obvious to work with for the tester. The users also need to create a more suitable imagination about the application structure, which will be easier to comprehend and remember.

We therefore decided to visualize and design this state machine as a hierarchy of menus. Key UP and DOWN move between the menu items, ENTER confirms the selected choice, ESC or BACKSPACE work the same as selecting the "back" choice.

7.2 Testing

The user will not interact directly with this prototype. The user will use the controls that actually have no effect and the tester will simulate the application behavior by moving his position in the prototype and reading the labels for the current state. The actual scanning or text file browsing were skipped, since it is not in the scope of the low fidelity prototype.

The prototype has been tested with three female participants (30—50 years old) from Prague with total loss of sight. We concentrated on a single scenario that covered all the key points of its functionality. The task for all the participants was to:

1. See what types of tests the application supports and select one,
2. Have the application countdown desired amount of time,
3. Insert an image of the performed test,
4. Interpret the test and get its result.

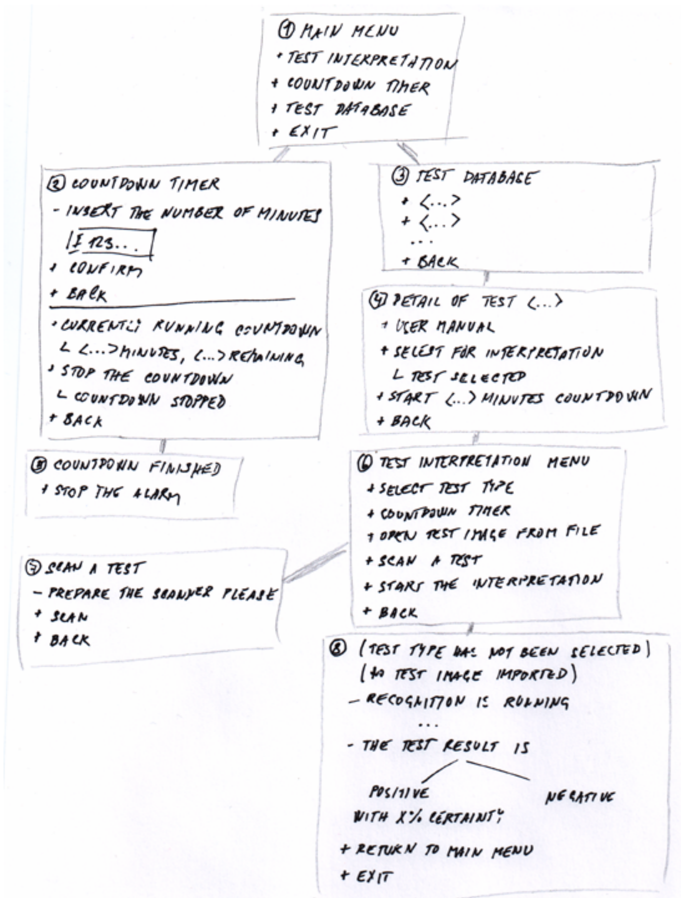


Fig. 1. Initial version of the prototype

User D. The testing took approximately an hour and was performed in a quiet gallery in a conservatory, where the User D spends much of her time and feels comfortable.

The testing was without problems until the user reached screen 4 (see Figure 1), where she could not identify the purpose of the button *Select this test type* and where she currently was in the application after selecting this. She did not remember the *Interpret test* option in the main menu and did not know how to continue. In this menu she also did not know what she will find exactly in the *User manual* option and thought that a more specific label should be used. She also wanted the application to start reading the manual after pressing the Enter key and not right upon the selection of this option.

She was confused by what happened when she pressed the Escape or Backspace keys, compared to the Back button – she got disoriented when she was back at the menu option where she came from instead of at the start of the menu.

The user did not understand the *Open image from file* option. After an explanation, she said, that she does know how to scan and save an image as an image file, because she only uses program which directly recognizes the text.

She suggested some change of labels to use words that are more commonly used by text to speech programs etc. She also suggested an option of finding an option in the menu by typing first letters of its name.

User E. The test with the second user took approximately 30 minutes and was performed at the user's home.

The User E was confused at some points when she did not know how to continue, and expected the application to guide her through by telling what keys to press for what option. She got disoriented by deeply nested submenus.

She suggested that the user manual should contain information about how to recognize what side of the pregnancy test she should put on the scanner. She also thought it could contain some hints how to scan the test without getting the scanner dirty and proposed for example using transparent foil.

She was also concerned about how the navigation in the longer text would work – whether she should use the key shortcuts she was used to from other applications or the application would have its own. (She is used to working with multiple text-to-speech programs because she also teaches other blind people to use them. She mentioned that each product utilizes slightly different sets of keyboard shortcuts).

The User E was happy about the fact that the application would also announce how clearly the result was visible, but she did not understand what the label *With certainty* means and suggested a more understandable label.

User F. The third test took about 20 minutes and was performed at the user's home**.**

The User F also suggested smaller depth of menu hierarchy. She was concerned about whether the application checks the stripe that indicates that the test was taken correctly and the result was correct.

After getting the test result, she wanted to save it, because she thought some women who would try to get pregnant and perform the test often would want to keep track about the dates and the results. She also said that she would like to have the images saved so that if she would not be sure about the correctness of the interpreted result by the application, she could show them to another person to check again.

7.3 Test Results and Recommendations

The depth of the menu hierarchy should be as small as possible and the application should guide the user in one way rather than leaving more options at one time open.

The application should give the user as much feedback as possible, so that she knows what is going on at all points. For example the countdown timer should give feedback every half minute, the interpreter of the results should announce whether the correct test indicator was found, and the application should announce what keys to press to continue at points where it is not completely obvious (e. g. "To go back to main menu, press enter, to exit the application, press Escape").

The revised structure of the application is shown in Figure 2. We will leave both options *Scan the test* and *Open test image from file* available. The Users E and F knew

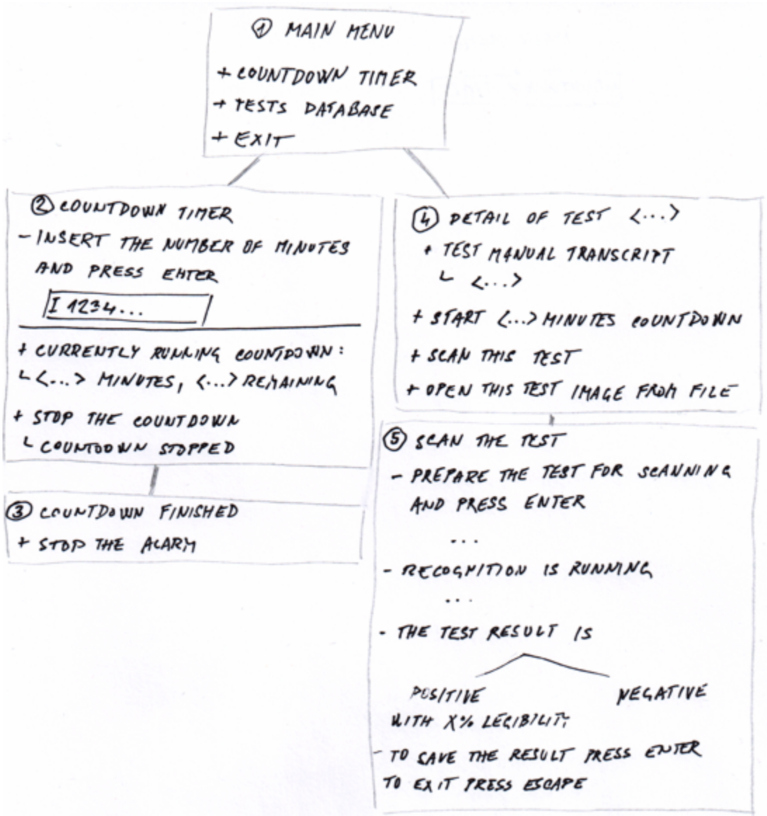


Fig. 2. Corrected version of the prototype, based on the results of the user test

how to save scanned image as an image file and thought they might need it someday. An option of saving the result will be added and tested in the next iterations.

Some labels were changed as suggested by all three users. They are used to some terminology used by many text-to-speech programs and know perfectly their meaning without having to think about it.

The control keys should also include TAB (with the same effect as the key DOWN) and typing of the first letters of a word should directly lead to a desired option. The *Back* option was removed because all the users were used to pressing Esc or Backspace keys and one extra item in menu delayed their navigation.

The user manual should provide more information specific to a visually impaired user than the one provided with the test, such as how to recognize what side of the test to put in the scanner.

The application should not repeat what menu the user just entered, because she already knows that from pressing the button. The first option should be selected after opening a new menu.

8 Conclusion

The paper aimed at informing the designers of a future system allowing the vision-impaired users to interpret the results of the commonly available pregnancy tests. The user research as well as the low-fidelity prototype testing provided useful insights into the given topic. The most important finding was that the need for a solution of the problem of pregnancy test interpretation by the vision-impaired users is real, since the currently available means of solving it reduce the privacy of the vision-impaired.

We found that we could achieve a solution by combination of common pregnancy tests and a scanner, a device commonly available and widely used among the target group. The software needed is a simple application that could interpret an image of a taken test. In this paper a prototype of a user interface of such an application has been described and as well as the related user research and testing.

Acknowledgments. This research has been partially supported by the MSMT under the research programs MSM 6840770014 and LC-06008 (Center for Computer Graphics). The topic for this project was suggested by a member of the target group.

References

1. Clearblue. Clearblue Easy Digital Pregnancy Test, `http://www.clearblueeasy.com/pdfs/clearblue-easy-digital-pregnancy-test-pdf.php` (visiterd February 2011)
2. Dix, A., Finlay, J., Abowd, G.D., Beale, R.: Human-Computer Interaction, 3rd edn. Prentice-Hall, Englewood Cliffs (2004)
3. Thang, T.C., Jung, Y.J., Ro, Y.M.: Effective adaptation of multimedia documents with modality conversion. Signal Processing: Image Communication 20(5), 413–434 (2005)
4. ThinkGeek. PTeq USB Pregnancy Test, `http://www.thinkgeek.com/stuff/41/pteq.html` (visited February 2011)

Effect of Spinal Cord Injury on Nonlinear Complexity of Skin Blood Flow Oscillations

Yih-Kuen Jan[1], Fuyuan Liao[1], and Stephanie Burns[2]

[1] University of Oklahoma Health Sciences Center, Department of Rehabilitation Sciences, Oklahoma City, Oklahoma, USA
yjan@ouhsc.edu
[2] Oklahoma City Veterans Affairs Medical Center, Department of Neurology and Rehabilitation, Oklahoma City, Oklahoma, USA

Abstract. This study investigated the effect of spinal cord injury (SCI) on nonlinear complexity of skin blood flow oscillations (BFO). Complexity of the characteristic frequencies embedded in BFO was described by the scaling coefficient derived by detrended fluctuation analysis (DFA) and the range of scaling coefficients derived from multifractal detrended fluctuation analysis (MDFA) in specific scale intervals. 23 subjects were recruited into this study, including 11 people with SCI and 12 healthy controls. Local heating-induced maximal sacral skin blood flow was measured by laser Doppler flowmetry. The results showed that metabolic BFO (0.0095-0.02 Hz) exhibited significantly lower complexity in people with SCI as compared with healthy controls ($p<0.01$) during maximal vasodilation. This study demonstrated that complexity analysis of BFO can provide information of blood flow dynamics beyond traditional spectral analysis.

Keywords: blood flow oscillations, complexity, detrended fluctuation analysis, multifractal detrended fluctuation analysis, spinal cord injury.

1 Introduction

Spinal cord injury (SCI) results in an interruption of the autonomic pathways from the brainstem and hypothalamus to the intermediolateral cell column of the spinal cord [1]. This interruption causes a loss or attenuation of modulation of the spinal autonomic reflexes in response to various stimuli at the level below spinal injury. Inability to increase skin blood flow to reduce tissue ischemia following SCI may increase risk for pressure ulcers [2].

Skin blood flow has been studied using two categories of methods: linear and nonlinear analysis. The linear analysis typically starts with the mean and standard deviation, focusing on Fourier spectral analysis or wavelet analysis. Spectral analysis of laser Doppler flowmetry (LDF) signals has revealed five characteristic frequencies [3, 4]. These characteristic frequencies have been associated with heart beats (0.4-2.0 Hz), respiration (0.15-0.4 Hz), myogenic activity (0.05-0.15 Hz), neurogenic

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 345–355, 2011.

(0.02-0.05 Hz), and metabolic (0.0095-0.02 Hz) activities, respectively [5]. Generally, the time averaged amplitude or power of the characteristic frequencies are used to evaluate the activities of the underlying mechanisms. However, these characteristic frequencies are not constant, but vary with time [4] (Fig. 1). This feature cannot be characterized by time averaged amplitude or power.

The time-varying feature of the characteristic frequencies of BFO may be described by their structural complexity [6]. This complexity arises from the interaction of structural units and regulatory feedback loops that operate over a wide range of temporal and spatial scales [6]. Research suggests that the output of healthy systems exhibits a type of complex variability, and that such complexity degrades with aging and diseases, reducing the adaptive capabilities of the individual [7-9]. Although the concept of complexity has been elaborated for several decades, the use of complexity for diagnostic and prognostic assessment remains a great challenge. Investigators have applied complexity analysis to evaluate variability in physiological signals, e.g. heart rate [10, 11], gait [12], and BFO [13, 14]. These studies provided evidences of clinical importance and potential use to indentify pathological conditions.

To better understand the role of microvascular dysfunction in pressure ulcers, we have conducted a series of studies examining BFO in response to various causative factors of pressure ulcers [14-18]. In this study, we examined the effect of spinal cord injury on microvascular function by quantifying complexity of BFO in response to local heat.

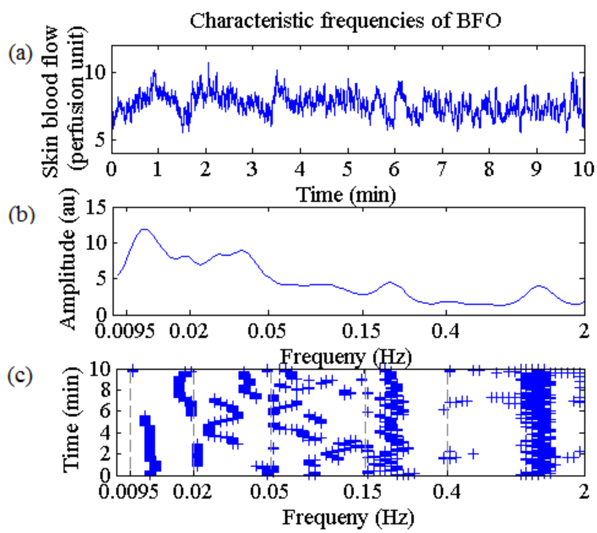


Fig. 1. Characteristic frequencies embedded in BFO. (a) A skin blood flow signal recorded from a healthy subject. (b) The time-averaged amplitudes of the continuous wavelet transform (WT) of the signal. (c) The local maxima of the WT.

2 Methods

2.1 Participants

We recruited 23 subjects into this study, including 11 people with SCI and 12 healthy controls. All subjects with SCI were wheelchair users, with the injury level between C4 and T12, and were at least 6 months post spinal injury. None of the subjects had cardiorespiratory disease, hypertension, or other pathological conditions, and none were taking medications that might affect cardiovascular function or diabetes mellitus. All subjects gave informed consent to this study approved by the University Institutional Review Board. The demographic features of the enrolled subjects are shown in Table 1.

Table 1. Demographic data of the research subjects

	SCI	Healthy controls
Number of subjects	11	12
Gender, M/F	7/4	4/8
Age (years)	35.4(12.4)	25.3 (5.4)
Duration of spinal injury (years)	7.6 (4.9)	-
Body mass index (kg/m^2)	25.1 (4.4)	23.2 (2.4)

Data are expressed as the mean (standard deviation).

2.2 Data Acquisition

After at least a 30 min quiet rest period to become acclimated to the room temperature (24±2°C), the subject was positioned in a prone posture. Sacral skin blood flow was recorded using the laser Doppler flowmetry (PF 5001, Perimed AB, Sweden) at a sampling frequency of 32 Hz. A heating probe (Probe 415-242, Perimed AB) was used to heat the skin to 42°C in 2 minutes and to maintain that temperature for 50 min. The protocol included a 10 min pre-heating period, a 50 min heating period, and a 10 min post-heating period [21].

2.3 Complexity Analysis

*Detrended fluctuation analysis (DFA).*The DFA was introduced by Peng et al. [19, 22] to quantify the long-range power law correlations of nonstationary time series. A key issue of the DFA is that the fluctuations driven by extrinsic uncorrelated stimuli are interpreted as a systematic "drift" or "trend," and are treated by removing the least-squares regression in each observation window [23]. To illustrate the DFA algorithm, consider the blood flow signal shown in Figure 1(a). First, the original time series, $\{x(i), i = 1, \dots, N\}$, is integrated after subtracting its mean value

$$y(k) = \sum_{i=1}^{k}(i) - \mu),\ k = 1, \dots, N \tag{1}$$

with μ as the mean of $\{x(i)\}$. The integrated time series is then divided into boxes of equal length, n (Fig. 2(a)). In each box, a least-squares line $y_n(k)$ is fitted to the integrated series, $\{y(k)\}$. Next, $\{y(k)\}$ is detrended by subtracting the local trend, $y_n(k)$. The root-mean-square fluctuation, $F(n)$, is calculated for all box sizes, n:

$$F(n) = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(y(k) - y_n(k))^2} \tag{2}$$

Finally, $F(n)$ is plotted against n in a log-log scale, and the slope of the regression line gives the scaling exponent, α (Fig. 2(b)). This exponent can be considered as a measure of the "roughness" of the original time series: the larger the value of α, the smoother the time series [8].

The scaling exponent derived by DFA is not always a constant [24], i.e. the log-log plot may exhibit more than one scaling region. A crossover can result from a change in the correlation properties of the signal at different time or space scales [13, 14]. However, a crossover may also arise from the nonstationarities of the signal and/or its trends without any transition in the intrinsic dynamics [25].

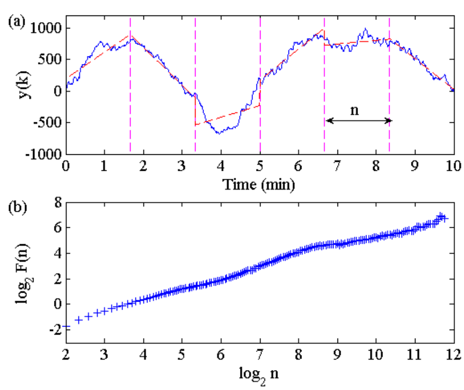


Fig. 2. Illustration of the DFA algorithm. The original time series is the blood flow signal shown in Fig. 1(a). (a) The solid curve is the integrated time series, $y(k)$. The vertical dash lines indicate boxes of size, n. The dashed line segments represent the trend estimated in each box by a linear least squared fit. (b) The deviations, $F(n)$, versus the box size, n, on a log-log scale. There are more than one scaling region.

Multifractal analysis. The DFA algorithm gives only one exponent to characterize a signal. This method, therefore, is most appropriate for the analysis of monofractal signals [8]. Monofractals have the same scaling properties throughout the entire signal. In contrast, for multifractal signals, different parts may have different scaling properties, requiring a large number of indices to characterize their scaling properties. It has been found that heart rate fluctuations exhibit multifractality in healthy human subjects [10, 26]. Moreover, heart rate recorded from patients with heart failure shows a breakdown of multifractality [10]. Recently, a few studies [27, 28] suggested that skin blood flow signals recorded by LDF also have multifractal properties.

The most popular methods for the detection of multifractal scaling properties of a signal are: the wavelet transform modulus maxima (WTMM) [29, 30] and the multifractal detrended fluctuation analysis (MDFA) [20]. It was suggested that in the case of lacking a priori knowledge of the properties of a process, the MDFA should be recommended [31]. MDFA is a generalization of the standard DFA [20], describing how the q-th order moments of fluctuation depend on the observation scale. For a time series, $\{x(i), i = 1, \ldots, N\}$, defined on a compact support, first, it is integrated after subtracting its mean value. Then the integrated series is divided into m non-overlapping segments of length n and the same procedure is repeated starting from the opposite end ($2m$ segments total). Next, for each segment v, the local trend is estimated by fitting a polynomial and subtracting it from the segment. Let the variance be $F^2(v,n)$, a q-th order fluctuation function is defined as

$$F_q(n) = \begin{cases} \{\frac{1}{2m}\sum_{v=1}^{2m}[F^2(v,n)]^{q/2}\}^{1/q}, & q \neq 0 \\ \exp\{\frac{1}{4m}\sum_{v=1}^{2m}\ln[F^2(v,n)]\}, & q = 0 \end{cases} \quad (3)$$

If the time series is long-range power-law correlated, it is expected that

$$F_q(n) \sim n^{h(q)} \quad (4)$$

Thus one obtains a family of exponents, $h(q)$. For $q = 2$, the standard DFA is retrieved.

The reason for this procedure is that, for positive values of q, $h(q)$ reflects the scaling behavior of the segments with large fluctuations, whereas for negative values of q, $h(q)$ reflects the scaling behavior of the segments with small fluctuations [20]. Thus, for a given signal, if the scaling behavior is identical for all segments, $h(q)$ is independent of q; if small and large fluctuations scale differently, there will be a significant dependence of $h(q)$ on q. For instance, in the scale range of n=1600~3368 (given the sampling frequency 32 Hz, this range is related to the frequency band 0.0095~0.02 Hz, see the following section), $h(q)$ decreases with q, indicating the existence of multifractal scaling behavior of BFO (Fig. 3).

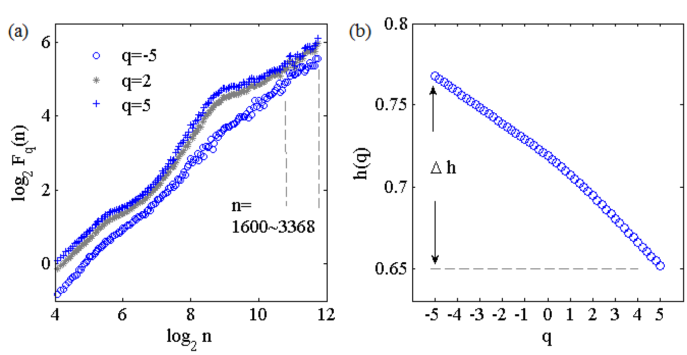


Fig. 3. Illustration of the multifractal detrended fluctuation analysis (MDFA) method. (a) The q-th order fluctuations versus observation scales on a log-log scale, showing different scaling behaviors for different values of q (-5, 2, 5). (b) Scaling exponent in the scale range of n=1600~3368 decreases with q.

2.4 Relationship between Scale (Box Size) and Frequency in DFA (MDFA)

Previous studies have demonstrated that DFA is closely related to power spectral density analysis [32, 33]. Suppose that the power spectral density (PSD) of the original time series, $\{x(i)\}$, is $P_x(\omega)$. The PSD of the integrated series (Eq. (1)), y, is then given by [32]

$$P_y(\omega) = \begin{cases} 0, & \omega = 0 \\ \frac{P_x(\omega)}{2(1-cos\omega)}, & \omega \neq 0 \end{cases} \tag{5}$$

The integrated series, y, is divided into non-overlapping boxes of length n and detrended. It can be numerically verified that the PSD of the detrended series obeys the following relation (Fig. 4)

$$P_y^d \approx \begin{cases} 0, & 0 \leq \omega \leq \omega_s/n \\ P_y(\omega), & otherwise \end{cases} \tag{6}$$

where $\omega_s = 2\pi f_s$ with f_s being the sampling frequency in Hertz. This relation means that detrending in boxes of length n is equivalent to applying a high-pass filter to the integrated series with a cut-off frequency f_s/n. Thus, the scaling behavior of metabolic, neurogenic, and myogenic BFO are respectively characterized by the scaling coefficients in the scale ranges: $f_s/0.02 < s < f_s/0.0095$, $f_s/0.05 < s < f_s/0.02$, and $f_s/0.15 < s < f_s/0.05$.

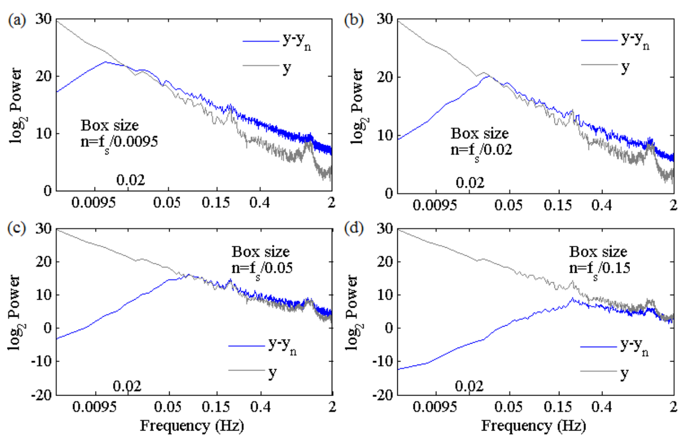


Fig. 4. Relationship between observation scales (box size) and frequencies in DFA and MDFA methods. In all the panels, y denotes the integrated series, $y - y_n$ denotes the detrended series, and n denotes the scale (box size). In the fitting procedure, quadratic polynomials were used. Detrending in boxes of length n is equivalent to applying a high-pass filter to the integrated series with a cut-off frequency f_s/n.

We used $\Delta h = \max(h(q)) - \min(h(q))$ in the scale interval $f_s/0.02 < s < f_s/0.0095$ to quantify the complexity degree of metabolic BFO, the scale interval

$f_s/0.05 < s < f_s/0.02$ to quantify the complexity degree of neurogenic BFO, and scale interval $f_s/0.15 < s < f_s/0.05$ to quantify the complexity degree of myogenic BFO. Comparisons between two groups in the same thermal condition were performed using the Wilcoxon rank sum tests. Comparisons between two thermal conditions in the same group were performed using the Wilcoxon signed rank tests. The significance level was set at p value less than 0.05.

3 Results and Discussion

Fig. 5 shows the DFA coefficient and complexity degree of metabolic oscillations. At baseline, the DFA coefficient showed significantly higher values in healthy controls (p<0.05). However, it did not show significant change either in healthy controls or in people with SCI due to local heating. Unlike DFA coefficient, the complexity degree of metabolic oscillations significantly increased in healthy controls in response to local heating (p <0.01), while not in people with SCI. During the maximal vasodilation, a significant lower degree of complexity was observed in people with SCI (p <0.01). For neurogenic and myogenic BFO, neither DFA coefficient nor complexity degree showed significant differences between healthy controls and people with SCI.

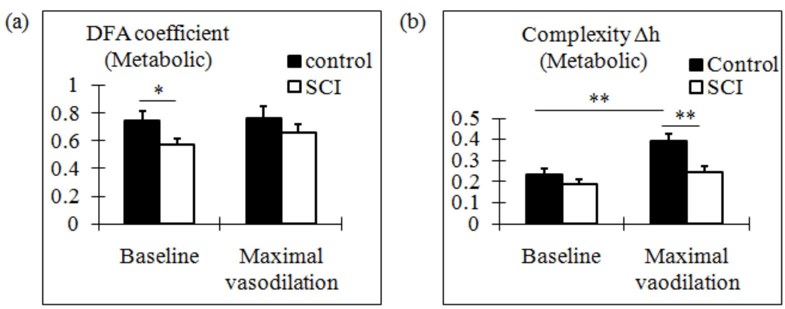


Fig. 5. Comparisons of DFA coefficient and complexity degree of metabolic oscillations between healthy controls and people with SCI. Values are expressed as means ± SE. * indicates p<0.05; ** indicates p<0.01.

We have shown that people with SCI had a lower complex behavior of metabolic oscillations. From our results, we may deduce that metabolic oscillations in healthy people have richer structures than in people with SCI. Our results support the concept that decreased complexity of a physiologic process is associated with pathological states [6, 7]. Previous studies have demonstrated that attenuated metabolic and neurogenic oscillations are associated with vasodilatory dysfunction [18]. Li et al. [34] observed a lower amplitude of metabolic frequency in people with SCI as compared with healthy controls. However, the dynamics of skin blood flow or of a characteristic frequency contain not only the amplitude, but also the variations in frequency and amplitude. The latter cannot be characterized by time-averaged amplitude or power. Although wavelet based time-frequency-amplitude presentation can describe how frequency and amplitude change with time, difficulties may arise from quantifying the changes [18].

Fig. 6 compares the dynamics of metabolic oscillations in a healthy subject with those in a subject with SCI. The blood flow signal shown in Fig. 1(a) and a blood flow signal recorded from a subject with SCI have almost the same time-averaged amplitude of WT in the frequency band, 0.0095-0.02 Hz. However, by comparing the time-frequency-amplitude presentations (Fig.6 (a) and (b)), it can be seen that the metabolic frequency in the healthy subject appears to have more complex behavior. This visual intuition can be verified by the MDFA. As shown in Fig. 6 (c) and (d), the metabolic frequency in the healthy subject has a wider range of scaling exponents, i.e., more complex dynamics.

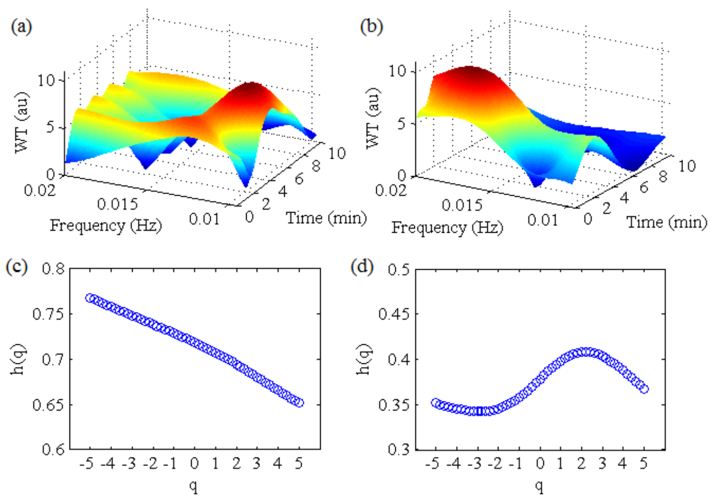


Fig. 6. Comparisons of the dynamics of metabolic frequency between a healthy subject and a subject with SCI. (a) Continuous wavelet transform (WT) of a blood flow signal in health (0.0095-0.02 Hz). (b) WT of a blood flow signal in SCI (0.0095-0.02 Hz). (c) MDFA of the blood flow signal in health (0.0095-0.02 Hz). (d) MDFA of the blood flow signal in SCI (0.0095-0.02 Hz).

4 Conclusion

Our results provided evidence that people with SCI have altered dynamics of skin blood flow: metabolic frequency of BFO in people with SCI exhibits less complexity compared to that in healthy people. Complexity analysis of BFO can provide information on blood flow dynamics beyond linear analysis, which might be used to study the interactions among blood flow control mechanisms in various pathological conditions.

Acknowledgements. This work was supported by the National Institutes of Health (R03HD060751) and the Christopher and Dana Reeve Foundation (JA2-0701-2).

References

1. Alexander, M.S., Biering-Sorensen, F., Bodner, D., Brackett, N.L., Cardenas, D., Charlifue, S., et al.: International standards to document remaining autonomic function after spinal cord injury. Spinal Cord 47(1), 36–43 (2009)
2. Nixon, J., Cranny, G., Bond, S.: Pathology, diagnosis, and classification of pressure ulcers: comparing clinical and imaging techniques. Wound Repair Regen 13(4), 365–372 (2005)
3. Stefanovska, A., Bracic, M., Kvernmo, H.D.: Wavelet analysis of oscillations in the peripheral blood circulation measured by laser Doppler technique. IEEE Trans. Biomed. Eng. 46(10), 1230–1239 (1999)
4. Bracic, M., Stefanovska, A.: Wavelet-based analysis of human blood-flow dynamics. Bull. Math. Biol. 60(5), 919–935 (1998)
5. Stefanovska, A., Bracic, M.: Physics of the human cardiovascular system. Contemporary Physics 40(1), 31–55 (1999)
6. Goldberger, A.L., Peng, C.K., Lipsitz, L.A.: What is physiologic complexity and how does it change with aging and disease? Neurobiology of Aging 23(1), 23–26 (2002)
7. Lipsitz, L.A., Goldberger, A.L.: Loss of Complexity and Aging - Potential Applications of Fractals and Chaos Theory to Senescence. Jama-Journal of the American Medical Association 267(13), 1806–1809 (1992)
8. Goldberger, A.L., Amaral, L.A.N., Hausdorff, J.M., Ivanov, P.C., Peng, C.K., Stanley, H.E.: Fractal dynamics in physiology: Alterations with disease and aging. Proceedings of the National Academy of Sciences of the United States of America 99, 2466–2472 (2002)
9. Goldberger, A.: Complexity loss, aging, and disease: Is there a dynamical "Theory of Everything Pathologic?". Journal of Critical Care 25(3), E2 (2010)
10. Ivanov, P.C., Amaral, L.A.N., Goldberger, A.L., Havlin, S., Rosenblum, M.G., Struzik, Z.R., et al.: Multifractality in human heartbeat dynamics. Nature 399(6735), 461–465 (1999)
11. Merati, G., Di Rienzo, M., Parati, G., Veicsteinas, A., Castiglioni, P.: Assessment of the autonomic control of heart rate variability in healthy and spinal-cord injured subjects: contribution of different complexity-based estimators. IEEE Trans. Biomed. Eng. 53(1), 43–52 (2006)
12. Harbourne, R.T., Stergiou, N.: Movement Variability and the Use of Nonlinear Tools: Principles to Guide Physical Therapist Practice Response. Physical Therapy 89(3), 284–285 (2009)
13. Esen, F., Esen, H.: Detrended fluctuation analysis of laser Doppler flowmetry time series: the effect of extrinsic and intrinsic factors on the fractal scaling of microvascular blood flow. Physiological Measurement 27(11), 1241–1253 (2006)
14. Liao, F., Garrison, D.W., Jan, Y.K.: Relationship between nonlinear properties of sacral skin blood flow oscillations and vasodilatory function in people at risk for pressure ulcers. Microvasc Res. 80(1), 44–53 (2010)
15. Geyer, M.J., Jan, Y.K., Brienza, D.M., Boninger, M.L.: Using wavelet analysis to characterize the thermoregulatory mechanisms of sacral skin blood flow. Journal of Rehabilitation Research and Development 41(6A), 797–805 (2004)

16. Jan, Y.K., Brienza, D.M., Geyer, M.J.: Analysis of week-to-week variability in skin blood flow measurements using wavelet transforms. Clin. Physiol. Funct. Imaging 25(5), 253–262 (2005)
17. Jan, Y.K., Brienza, D.M., Geyer, M.J., Karg, P.: Wavelet-based spectrum analysis of sacral skin blood flow response to alternating pressure. Arch. Phys. Med. Rehabil. 89(1), 137–145 (2008)
18. Jan, Y.K., Struck, B.D., Foreman, R.D., Robinson, C.: Wavelet analysis of sacral skin blood flow oscillations to assess soft tissue viability in older adults. Microvasc. Res. 78(2), 162–168 (2009)
19. Peng, C.K., Buldyrev, S.V., Havlin, S., Simons, M., Stanley, H.E., Goldberger, A.L.: Mosaic Organization of DNA Nucleotides. Physical Review E 49(2), 1685–1689 (1994)
20. Kantelhardt, J.W., Zschiegner, S.A., Koscielny-Bunde, E., Havlin, S., Bunde, A., Stanley, H.E.: Multifractal detrended fluctuation analysis of nonstationary time series. Physica a-Statistical Mechanics and Its Applications 316(1-4), 87–114 (2002)
21. Minson, C.T., Berry, L.T., Joyner, M.J.: Nitric oxide and neurally mediated regulation of skin blood flow during local heating. Journal of Applied Physiology 91(4), 1619–1626 (2001)
22. Peng, C.K., Havlin, S., Hausdorff, J.M., Mietus, J.E., Stanley, H.E., Goldberger, A.L.: Fractal mechanisms and heart rate dynamics - Long-range correlations and their breakdown with disease. Journal of Electrocardiology 28, 59–65 (1995)
23. Meyer, M., Stiedl, O.: Self-affine fractal variability of human heartbeat interval dynamics in health and disease. Eur. J. Appl. Physiol. 90(3-4), 305–316 (2003)
24. Chen, Z., Ivanov, P., Hu, K., Stanley, H.E.: Effect of nonstationarities on detrended fluctuation analysis. Phys. Rev. E Stat. Nonlin. Soft. Matter Phys. 65(4 Pt 1), 041107 (2002)
25. Hu, K., Ivanov, P.C., Chen, Z., Carpena, P., Stanley, H.E.: Effect of trends on detrended fluctuation analysis. Physical Review E 6401(1) (2001)
26. Havlin, S., Amaral, L.A., Ashkenazy, Y., Goldberger, A.L., Ivanov, P., Peng, C.K., et al.: Application of statistical physics to heartbeat diagnosis. Physica a-Statistical Mechanics and Its Applications 274(1-2), 99–110 (1999)
27. Humeau, A., Chapeau-Blondeau, F., Rousseau, D., Rousseau, P., Trzepizur, W., Abraham, P.: Multifractality, sample entropy, and wavelet analyses for age-related changes in the peripheral cardiovascular system: Preliminary results. Medical Physics 35(2), 717–723 (2008)
28. Humeau, A., Chapeau-Blondeau, F., Rousseau, D., Tartas, M., Fromy, B., Abraham, P.: Multifractality in the peripheral cardiovascular system from pointwise holder exponents of laser Doppler flowmetry signals. Biophysical Journal 93(12), L59–L61 (2007)
29. Arneodo, A., Bacry, E., Muzy, J.F.: The Thermodynamics of Fractals Revisited with Wavelets. Physica a-Statistical Mechanics and Its Applications 213(1-2), 232–275 (1995)
30. Bacry, E., Muzy, J.F., Arneodo, A.: Singularity Spectrum of Fractal Signals from Wavelet Analysis - Exact Results. Journal of Statistical Physics 70(3-4), 635–674 (1993)
31. Oswiecimka, P., Kwapien, J., Drozdz, S.: Wavelet versus detrended fluctuation analysis of multifractal structures. Phys. Rev. E Stat. Nonlin. Soft. Matter Phys. 74(1 Pt 2), 016103 (2006)
32. Heneghan, C., McDarby, G.: Establishing the relation between detrended fluctuation analysis and power spectral density analysis for stochastic processes. Physical Review E 62(5), 6103–6110 (2000)

33. Willson, K., Francis, D.P., Wensel, R., Coats, A.J.S., Parker, K.H.: Relationship between detrended fluctuation analysis and spectral analysis of heart-rate variability. Physiological Measurement 23(2), 385–401 (2002)
34. Li, Z., Leung, J.Y., Tam, E.W., Mak, A.F.: Wavelet analysis of skin blood oscillations in persons with spinal cord injury and able-bodied subjects. Arch. Phys. Med. Rehabil. 87(9), 1207–1212 (2006), quiz 1287

Developing Protégé to Structure Medical Report

Josette Jones, Kanitha Phalakornkule, Tia Fitzpatrick,
Sudha Iyer, and C. Zorina Ombac

Indiana University-Purdue University Indianapolis, School of Informatics,
535 West Michigan Street, Indianapolis, Indiana, 46202, USA
jofjones@iupui.edu

Abstract. The paper presents a creation of an Ontological framework for the capture, integration and presentation of clinical information embedded in medical reports. The project used a hybrid method of top-down and bottom-up method started by abstracting the concept healthcare event resulting in reusable knowledge structure. The tools employed in the project are Protégé 3.4, add-ins, and MySQL, which were used to structure annotated radiology reports with the focus on the reports pertaining to the heart. Furthermore, the project demonstrated how the relational database was developed from an existing ontology and how its database schema was imported to the ontology framework through Protégé's DataMaster Plug-in. Then, we presented how to query the knowledge taxonomy in Protégé.

1 Introduction

Knowledge is what we learn and believe about an entity in a domain. A knowledge representation is a surrogate for something that can be seen and/or imagined. Or, it can be referred to as a modeling of a widespread belief about something [23]. It allows people to share, think, organize and reason about the thing rather than acting on the actual thing [5]. Domain knowledge is formally specified through definitions of classes, relations, functions, and other objects and is represented in ontology [9].

Ontology is a philosophical term adapted for use in modern times in the field of computer science. Originally, ontology is the study of what exists in the universe or a description of that which we know to exist as presented in theories (concepts) and mathematical formulations. Ontology frameworks have been accepted more and longer in Web and biology applications. It is a key of the solution in semantic heterogeneity and in reusing domain knowledge [12]. Its well-known contribution is in text-mining across different data types and knowledge domains and foremost in W3-applications. Ontology in healthcare is also mainly focused on the representation and organization of health-care terminologies and a few others, but not in database and system design. For instance, the ontology or terminology servers like UMLS, SNOMED, and LOINC provide common structures and consistent representation of clinical information for standardized vocabularies and communication [1]. This concepts can be applied to knowledge from medical reports and protocols will be modeled this knowledge in a specific-purpose knowledge representation language, and finally formalized this knowledge representation in terms of temporal logic and

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 356–365, 2011.

parallel programs. The value of this formalization process, called Ontology, is that each successive formalization step has contributed to improving the quality of the original medical protocol and report. Therefore, Ontology has successfully been used for integration between heterogeneous databases across various unrelated systems and organizations [9]; healthcare application is emerging. Three concepts built-in the ontology in healthcare system are belonging to clinical conditions (diagnosis), disease treatment pattern (procedures) and patients' preferences. Another benefit of Ontology can be found in document searching and reporting.

In the past, much work in healthcare ontology has been focused on the representation and organization of health-care terminologies and related biomedical concepts but not so much so in logical models for database design and/or reusable knowledge structures. Furthermore, biomedical ontology development for UMLS terminologies (e.g., SNOMED-CT, LOINC) is derived from knowledge structures [1], [21] and is not built from clinical documents and protocols.

In this study, we propose a medical radiology ontology based on description logic that is easy to use and understand for both clinicians and system developers and which has the potential for expansion beyond the specialty of radiology for the purposes of understanding and communicating medical information and of modeling an application. Then we proved that the ontology framework can be connected to a rational database using Protégé

1.1 Background

The clinical documents are heterogeneous and recorded in unstructured free texts, known as natural language [23]. It is the most expressive and powerful knowledge representation method available due to its ability to express subtle shades of meaning and to reason [21]. Knowledge representation is presenting or representing symbolically that which exists in reality that can be seen and/or imagined. On the other hand, natural language attempts to do so in the form of words and sentences. However, natural language's strengths of expressing subtle shades of meaning and of reasoning could also be considered a weakness in that inherent ambiguity exists and a computable data structure is absent.

Since a significant portion of medical knowledge is largely embodied in narrative documents as free text, capturing and representing the semantic meaning of these documents would enable to create a bridge between clinical activities, evidence-based practice, and research. As Friedlin et al states, Natural Language Processing (NLP) system may represent this knowledge in various ways using one or more symbols. But the key point is that, no matter what formal knowledge representation system is used, each of these sentences should be represented by the *exact same symbols* since they are semantically equivalent. To this effect a novel approach to NLP of large medical corpi was devised and validated (Grant 9R44RR024929-02-National Center for Research Resources (NCRR)). The NLP processing used the *sentential proposition* as the atomic unit of semantic annotation. This unit captures the semantic meaning within a simple or complex sentence through propositional calculus or sentential calculus, a formal system of logic to obtain semantic equivalency [7].

Expanding on previously mentioned research, in this project we propose to develop and test an ontological framework as a structural model for a relational database system and associated clinical knowledge representation of semantically equivalent sentences while evaluating Protégé's ability to allow us to do so. Selected semantically annotated radiology reports from the study described earlier, in particular reports associated with the heart, provided a test vehicle to determine how well our framework represents semantically equivalent sentences in the knowledge structure. Furthermore, the model proposed has a potential for expansion beyond the specialty of radiology for the purposes of understanding and communicating medical information.

2 Methodology

2.1 Building Structural Knowledge Model with Protégé

The ontology in this project was built similarly to the four steps used in the TOVE (Toronto Virtual Enterprise) project [26]:

- Define ontology requirements.
  - What are the key concepts and relationships in the domain of interest?
- Define the terminology of the ontology (objects, attributes, and relations).
  - What are the terms that refer to the concepts?
- Specify the definitions and constraints on the terminology.
- Test the competency of the ontology.

The knowledge structural model was initiated by a high level abstraction of a "healthcare event" as experienced by a consumer and a provider with the concept healthcare events structured in triples of *subject* -> *predicate* -> *object* and based on the following description logic premise:

An individual (consumer) may experience signs and symptoms. Signs and symptoms may initiate a health event. A health event is serviced by an individual (provider and/or caregiver). Signs and symptoms are assessed using diagnostic evaluation studies (i.e.: medical imaging, medical history, lab, physical exam, mental exam). Diagnostic evaluation studies (ie: medical imaging – radiology, in this case) gather findings. Findings generate a diagnosis. A diagnosis helps identify intervention and treatment strategies. Intervention and treatment strategies impact the signs and symptoms. The intervention and treatment strategies produce certain outcomes. The outcomes are based on outcome criteria.

To transform the knowledge framework to the Ontology model, we formed a taxonomy defining the classes of our ontology first. In this step, the taxonomy had been reformed for multiple times as we began to develop the subclasses and slots. In our final version, there are eight main classes: (1) Body_Physiology; (2) Body_Systems; (3) Clinical_Documents; (4) Health_Event; (5) Individual; (6) Role; (7) Semantic_Propositions; and (8) Signs_and_Symptoms (Figure 1).

In Protégé the subclass relation is visualized in a tree. Subclasses represent concepts that are more specific. Also, if a class A is a subclass of a class B, then every instance of A is also an instance of B. In other words, as in any hierarchy, the subclasses are at the child level while the classes are at the parent level. Therefore, the child inherits the features of the parent. For example, radiology is a subclass of diagnostic evaluation. Therefore, radiology is a diagnostic evaluation study and will inherit the qualities of a diagnostic evaluation study. Once the classes and subclasses have been developed, they need to be described to identify the internal structure of concepts. Slots in Protégé can be used to describe the internal structure of concepts. That is, they can be used to describe the properties of classes. And, properties of the classes as described by slots automatically get inherited by the subclasses.

As we know that the heathcare is a complex system, the HEALTH_EVENT class was the most complicated and has the deepest level of subclasses. We organized the concepts and relationships using a logical diagram that can be described as:

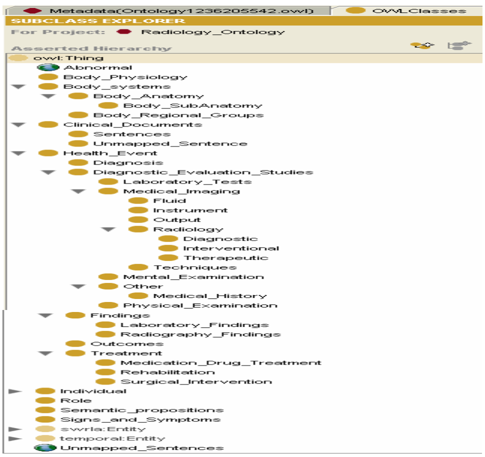


Fig. 1. Development of Main Classes (Concepts)

An individual has a role. An individual experiences signs and symptoms. Signs and symptoms are assessed using diagnostic evaluation studies. Diagnostic evaluation studies gather findings. Diagnostic evaluation studies are reported in clinical documents. Findings generate diagnosis. Diagnoses are also reported in clinical documents. A diagnosis helps to identify treatments. Treatments are also reported in clinical documents. Treatments have outcomes.

Clinical documents consist of sentences. Sentences can be expressed as propositions. If a sentence cannot be expressed as a proposition, then the sentence can remain as a sentence and hence unmapped. And the diagram of the Healthcare Events Concepts is presented in Figure 2 as follows.

Protégé provides Unified Medical Language System (UMLS) plug-in that allows users to search and retrieve concept name and id, semantic types, synonyms and terms from the UMLS Knowledge Source Server. We tried the UMLS plug-in for our project but we were unable to retrieve any results. However, we were able to search for Concept Unique Identifier (CUI) numbers for some if not all of the medical terms used in our ontology from the UMLS website http://www.nlm.nih.gov/research/umls/. For those medical terms that we could not find CUI numbers, we either left blank or assigned our own number.

At this point, our Ontology was developed in Frame-based and developed in Protégé Resource Description Framework (RDF). The RDF format supports a scalable representation of domain specific entities (e.g. metaclasses), relationships and constraints; including customized data entry formats and visualization of the knowledge structure using Jambalaya tool. Next, once the structured model developed was stable and validated using real life examples of diverse clinical events, the RDF format model was imported into Protégé and converted to the Protégé-OWL format to build logic into our ontology.

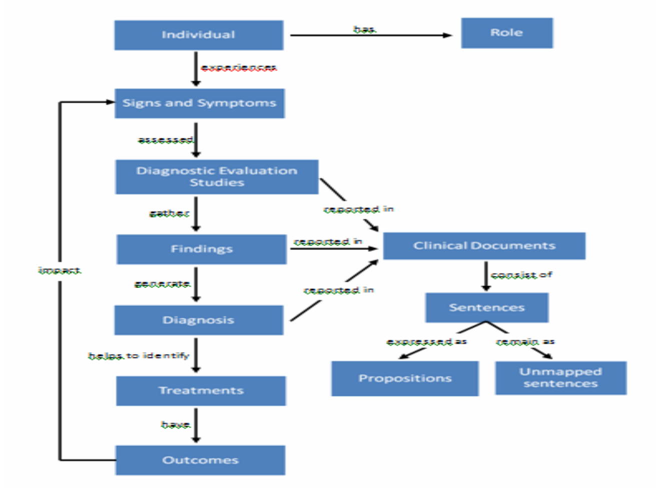


Fig. 2. The Concept of Healthcare Events

The benefits of Protégé-OWL format are to support consistency checking of the domain coverage; define logical class characteristics, the derivation a logical data model. The OWL format also has been extended with the Semantic Web Rule Language SWRL, which combines description logic constructs with capabilities similar to logic programming [19].

2.2 Constructing Logical Data Model and Relational Database

Relational databases are mainly utilized for storing and querying high volumes of data, while Ontology presents the structure of knowledge in a specific domain that can be shared and reused regardless of database systems. Moreover, this Ontology knowledge framework is useful for designing a database because it can suggest what entities and terms to include and how they are related to each other.

The purpose in this step is to link our Ontology framework in Protégé to the relational database, as well as to store processed data from the clinical reports since Protégé was not designed to store large amounts of data or have backing data functionality. There are some similarities between database schemas and ontology. A table in a database and a class in ontology both categorizes things in a set of objects with similar properties and relations that relate to other objects. And, they can be used to retrieve data and integrate data from multiple sources [14].

An Ontology Model consists of classes, properties (object and data type), inheritance, restrictions and other semantics. Classes are organized into a hierarchy in which subclass inherits object and data type properties from its superclass. A relational database is relational model that consists of tables, data types, constraints and columns [3]. Each column in the table contains a defined data type value and can only have one single value type. Tables and columns can have the following constraints null, not null, unique, check, primary and foreign key. The transformation of relational databases to ontology is based on a set of rules called mapping rules. These rules specify how to map related constructs of a relational model to ontological model [2], [22].

In our project we took the knowledge schema from OWL ontology and created a relational database using MySQL and then imported the database schema into Protégé 3.4 using DataMaster tool. The main purpose was to evaluate DataMaster tool and its functionality between relational database and Protégé. In this step, all OWL classes and subclasses were transformed into individual tables. Properties depending on their functionality in the knowledge structure were converted to junction tables. Modifiers and relationships became junction tables linking two tables of classes and/or subclasses and individuals (instances) were stored in its corresponding table as a row as illustrated in Figure 3.

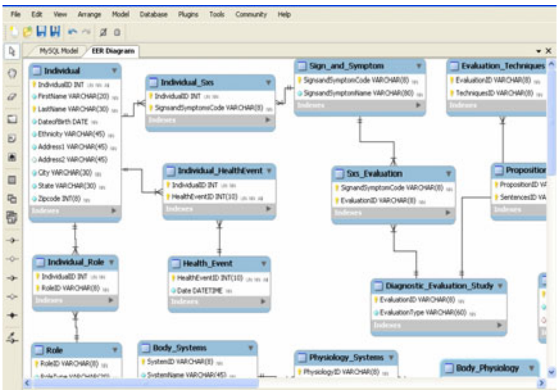


Fig. 3. EER Diagram in MySQL WorkBench

After the relational database was created, the tables were then exported using forward engineer to SQL script. Once the SQL script was saved, the script file was then opened in MySQL Query Browser. The Query Browser was used to execute the SQL script file to make sure no errors were found in our database.

Next, the database schema was imported into Protégé though MySQL Connector/ JDBC driver. This is the first step in building the knowledge framework from an existing relational database.

DataMaster is a relational-to-OWL importation tool that creates a knowledge level description of a relational schema in form of an OWL schema [16]. This tool automatically converts relational database tables into classes, columns as data property, and junction tables are mapped as object property. Rows within a table will be interpreted as ontology individuals (instances). The DataMaster plug-in can be found under Project Tab in Protégé. This plug-in is not a database backend and it currently does not have the capability to move data from Protégé back to a relational database.

Tables. Protégé will import raw data of that table to the assigned location. Using DataMaster plug-in, the tables are automatically mapped to a class unless all of the columns are foreign keys to two other tables then it is mapped to two object properties which become inverse of one another. For a junction or bridge table in a relational database, it will turn to be inverse object properties in Protégé.

Columns. For a column in a relational database, it is mapped to an OWL data type property. Databases have built in SQL data types where as OWL uses XSD which is the standard schema language for XML data types. Unlike SQL, OWL does not have any built-in data types. Instead, it uses XSD and RDF literals to map OWL data type to SQ.

In our database we created a lookup table (named Course), a table that contains values that is used in the main table, with the above allowed values and we added the primary key from the Course table which became a foreign key in the Signs and

Symptoms, Sentences and Semantic Proposition tables. Currently, Protégé does not populate data from lookup tables from database into the allowed values box. This is the reason why we created lookup tables to be able to select any allowed value that pertains to the selected instance from a class (domains).

Primary Keys and Foreign Keys. A constraint primary key when used as a column constraint is mapped to a data type property. A foreign key is a table constraint and a reference is a column constraint. Foreign keys identify parent and child entities of a relationship. A foreign key constraint is mapped to an object property unless it is also a primary key. Then it maps to class inheritance. As we previously mentioned, a subclass will automatically inherit object and data type properties from its superclass,

Queries tab in OWL allows users to search for instances based on class and slot values. For each class selected, a user will only be able to select the slots and instances connected to that class. The created queries can be saved and added to the query library. We were able to add the query, to the Query library tab at the bottom of the screen, but we were not able to find where Protégé stores the saved queries once we close the queries tab.

2.3 Implementing Reasoning Rules

One of the goals for this project was to enable decision making. To infer knowledge between classes, subclasses and instances and to build relationships between them, a rules engine is required. Rules are a form of expressing knowledge in the domain of interest that is a result of a consequence. For this project, the initial step was to pick twenty-six sentences and twelve semantic propositions from the radiology database that related to the heart. The sentences and semantic proposition criteria was formed using a horn like "if then" rules in SWRL. These rules represent the logic between the classes, properties and instances. These rules would enable automatic identification of classes in the knowledge base.

Initially the SWRL rules were used to query the ontology. We have used the SWRL rules to further map each sentence to its equivalent meaningful proposition. The information in this project is from a set of radiology reports which consists of sentences extracted from the reports and mapped them to semantically logical propositions. The following rule displays each sentence that is mapped to its semantically meaningful proposition.

Clinical Documents consists of sentences ***And*** *which are mapped to Propositions* ***And*** *has degree* ***And*** *as laterality* ***And*** *has shape* ***And*** *has position* ***And*** *relates to anatomy* ***And*** *related to sub anatomy*
Then *display*

In SQWRL syntax it is represented as:

Clinical_Documents(?document) **Λ** *consists(?mappedsen, ?document)* **Λ** *expressed_as(?mappedsen, ?meaning)* **Λ** *has_degree(?mappedsen, ?degree)* **Λ** *has_laterality(?mappedsen, ?laterality)* **Λ** *morphology_shape(?mappedsen, ?shape)* **Λ** *has_position(?mappedsen, ?position)* **Λ** *relates_to_anatomy(?mappedsen, ?anatomy)* **Λ** *relates_to_subanatomy(?mappedsen, ?subanatomy) → sqwrl:select(?mappedsen, ?meaning, ?degree, ?laterality, ?shape, ?anatomy, ?subanatomy, ?position)*

The more meaningful purpose of ontology would be to create inferences and in the future support decision support. Ontology can produce some intelligent inference using SWRL rules from existing knowledge [18]. Sometimes there could be rules which have a combination of both abnormal and normal.

For example, tortuous aorta otherwise normal (Sentence) is mapped to two propositions:

- The heart is normal (Proposition)
- The aorta is tortuous (Proposition)

In such a case, the property has classification would have two values normal for the sentence heart is normal and abnormal for the sentence the aorta is tortuous. So the result displays twice under the rule normal and twice under the rule abnormal. This is one of the challenges we faced in writing SWRL rules. A possible solution to such a mixed rule would be to split the sentence into two different sentences. The sentence can be split into two sentences such as tortuous aorta and Normal aorta. This would require some additional programming capabilities. The above rules and reasoning displayed some inferences which could enable decision support. The rules have been classified as normal, abnormal and procedures. This could highlight documents which are abnormal or normal. The next level of reasoning could be to create rules at the individual level. That would be a more granular level and could tie each individual's clinical document, symptoms, procedures and status.

2.4 Checking Consistency of Ontology Using Racer

Our final step was to check for the consistency of our ontology. For this purpose, we used Renamed ABox and Concept Expression Reasoner Professional (RacerPro).The origins of RacerPro are within the area of description logics [20]. Despite its various capabilities, it is used primarily for checking the consistency of OWL ontology and a set of data descriptions as shown in Figure 5.

3 Conclusion

The paper presented how we developed the relational database based on expert's knowledge in Ontology with Protégé. We proved that relational database can be connected to Ontology with Protégé tools. As a result, we can gain benefits from both Ontology and RDB. Ontology offers the knowledge framework that is easy to understand and be applied to different domains, while RDB offers a good data-storage along with other built-in functions in RDBMS.

The constructing of ontology is time consuming since it is a manual process and requires domain expertise knowledge in obtaining terms and concepts.

Currently, the version of DataMaster does not allow exporting schema structure or data from Protégé to a relational database. Therefore, the common use is for importing database schema before applying domain-knowledge framework to Ontology. Moreoever, a relational database that is compatible with Ontology frame work in Protégé must have junction tables. In most commercial EMR database, their schemas have designed with junction tables. However, for small-scale databases

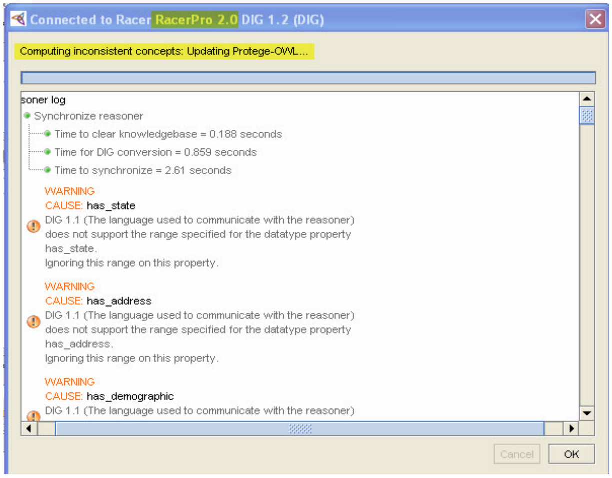


Fig. 4. RacerPro Checking for Consistency in the Ontology

might not include any junction tables. So researchers should manually implement junction tables in RDB or add properties of classes Ontology after importing RDB tables in DataMaster.

The main benefit of using Ontology is to share the knowledge database. In this project, through the Protégé project in RDF project can be transformed into a new ontology and model by others who wish to do so in the future. For a future project, the next level of reasoning could be to create rules at the individual level. That would be a more granular level and could tie each individual's clinical document, symptoms, procedures and status which leads to a more-precise decision support function.

Acknowledgement. Thanks to Drs. Jamieson P.(Logical Semantics, Inc.), Mahoui, M., and Friedlin, J. for their support.

References

[1] Ammon, D., Hoffmann, D., Jakob, T., Finkeissen, E.: Developing an architecture of a knowledge-based electronic patient record. In: Proceedings of the 30th International Conference on Software Engineering, pp. 653–660. ACM, Leipzig (2008)

[2] Astrova, I.: Rules for Mapping SQL Relational Databases to OWL Ontologies (2009)

[3] Astrova, I., Korda, N., Kalja, A.: Storing OWL Ontologies in SQL Relational Databases. World Academy of Science, Engineering and Technology 29 (2007)

[4] Cuenca Grau, B., Horrocks, I., Motik, B., Parsia, B., Patel-Schneider, P., Sattler, U.: OWL 2: The next step for OWL. Journal of Web Semantics 6(4), 309–322 (2008), doi:10.1016/j.websem.2008.05.001

[5] Davis, R., Shrobe, H., Szolovits, P.: What is Knowledge Representation? AI Magazine 1993, 17–33 (1993)

[6] Embley, D.W., Olivé, A., Ram, S.: Conceputal modeling. In: ER 2006: Proc. of the 25th International Conference on Conceptual Modeling, Tucson, AZ, USA, November 6-9 (2006)

[7] Friedlin, J., Jones, J., Mahoui, M., Jamieson, P.: Semantic Equivalence and Natural Language Processing of Biomedical Text. JAMIA under review
[8] Grimm, S., Hitzler, P., Abecker, A.: Knowledge Representation and Ontologies
[9] Gruber, T.: Toward principles for the design of ontologies used for knowledge sharing. In: Guarino, N., Poli, R. (eds.) Formal Ontology in Conceptual Analysis and Knowledge Representation, Substantial revision ed. Kluwer Academic Publishers, Dordrecht (1993)
[10] Horrocks, I., Patel-Schneider, P.F., Boley, H., Tabet, S., Grosof, B., Deab, M.: SWRL: A Semantic Web Rule Language Combining OWL and RuleML. W3C, May 21 (2004)
[11] Jovic, A.: Ontologies in Medical Knowledge Representation
[12] Mao, Y., Wu, Z., Tian, W., Jiang, X., Cheung, W.: Dynamic sub-ontology evolution for traditional Chinese medicine web ontology. Journal of Biomedical Informatics (2008)
[13] Minsky, M.: A framework for representing knowledge. Psychology of Computer Vision
[14] Mostowfi, F., Fotouhi, F.: Improving Quality of Ontology: An Ontology Transformation Approach (2005)
[15] Nyulas, C., O'Connor, M., Tu, S.: Datamaster-a plug-in for importing schemas and data from relational databases in protégé (2007)
[16] O'Connor, M.J., Shankar, R., Nyulas, C., Tu, S., Das, A.: Developing a Web-Based Application using OWL and SWRL. Stanford Medical Informatics
[17] O'Connor, M.J., Das, A.K.: A lightweight Model for representing and reasoning with temporal information in Biomedical Ontologies (2010)
[18] O'Connor, M., Knaublauch, H., Tu, S., Musen, M.: Writing Rules for the Semantic Web Using SWRL and Jess. Stanford Medical Informatics (2005)
[19] Obrst, L., Ceusters, W., Mani, I., Ray, S., Smith, B.: The Evaluation of Ontologies. Semantic Web, 139–158 (2007)
[20] Racer Systems GmbH & Co. KG. RacerPro User's Guide Version 1.9.2 (October 2007), http://www.racer-systems.com
[21] Rothwell, D.J., Cote, R.A., Cordeau, J.P., Boisvert, M.A.: Developing a Standard Data Structure for Medical Language - The SNOMED Proposal. AMIA, 695–699 (1994)
[22] Sane, S., Shirke, A.: Generating OWL Ontologies from a Relational Databases for the Semantic Web. In: International Conference on Advances in Computing, Communication and Control (ICAC3 2009) (2009)
[23] Schulz, S., Stenzorn, H., Boeker, M., Smith, B.: Strengths and Limitations of Formal Ontologies in the Biomedical Domain. RECIIS Rev. Electron. Comun. Inf. Inov. Saude, 1–27 (2009)
[24] Stevens, R.: Ontology Based Document Enrichment in Bioinformatics. Comparative and Functional Genomics 3, 42–46 (2002)
[25] Sure, Y., Staab, S., Studer, R.: Methodology for Development and Employment of Ontology Based Knowledge Management Applications. SIGMOD Record 31(4), 18–23 (2002)
[26] Uschold, M., et al.: The Enterprise Ontology. The Knowledge Engineering Review 13, 1–18 (1998)
[27] W3C, Resource Description Framework (RDF), World-Wide Web Consortium (2000), http://www.w3.org/RDF/
[28] Zhang, S., Bodenreider, O.: Experience in Aligning Anatomical Ontologies. SInt J. Semant. Web Inf. Syst. 3(2), 1–26 (2007)

Increasing Physical Activity by Implementing a Behavioral Change Intervention Using Pervasive Personal Health Record System: An Exploratory Study

Hadi Kharrazi and Lynn Vincz

535 W Michigan, Indianapolis, IN, 46236, USA
kharrazi@iupui.edu, lvincz@imail.iu.edu

Abstract. The obesity epidemic is a major health concern. More than two thirds of Americans and more than one third of children and adolescents are overweight or obese. Physical inactivity and sedentary lifestyle are contributing factors of overweight and obesity. Increasing physical activity in overweight and obese individuals can potentially improve health status and lower the risk of major health problems. Based on the theoretical frameworks of behavioral change models, health information technologies, and pervasive interactive technologies, a conceptual model has been developed. The proposed model is an intervention method combining a behavior change model and a personalized Health IT system using an interactive accelerometer-based pervasive technology component, to increase physical activity potentially in an overweight and obese population. The accelerometer automatically synchronizes with the personal health record database over wireless networks; the decision engine utilizes the theory of planned behavior model to form a feedback; and; proper interactive feedback is provided directly to the users in the personal health record interface. A prototype was developed based on the conceptual model. The prototype measures and stores the number of steps, type of steps and caloric consumption on a daily basis with an interval of 20 minutes by an advanced accelerometer that can simply be attached to any footwear. The accelerometer synchronizes automatically the stored data with a central database as soon as the participant enters a designated wireless area. Virtual characters that are embedded inside a personal health record system convey the feedback to the users. The feedback is generated based on the theory of planned behavior decision model utilizing the stored daily physical activity data. The interface prompts knowledge-based information, presents self-monitoring trend lines, and shows virtual character changes over time such as becoming obese, to reflect the tailored feedback message. In addition, the interface is integrated in a popular social-networking website where users can review other virtual characters and generate peer pressure. The purpose of this study is to explore the feasibility of implementing a behavioral change intervention using pervasive health information technology among overweight and obese population. This research explains the development of the conceptual model, integration of wireless physical activity self-monitoring devices into personal health records, and the application of rich interactive media to convey messages to the users. The final prototype will be presented and preliminary findings will be discussed.

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 366–375, 2011.

1 Introduction

Recent debates and discussion about health care reform have sparked interest in taking a closer look at some of the factors related to the skyrocketing cost of health care. There has been extensive peer reviewed research about some of the factors related to the high cost of health care, such as a sedentary lifestyle that can lead to overweight and obesity. These factors increase the risks of major health problems such as high blood pressure, diabetes and heart disease.

The obesity epidemic is a major health concern because two thirds of American adults (66%) are overweight or obese, and one third (17%) of children and adolescents are overweight or obese (Wang, 2007). Children and adolescents who are overweight are most likely to be obese as adults. The medical costs attributed to overweight and obesity in most U.S. adults is projected to be more than $950 billion by the year 2030, and obesity prevention could potentially reduce by 9.1%, the current total U.S. expenditure related to obesity and overweight (Anderson, 2005). The later study found obesity factors to be associated with 23% of health plan health care charges and 27% of the national health care charges, and also that health behaviors and health charges related to obesity have a significant association.

The use of information technology such as the interactive personal health record (PHR) can be used to address the issue of a sedentary lifestyle. The PHR can be useful in facilitating self-management of a sedentary lifestyle by empowering individuals to take control of their health, by tracking individual physical activity and progress in real time, and providing health education materials in the format of an enriched interactive health multimedia. In addition, PHR interactivity can facilitate behavior modification by supporting goal setting and providing feedback on the success of the behaviors. The purpose of this study is to review and propose a conceptual framework that increases physical activity by implementing a behavioral intervention using PHR in a sample target population of college freshmen.

2 Background

There is a high cost per person for health care in the U.S. and almost two trillion dollars are spent in health care on an annual basis. Studies have identified that “PHR’s have the potential, if designed appropriately and adopted widely, to reduce costs and simultaneously improve quality and safety of care” (Kaelber, 2008, p. 733).

Several studies (Ahern, 2007) have targeted increase in physical activity and reduction of BMI (Body Mass Index) through a reduction in sedentary lifestyle. For example, a study focused on limiting the amount of television time for young children found that the intervention group had a decrease in BMI (Ahern, 2007). According to a study in exploring the role that technology can play in reducing the incidence and prevalence of childhood obesity (Ahern, 2007), one of the key areas noted for further exploration included the use of e-health to increase physical activity at the individual level. For example “mobile/wearable/portable technologies were seen as the most likely to engage youth and applications must be interactive and provide feedback” (Ahern, 2007, p. 27). Thus, the use of online interactive interventions could be a potential method for promoting physical activity and decreasing sedentary behaviors.

More evidence is needed about optimal incorporation of theoretical frameworks for weight management programs for children and youth via the Internet (An, 2009). One study reported that inclusion of theory-based intervention strategies for physical activity through the Internet was at a low level (An, 2009), despite a need to address eating and activity habits during childhood (Jain 2004). Internet-based physical activity intervention studies found evidence that this type of intervention was effective, however "more studies are needed that use appropriate research designs, in which the only difference between the intervention and control groups is the addition of a specific component, such as personalized supervision" (Vanden Berg, 2007, p4).

The Theory of Planned Behavioral (TPB) model examines the relations between the individual's beliefs, attitude, intentions, behavior and perceived control over the behavior (Hardeman, 2002). TPB is the most extensively studied social cognition theory that is relevant to both intention and behavior change, and it is also the most promising behavior change model in preventing weight gain (Daddario, 2007). A study on review of the use of the health belief model for weight management found that The Theory of Planned Behavior (TPB) was one of the models that could predict the variance in intention to diet, and fasting behavior, but future studies are needed to address more questions related to weight loss and obesity (Daddario, 2007).

Despite the widespread interest and activity by many individuals, there has been very little interactive PHR research done. The use of an interactive PHR based on a behavior change model such as TPB to increase physical activity has not been addressed in the current literature.

3 Methods

Peer reviewed published articles were considered in this review. A search for systematic reviews using combinations of key words: obesity and health IT, behavior change model, Personal Health Record (PHR), physical activity, and exercise adherence were applied in this review. A search in PubMed, PsychInfo, Ovid, and EBSCO did not reveal any publications comprising all interventions in one solution.

Due to the lack of publications in combining Health IT (e.g., PHR), Interactivity (e.g., health games), behavioral change models (e.g., TPB) in the context of Obesity prevention (e.g., physical activity), this review explores and proposes a methodology to develop and implement such a solution by effectively integrating its various components. Followed is a discussion about a proposed conceptual framework that utilizes these components.

3.1 Components

Combining health information technology, interactive media, medicine, and psychology as an intervention method for a disease management program could be effective in obesity prevention. The following lists various components that can be employed to deliver an effective intervention:

Health Information Technology (e.g., PHR): A core system is required to store, manage, and analyze the health data that is being collected from the users while using it. For example, collected physical activity levels, weight, and BMI should be stored and proper feedback should be provided to the users through motivating elements. There are multiple commercial (Open API) and non-commercial (Open-source) PHR systems that can be utilized to fulfill this task.

Pervasive Data Collection (e.g., wireless sensors): In addition to the management of health information in PHR systems, due to the nature of obesity prevention through physical activity, daily health data should be collected. Data (e.g., physical activity levels) should be captured effortlessly, accurately (valid and reliable), and automatically. This can be achieved by various wearable and portable sensors that capture the necessary data continuously. Data should be synced, preferably without any efforts (e.g., wirelessly), with the central PHR system on a daily basis.

Interactive Interface (e.g., Health Games): Interactive interfaces can deliver the PHR consumer guideline feedback to the users effectively. The system should be fun, motivating, and empowering; otherwise the users may terminate their interaction prematurely. An integrated health game can simulate various interventions (based on a particular behavior change model) in a virtual world. For example, peer pressure can be achieved by sharing the status of a user's virtual character with other players.

Behavior Change Model (e.g., Theory of Planned Behavior): Behavior change models can be used to tailor intervention strategies to a specific population need. For example, these models can address the individual's readiness to change and it is used in disease management program interventions. The TPB model is a behavior change model that can identify 50 percent intention and 25 percent behavior change (Hardeman, 2005) and thus a suitable model to shape the PHR interventions.

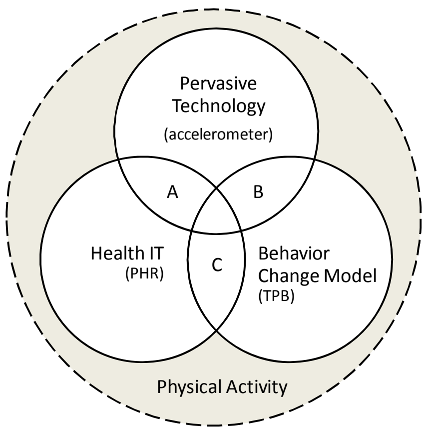


Fig. 1. Integration of HIT (PHR), BCM (TPB), and Pervasive Technology (accelerometer) (A) Automatic sync of the accelerometer with the PHR database, (B) Providing feedback to the TPB decision model, (C) Presenting proper interactive TPB-based feedback in PHR

3.2 Theoretical Perspective

The use of health information technology can improve prevention and treatment of obesity at the individual level through the use of a PHR to track physical activity level

(Kaelber, 2008). Behavior change models are important in preventing weight gain, and the TPB is the most promising behavior change model in preventing weight gain (Daddario, 2007). TPB is most studied in young adults. TPB is the most extensively studied social cognition theory, and is relevant to both intention and behavior change (Hardeman, 2002). Health games (i.e., highly interactive media) have been shown to be effective in adapting behavioral models in a number of healthy behaviors such as adherence to medication among diabetics (Kharrazi, 2009) and patients with asthma (Lieberman, 2006).

4 Proposed Development

A technical solution that combines the aforementioned components based on the theoretical perspectives of the TPB model is currently under development. Following is a discussion of the current developments/plans:

Hypothesis: The alternative hypothesis (H_A) is that the increase of physical activity measured in steps by an accelerometer in the target population that interacts with the integration of a TPB based behavioral intervention in an interactive PHR will be greater than the control group that interacts only with the standard PHR.

Target Population: A population of 60 college freshmen students aged 18 to 35 yrs will be recruited from IUPUI campus. Volunteer subjects will not be included if they self report any underlying health conditions that limits their functional levels or restricts them from physical activity.

Study Design: The study design is a between group study based on independent and dependent variables predicting an alternative hypothesis. The independent variables are the existence of the interactive versus the standard PHR in the proposed information technology solution, and the dependent variables are the number of steps per day measured in unit of time, by an accurate accelerometer, and the intention, obtained from the TPB questionnaire, (i.e. behavior), and the physical measurements including height, weight, blood pressure and pulse.

Students from both classes will be in either the treatment or control group. All students will receive an orientation providing detailed information about the study and that each participant will be provided with an accelerometer to track steps taken for a specified period of time. Participant in the treatment group will have access to an online PHR through a Facebook© application to review and track their physical activity information, and they may also receive educational and coaching information about healthy lifestyle and the benefits of physical activity. Facebook© was selected as the container for this interactive process due to its high usage frequency among the participants (e.g., daily logins), which in turn ensures a higher interaction rate with the proposed intervention. A validated TPB questionnaire will be adopted and adjusted for use in this study. The TPB questionnaire will identify past behavior (self-reported), perceived behavioral control, subjective norm, attitude and intention, motivation to comply, behavioral beliefs, control beliefs, power of control factors, and normative beliefs. Data will be collected throughout the study and cross-sectional measurements will be collected and recorded four times throughout the study; prior to the start of the baseline period, prior to the start of the intervention period, prior to the start of the follow-up period, and at the end of the follow-up period. During the intervention period, steps taken will be measured in units by an accurate sensor (i.e., accelerometer). As discussed, the treatment group will receive an interactive PHR while the control group will receive a standard PHR with no interactive feedback. During the follow-up period, both groups will receive the standard PHR

Study Details: Participants will receive an orientation and baseline information will be collected prior to the start of the study. The TPB questionnaire will be given to participants to complete. Measurements of height, weight, blood pressure, and pulse rate will be obtained and recorded at the beginning of the baseline period and measurements of weight, blood pressure and pulse will be measured and recorded at three additional intervals during the study. Participants will be given an accelerometer to wear at the beginning of the baseline period. The number of steps taken will be measured in units by an accurate accelerometer for one week. At the end of the baseline period, weight, blood pressure, and pulse rate will be measured prior to start of the intervention period. Goals will be set based on the individual participant level, prior to the start of the intervention period. During the intervention period the treatment group will receive the interactive PHR and the control group will receive the standard PHR for four weeks. At the end of the intervention period, weight, blood pressure and pulse rate will be measured prior to the start of the follow-up period. During the follow-up period, both groups will receive the standard PHR for two weeks. At the end of the follow-up period weight, blood pressure and pulse rate will be measured. The number of participant steps taken will be automatically uploaded to an online PHR through a wireless access point.

The Accelerometers provided to be worn on the top of the participant's shoe will accurately count the number of steps taken in units, and a wireless access point will be used to offload data to a web site automatically when the participant is within range of the unit, approximately 300 feet. Offloading of the data will be seamless to the participants. The ActiPed accelerometer by Fitlinxx was selected for use in this study based on reports of accuracy and reliability as well as testing of this device. "The ActiPed is a very small, wireless activity sensor that clips onto any shoe and accurately tracks steps, distance traveled, calories burned and activity time" (Fitlinxx.com, 2009). The ActiPed is different than a pedometer in that it measures foot contact time with the ground. It measures the amount of time the foot is making contact with the ground from the time the foot strikes the ground to the time it lifts off the ground. The ActiPed contains some complex algorithms that can determine distance, calorie burn, and activity time as well as count number of steps. Most pedometers count steps by using a pendulum method that can be fooled by any slight undulation that will cause the pendulum to swing and count a step. Shaking the ActiPed will not be able to fool the device; therefore the ActiPed accelerometer is much more accurate than a pedometer. The ActiPed can be worn on different types of shoes and sneakers. It is ideal to wear it flat on the top of the foot however it can be worn on the side of the shoe. It can also be place on the top part or opening of a boot, however it may only capture 75 percent of activity when worn perpendicular to the ground. It does need to be placed snug on the shoe with the open part of the clip facing down toward the shoe.

Fig. 2. ActiPed (Fitlinxx, 2009)

The Intervention: The intervention includes collecting baseline data initially and at various stages, including measurement and recording of height, weight, blood pressure, and pulse rate prior to the onset of the study. BMI will be calculated based on height and weight, and recorded. A physical activity questionnaire will be given to participants to complete. Daily and weekly activity goals will be set for participants based on their response to the physical activity questionnaire. Goals will be based on a percentage of improvement for each participant based on individual goals, for example, a participant who normally takes 3000 steps per day will have a goal of 5000 steps per day, and participant who normally takes 5000 steps per day will have a goal of 7000 steps per day.

A reward system will be implemented. The reward (Medal; bronze, silver, gold) will be based on a percentage of increase in physical activity from the participant's baseline, for example 25% increase, 50% increase, or 75% increase in physical activity. For example, a participant at the 3000 steps per day level with a goal of 5000 steps per day will receive a reward at the 50 percent (silver medal) level for achieving 4000 steps per day. A participant at the 5000 step level with a goal of 7000 steps per day will receive a reward for achieving 6000 steps per day. The PHR will track steps, calories, distance/miles, active minutes, date/time of day, activity type such as walking, running, or other and will also track, daily, weekly, monthly, and cumulative totals.

The online PHR will be accessed through a Facebook© interactive interface application (i.e., embedded health game). Avatar characters will be created and displayed on the Facebook© page of each participant, and visual feedback will be provided to the participants by a display of emotions and/or size of the Avatar. The PHR will be interactive with ability for participants to view their progress and make changes in the appearance of their avatar (e.g. change colors, or clothing etc.). Participants will also be able to view activity of peers related to percentage of increase in physical activity but not the actual number of steps taken by peers. Participants will be able to view their own actual number of steps they have taken. Participants will be able to compare their own individual progress with the group progress and also compare their progress as a class with the progress of the other class. All PHR logs on activity through the Facebook© interactive application will be monitored and recorded. Medals will be awarded to participants for achieving 50% increase in weekly activity. At the end of the study a Trophy will be awarded to the class that achieved the highest overall percentage of increase in physical activity.

Table 1. Intervention; Affect of TPB and Game element

Intervention	Target TPB Element	Game Element
Visual Feedback	Behavioral Belief/Attitude	Character specifications (e.g., emotions, size)
Motivation	Normative Belief/Peer Pressure	Compare progress, Trophy status, Medal
Knowledge	Control Belief	Tracking step numbers, Educational material

Measurements: Blood pressure will be measured and recorded at the start and at the conclusion of the intervention period. Weight and blood pressure will be measured and recorded again at the conclusion of the study. Participant steps per day will be measured by an accelerometer and will be automatically uploaded to the Interactive Web-based PHR through a wireless access point. Each accelerometer offload date and time will be captured and stored automatically.

Technical Design: ActiPed accelerometer will sync wirelessly with the ActiHealth database (ActiPed commercial database) seamlessly. The custom PHR database will pull the data from ActiHealth database. Based on the pulled data, the content of the Facebook© application (e.g., health game) will be updated.

Limitation: Accuracy of reporting is a consideration, as well as adherence to using the accelerometer and placing it appropriately on footwear for optimal functioning. Improper attachment of the device to the participant's footwear as recommended by the manufacturer could result in technical issues related to accelerometer malfunction.

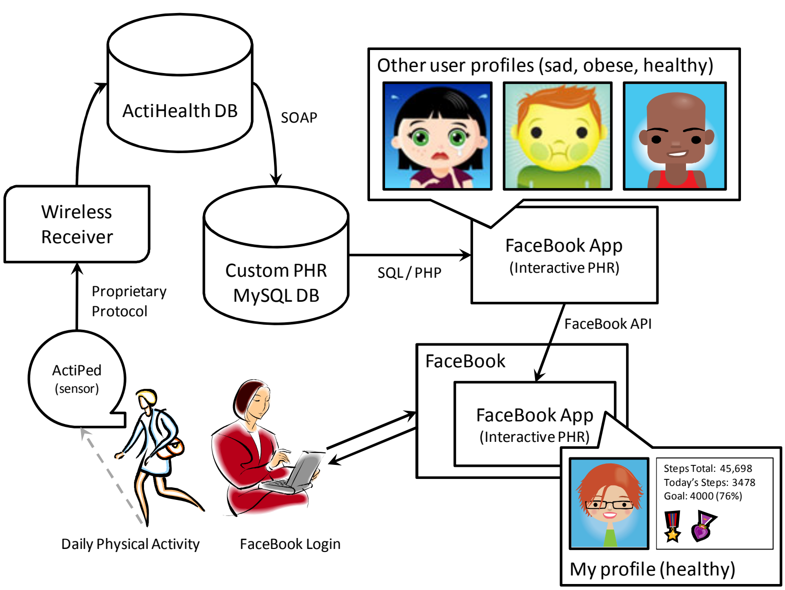


Fig. 3. Technical Design

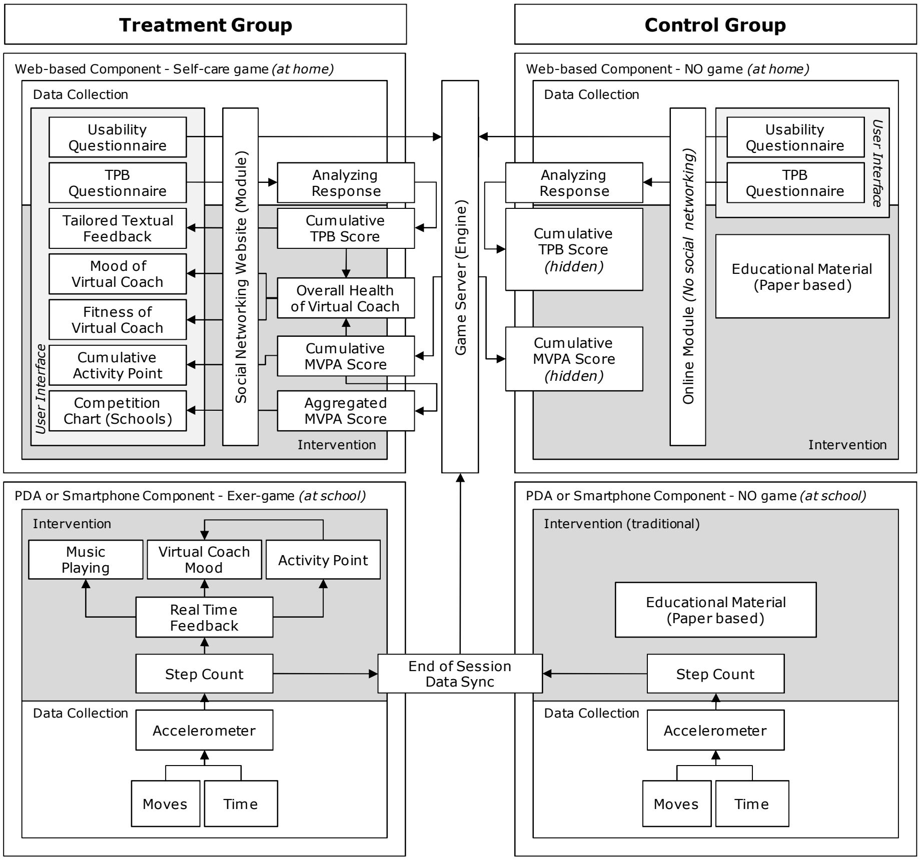


Fig. 4. Proposed Research Design / IT Interaction Framework

The assumption is that there will be minimal user bias. Issues with the methodology, design and sampling could present potential issues, including participant dropout rate. Technical issues related to server down time would require a back-up plan.

5 Conclusion

Increasing physical activity in overweight and obese individuals can potentially improve health status and lower the risk of major health problems such as high blood pressure, diabetes, and heart disease. The purpose of this study is to explore the implementation of a TPB based behavioral intervention using an integrated health game and health information technology system in a target population of urban college freshman. Expected outcome is that the TPB enhanced PHR intervention method using HIT will have a positive impact on sedentary lifestyle and increase physical activity in the target population, as evidenced by the accelerometer steps measured.

References

1. Ahern, D.P.: Childhood Obesity Prevention and Reduction: Role of e-Health: Exploring the role that technology can play in reducing the incidence and prevalence of childhood obesity. Health e-Technologies Initiatives, 1–44 (2007)
2. An, J.H.: Web-Based Weight Management Programs for Children and Adolescents. Advances in Nursing Science, 222–240 (2009)
3. Anderson, L.M.: Health Care Charges Associated with Physical Inactivity, Overweight, and Obesity. Preventing Chronic Disease, 1–12 (2005)
4. Daddario, D.K.: A Review of the Use of the Health Belief Model for Weight Management. Med. Surg. Nursing, 363–366 (2007)
5. Fitlinxx.com (2009). Activity and Health Tracking Technology (retrieved April 21, 2010)
6. Hardeman, W.J.: Application of the Theory of Planned Behavior Change Interventions: A Systematic Review. Psychology & Health, 123–158 (2002)
7. Jain, A.: What Works for Obesity? A summary of the Research Behind Obesity Intervention. BMJ Publishing Group, London (2004)
8. Kaelber, D.J.: A Research Agenda for Personal Health Records. Journal of the American Medical Informatics Association, 729–736 (2008)
9. Kharrazi, H.: Changing the Behavior of Type 1 Diabetic Patients by Digital Games. In: 5th Annual Games for Health Conference, Boston, USA (2009)
10. Lieberman, D.: What can we learn from playing interactive games? In: Vorderer, P., Bryant, J. (eds.) Playing Video Games: Motives, Responses, and Consequences, pp. 379–398. Lawrence Erlbaum Associates, Mahwah (2006)
11. Vanden Berg, M.S.: Internet-Based Physical Activity Interventions: A Systematic Review of the Literature. Journal of Medical Internet Research, 1–16 (2007)
12. Wang, Y.B.: The Obesity Epidemic in the United States - Gender, Age, Socioeconomic, Racial/Ethnic, and Geographic Characteristics: A Systematic Review and Meta Regression Analysis. Epidemiologic Reviews Johns Hopkins Bloomberg School of Public Health, 1–23 (2007)

Exploring Health Website Users by Web Mining

Wei Kong and Josette Jones

Health Informatics, School of Informatics, Indiana University,
Indianapolis, Indiana, Usa
weikong@umail.iu.edu, jofjones@iupui.edu

Abstract. With the continuous growth of health information and users of the Internet, how to build an easy-to-use interface for different users is one fundamental desideratum when constructing a health website. The goal of this paper is to explore different information needs by examining the search terms of different Internet user groups. In order to deeply investigate the information, five months' daily access weblog files from one particular health provider's website are collected and the log data is analyzed by Web Mining technique. Based on the mining results, the paper also gives some suggestions of how to redesign the interface to be more intuitive to users.

Keywords: Web Mining; Web Log; Internet; User-Computer Interface; Search Term; Information Seeking Behavior.

1 Introduction

With the rapid development of the Internet and technologies used in the health area, people have more opportunities than ever to resort the Internet for health information seeking. Surveys [1-2] have shown that more than half of patients have used the Internet to access health information. In addition, more than 70% of Internet users prefer to use search engines rather than medical portals or libraries to start the information searching[3]. Several studies [4-7] have also described the importance of the use of the World Wide Web (WWW) as a source for health information and demonstrated that individuals who seek health information on line for decision-making promoted disease management and thus improving quality of life. It's very clear that the Web is progressively playing a significant role in patients' healthcare, and the impact of the Internet cannot be overlooked.

However, the information on the Web is huge and overwhelming. Although general search engines, such as Google and Yahoo are good starting points for users, the precision of the information retrieval results still need to be improved [8-10]. Obtaining the maximal benefits from the Internet must be built on understanding the users' interests, characteristics, and preference first. Understanding the users' preference is the first step to provide the tailored health service and user-friendly Web interfaces. The purpose of this study is to examine users' information seeking behaviors based on different user groups by extracting their search terms with Web mining technology.

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 376–383, 2011.

Like Lambert & Loiselle mentioned, "*Seeking information about one's health is increasingly documented as a key coping strategy in health-promotive activities*"[11]. Some studies have shown some progress of Web design by mining the user behaviors in health domain. Chen and Cimino[12] analyzed a Web-based clinical information system's (New York Presbyterian Hospital) logs to discover patient's pattern of usage. The result of mining data indicated that users usually view radiology and laboratory results in one session. Hence, they suggested adding "shortcuts" in these Web pages to provide patients a quicker access to the information. Graham and Tony[13] did navigation research on a consumer health website (ClincialTrials.gov) and one of their findings showed that majority of the users were referred by general search engines to access the webpage. Therefore, they suggested increasing the use of general search engines like Google. Rozic-Hristovsk and Hristovski[14] investigated the usage of the central medical library of their University by exploring weblog files and decided to reconstruct more intuitive reference pages to fulfill the increasing visitors. These applications have represented that Web log analysis is a powerful tool for researchers to explore the users' usage patterns and correlate these characteristics to website construction.

However, most of the applications focused on all Web users instead of specific users. As Rozic-Hristovsk and Hristovski[14] stated in the limitation discussion in their study that "*the analyse adequaletly reveal overall usage patterns but can only provided estimated of individual user characteristics*". Although it is hard to predict every single user's preference, it is useful that users can be divided into groups to examine their information needs.

In past studies, some researchers investigated health information seeking behavior from either a patient's or a physician's perspective. A study[15] of cancer patients' information needs showed that all participated patients just want basic information rather than every detail information at all stages of their illness, while other study[16] has shown that primary physicians would like to spend more time gathering information focused on the diagnose and treatment. Also, people would like to assume that there is a different preference between the Web users when they seek information. For example, HON.ch [17] provides different search options and the result are different even we search a same term. For patients, it provides some consumer health website links to other websites like MedlinePlus, WebMD and family doctor. While for health professionals, the result is more focused on professional articles from journals and knowledge base. For example, it provides some articles from eMedicine, which is an online clinical medical knowledge maintained by WebMD. Although users are "assumed" differently, limited research and comparison was done to prove this assumption quantitatively. In this study, we will investigate different information needs of patient and doctor group based on the Web query analysis.

2 Method

The website of Clarian Health [18] was used for this study. Clarian Health, now renamed as "IU Heath", was first formed in 1997 and it is a private, nonprofit organization owns more than 20 hospitals and health centers throughout Indiana.

Daily access log files of this website were collected and analyzed for pattern detection in search behaviors of different user groups such as patients, providers and occasional visitors. As is suggested from the literature and communication with users, their needs are complementary but different.

Web mining technology was used for data analysis. WUM-prep and Perl script were used to clean the original data; descriptive statistics described the features of the logs and user groups; association rule was employed to discover the frequency and patterns of the users; auto-classification categorized the users even without logged information.

2.1 Data Collection

The study was based on five months' weblog collected in 2007. The usage data of this website is considered sufficient to provide some trends, and also the website has already built up the navigation bars for patients, physicians and the visitors, which would greatly facilitate the user classification. Figure1 is a screenshot of this website.

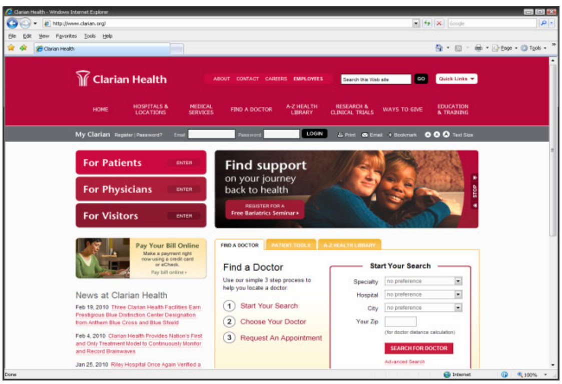


Fig. 1. Screenshot of Clarian's website

2.2 Data Preparation

Raw log data was processed to reduce the noisy. After removing the identification information of the users, we got rid of the Web spiders, irrelevant and duplicate records. Web spiders, also called Web robots or Web crawlers, are programs that automatically collect relevant contents from Web pages, so the search queries generated by these spiders do not represent the actual information needs of the real users. And then user sessions were made by cookie and 30 minutes time constrain. The user groups of patients, doctors and visitors were separated by URL. The last step was to extract the search terms of the users. Figure 2 describes the whole process.

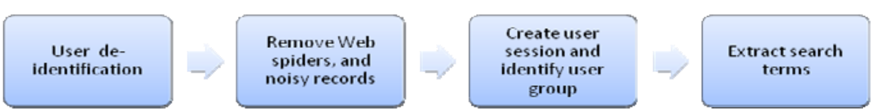


Fig. 2. Process of Data Preparation

3 Result

3.1 Log File Statistics

Figure 3 describes the statistics of the log files. There were totally 11 million original log records, but after the cleaning process, there were only 58% left to be used to process the study. And during the five months' time period, April and May have the top visit amount. This may be due to the season's reason or some event happened during that time period.

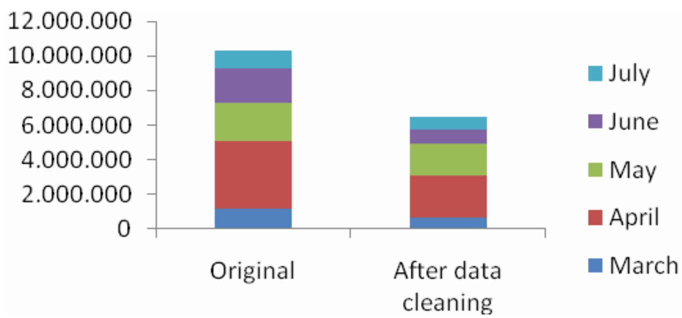


Fig. 3. Log file statistics

3.2 User Session

Patients have around 200 thousands user sessions, which are almost 10 times of doctors. No user logged as a visitor during five months period. 73% percentage of the users did not log as any of the user groups when they were surfing the website. Although this website has built the log in button, majority of the users still didn't logged, so they might not have the special service provided based on user groups.

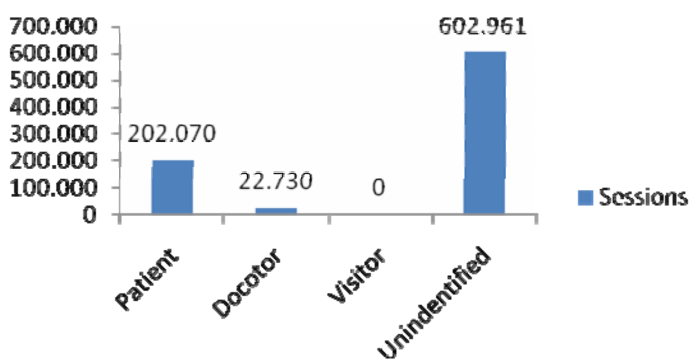


Fig. 4. User sessions

3.3 Search Engine Distribution

We examined the query request from four most popular search engines, Google, Yahoo, MSN, now is Bing, and the Clarian's site search. The result has shown that 30% of the users employ search engine to look for health information and google.com is the most popular search engine for both users logged as patients or doctors. It can be seen that intranet is well used through Web and we recommend that this website could consider increase the server support ability and optimize the website to Google.

Another finding we can see from the result is that doctors relied on site search engines much more than patients. It can be predicted that when doctors were browsing this website, they usually cared about specific topics rather than general ideas. So they preferred to search directly in Clarian's site.

Figure 5 give the search engine distribution chart for both patient and doctor group.

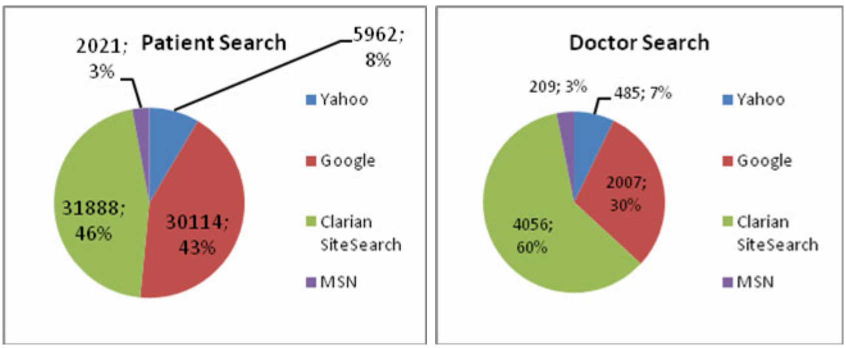


Fig. 5. Search engine distribution

3.4 Search Term

We extracted the top 20 search terms from Clarian's site search engine according to patients and doctors groups.

For patient group, it is surprise to see that more than half of the terms are related to employment and education information. In another word, users logged as patient cared about jobs and training rather than health information. This prediction may be true, but there may be another reason that the users logged as patients are not the real "patients". They might be some "seekers" looking for employment or training. As they didn't know where to go through the home page, they clicked on "patients" to start. In this case, we can see that the homepage interface is not intuitive for these users. They were confused and mislead by the webpage.

Compared to patient group, doctors used more medical terms to search, like pathology, pain. However, they were more likely to use this website as a handy tool to search auxiliary information, as doctor detail information (by search Dr.name), patient medical record (by careweb) and lab or surgery data.

Table 1. Top 20 Clarian Search Terms with Term Frequency

Patient		Doctor	
job	814	dr.	214
clarian	666	center	75
center	641	clarian	69
methodist	616	medical	68
employment	598	care	58
health	582	methodist	55
medical	571	health	41
care	480	*pulse*	40
human	409	john	39
patient	387	physician	38
program	355	*surgery*	35
resource	349	clinic	34
address	323	doctor	33
hospital	294	careweb	33
employee	293	*laboratory*	31
nurse	289	*lab*	28
class	284	*cancer*	27
career	241	*pathology*	26
dr.	282	*transplant*	26
service	254	*pain*	23

3.5 Association Discovery

In this study, we also applied market basket analysis[19] to find the associations of the popular search terms. As we know which terms patients or doctors are most likely to search together, we can better understand the users' needs, and also provide dynamic site search suggestions to users to promote their search. Table 2 lists some of the associations found with high confidence rate.

Table 2. Associations for patient and doctor groups

Patient			Doctor		
Term1	**Term2**	**Confidence**	**Term1**	**Term2**	**Confidence**
human	resources	92.83%	order	sets	95.24 %
therapy	physical	84.80%	women's	health	84.62 %
phone	number	83.33%	west	clarian	84.62 %
life	child	93.52%	group	medical	70.00 %
information	patient	56.57%			
community	plunge	91.30%			
records	medical	92.31%			
people	mover	97.44%			
financial	assistance	72.97%			
occupational	health	69.23%			

3.6 User Group Classification

As we seen before, 73% of the users did not log as any group when browsing, so any "tailored" service would not be available to them. Nevertheless, if a classifier can be built to automatically identify the user role based on the search terms users input, majority of the users can still get benefit even they don't log in.

In this study, we tried several popular classifiers, like naïve Bayes and neural network. Among these, we found the binary linear classifier, Support Vector Machine (SVM), has the best F-score. The classifier was tested based on 600 patient queries and 600 doctor queries randomly selected from the five month data. With this classifier, we are able to categorize the users. So when people search information, they can be suggested or directed according to their user roles, no matter whether they remembered to log in or not.

Table 3. Performance of SVM classifier

		Patient (F = 74.57%)	**Doctor (F = 84.00%)**	
Predict	Patient	True Positive = 447	False Positive = 96	Precision=82.32%
	Doctor	False Negative = 153	True Negative = 504	Precision = 76.71%
		Recall:= 74.50%	Recall:= 84.00%	

4 Conclusion

The previous results prove that users logged as patients or doctors have different information preferences. Web mining technology can help us understand what information the users are really interested. Based on the results we found, we would purpose some suggestions to redesign the website and build more user-friendly interface for different users. The suggestions are summarized as below.

- For the homepage, remove the "visitor" log portal, and instead of it, build a log portal for employment careers, like "employer" or "future employee".
- Differentiate the entry pages for different user groups. For patient group, build friendly links to training, education programs and general information.
- For doctor group, build intuitive links to people contact directory, knowledgebase, and auxiliary medical data access.
- For search engine, provide dynamic searching suggestions and implement the SVM classifier to promote the search criteria.

5 Limitations and Future Study

This study only represents the users' seeking pattern from one website and the results can only be used as an estimate data for other health websites. The user separation is based on the log in information, so we cannot say the user groups of patients or doctors are the real patients and doctors in life. As we have pointed out that the users logged as patient may be some "employment seeker" who are just look this website for jobs. Only we can get rid of these noisy data from patient user group, can we future investigate more topics of the patient group. Another study may be more focus on the navigation pattern of the different groups, like what is the path to find the same topic, are there any waste steps in the process to get the final information. Get to know that, we can redesign and refine the website for more convenient use.

References

1. Ayantunde, A.A., Welch, N.T., Parsons, S.L.: A survey of patient satisfaction and use of the Internet for health information. Int. J. Clin. Pract. 61(3), 458–462 (2007)
2. Trotter, M.I., Morgan, D.W.: Patients' use of the Internet for health related matters: a study of Internet usage in 2000 and 2006. Health Informatics J. 14(3), 175–181 (2008)
3. Eysenbach, G., Köhler, C.: How Do Consumers Search For And Appraise Health Information On The World Wide Web? Qualitative Study Using Focus Groups, Usability Tests, And In-Depth Interviews. BMJ: British Medical Journal 324(7337), 573–577 (2002)
4. Eysenbach, G.: The Impact of the Internet on Cancer Outcomes. CA Cancer J. Clin. 53(6), 356–371 (2003)
5. Fox, S., Fallows, D.: Internet Health Resources. Internet & American Life Project (July 16, 2003) [cited August 30, 2003)]; report/survey], `http://www.pewinternet.org/`
6. Fox, S.: Health Information Online. PEW Internet & American Life Project (2005)
7. Rice, R.E.: Influences, usage, and outcomes of Internet health information searching: Multivariate results from the Pew surveys. International Journal of Medical Informatics 75(1), 8–28 (2006)
8. Chang, P., et al.: Are Google or Yahoo a good portal for getting quality healthcare web information? In: Proc. of AMIA Annu. Symp., 2006, p. 878 (2006)
9. Morita, T., et al.: A study of cancer information for cancer patients on the internet. Int. J. Clin. Oncol. 12(6), 440–447 (2007)
10. Wu, A.S., et al.: Evaluation of Negation and Uncertainty Detection and its Impact on Precision and Recall in Search. J. Digit Imaging (2009)
11. Lambert, S.D., Loiselle, C.G.: Health information seeking behavior. Qual. Health Res. 17(8), 1006–1019 (2007)
12. Chen, E.S., Cimino, J.J.: Automated discovery of patient-specific clinician information needs using clinical information system log files. In: Proc. of AMIA Annu. Symp., 2003, pp. 145–149 (2003)
13. Graham, L., Tse, T., Keselman, A.: Exploring user navigation during online health information seeking. In: Proc. of AMIA Annu. Symp., 2006, pp. 299–303 (2006)
14. Rozic-Hristovsk, A., Hristovski, D., Todorovski, L.: Users' information-seeking behavior on a medical library Website. J. Med. Libr. Assoc. 90(2), 210–217 (2002)
15. Leydon, G.M., et al.: Cancer patients' information needs and information seeking behaviour: in depth interview study. BMJ 320(7239), 909–913 (2000)
16. Gonzalez-Gonzalez, A.I., et al.: Information needs and information-seeking behavior of primary care physicians. Ann. Fam. Med. 5(4), 345–352 (2007)
17. Health On the Net Foundation, `http://www.hon.ch/`
18. Clarian Health, `http://www.clarian.org`
19. Agrawal, R., et al.: Mining association rules between sets of items in large databases. In: Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, pp. 207–216. ACM, Washington, D.C (1993)

Double Visual Feedback in the Rehabilitation of Upper Limb

Liu Enchen[1,2], Sui Jianfeng[1], Christopher Young[1], and Ji Linhong[1]

[1] Division of Intelligent and Biomechanical System, State Key Laboratory of Tribology, Tsinghua University, 100084 Beijing, PRC
[2] Department of Mechanical Engineering, Guangxi University of Technology, 545006 Liuzhou, PRC

Abstract. Stroke patients may have limb paralysis or disturbance of consciousness. In the rehabilitation system, the main task of human-computer interaction focused on visual interaction. In this study, it has been confirmed that the visual feedback system in the rehabilitation has played an important role. This robotic system combines a dual video to meet training needs. In the experiment, trainees trains in accordance with the planned trajectory. Through the combined visual system, motion analysis feeds back to trainees in real time. iIn comparing to traditional training without feedback, the combination of Visual feedback can enhance the trainee's consciousness of balance, train the neural sensitivity of the brain and nerve system, and can restore the movement ability.

Keywords: Visual Feedback, consciousness, rehabilitation.

1 Introduction

Patients caused by cerebrovascular disease may have barrier in language, cognition, intelligence and other respects. Various characteristics of stroke patients such as disturbance of consciousness and limb paralysis should be considered when using robotics system rehabilitation therapy. In addition, human-computer interaction is proved to be essential to meet the needs of patients.

The main clinical manifestation of hemiplegic patients is that their movement is limited and disordered. Studies have indicated that there are three possibilities to make movement discontinuity: signals disturbance generated in the feedback path, the forward path, or the bypassed forward path during the exercise. The purpose of using neurological rehabilitation robots is to help nervous system to recover, because patients' nervous systems are constrained. Central nervous system has function of plasticity, so it can be recovered by movement stimulation and appropriate visual feedback form. Strengthening visual feedback with treatment progress can speed up the process of rehabilitation.

Hemiplegic rehabilitation robot can generate movement therapy as a means of rehabilitation to stimulate nervous system, using active, passive, resistance and assistance training modes to help patients in rehabilitation training. Human-computer

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 384–388, 2011.

interaction helps patients perceive environment that. The interaction between patients and robot provides patients with a much more appropriate training methods and robot learn simultaneously with feedback on patients' conditions. Human-computer interaction means robot can accurately convey critical information to patients, while patients' information can be input to robot simultaneously, which in result help achieving better therapeutic effect. It's necessary to design human-computer interaction system according to physiological and psychological characteristics of patients to accomplish above goals, including information systems interaction, feedback interaction, and visual interaction.

2 Method

Rehabilitation robot can help patients understand and analysis movement and stimulate their corresponding region of the brain and nerve. Using redundancy in the brain, robot help patients gradually repair and reconstruct the corresponding functional areas to achieve the rehabilitation objectives. Since with neurologic injury, patients receive loss in sense of touch. Visual interaction plays an important role in the rehabilitation process. During rehabilitation, they make communication with robot depending on the vision mainly. Using two sets of visual display feedback system meets the need of visual interaction. Trajectory of the system displays exercises that are stored in the system previously. According to the degree of recovery of patients, robot changes position and angle to help patients do more advanced training. The system provides feedback to patients in time to achieve interaction. The changes in the environment can cause changes in excitatory of membrane, thereby improving the training effect. Visual feedback system has a significant role in nerve repair compared with no visual feedback system.

2.1 Composition of the Visual System

Implementation of visual interaction uses two sets of visual display feedback system. Based on ergonomic size and site of injury in patients, visual feedback system is made up of two sets of display interface with different location and size, which can be carried out with different orientations and angles adjustment. Thus trainees can get quantitative visual feedback at any time to achieve accurate and accessible human-computer interaction in training.

2.2 The Display Part of the Visual System

The trajectory of patients is determined based on the maximum range of motion, which is displayed on the screen of the visual feedback system. When trainees are in training, motion information as input factors conveys some relevant factors to the robot, which adjusts training model for trainees. Then robot compares the actual trajectory of patient with target trajectories which are previously stored in the memory, and gap is fed back to patient through screens of the visual feedback system. Then patients adjust their direction of motion and force size according to tips. This process requires that robot constantly makes judgments in patients' sports and patients adjust their motion timely. Therefore, this process achieves a human-computer interaction.

Visual feedback system in human-computer interaction has easy and simple display interface and expresses concise and accurate information, so it can attract the attention of patients and cause nerve excitability. Rich colors stimulate nerve and maneuver patients' mood, so combination of various visual elements improves training effect.

In experiment, the target trajectory and feedback of the trainees' training information can provide trainees stimulation and challenges, so trainees can timely adjust direction and size of their force, rather than a simple random movement. In result, patients' movement become consistent while improving the limbs' motor skills. Visual feedback can deliver location information, to enable trainees to be active in the force balance on the effective control and improve the accuracy of balance.

2.3 Characteristics of the Visual System

Healthy human receive 83% of external information from the visual feedback. Limited by scope of their activities, hemiplegic patients gain greater percentage of information through visual feedback. So it is very significant to encourage visual feedback with treatment progress. For example, patients can access directly target trajectory information by visual feedback and learn deviation between actual trajectory and the target trajectory. Constant visual feedback stimulates patients to approach to target trajectory, making nervous system repaired or rebuilt.

Visual feedback system can record patients' performance in detail and collect data of each trainee's training progress. This makes each experiment comparable and unity, and helps establishing a standardized evaluation system. At the same time this system can save each patient's track and date of each training progress which could contributes to observing patients' rehabilitation status. Rehabilitation physicians adjust training time and intensity according to rehabilitation status.

3 Experiment

Eight patients(age: 55±5.25 years, time after stroke: four weeks) with hemiplegia training for this experiment. Two sets of Experimental data were collected as with visual feedback and with no visual feedback.

The first set of data (Figure 1,2) is the trainees for training in the environment of a straight line trajectory. The second set of control data (Figure 3,4) is the training for the environment of circular trajectory.

In the figure, the sampling interval is constant .The sampling points into the curve which is the path of movement. Straight line is connection of the sampling point and a fixed point, to visualize the number of sampling points. That is to say, the number of lines is equal to the number of sampling points. By the number of line segments and intensity can be seen: after the same length of track to spend at different times. Figures

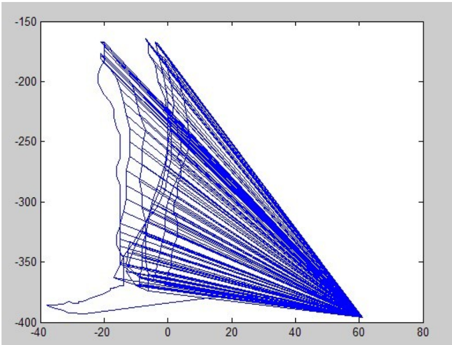


Fig. 1. With visual feedback, the training results of the linear motion

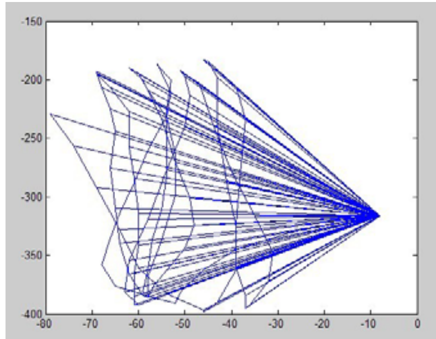


Fig. 2. Without visual feedback, the training results of the linear motion

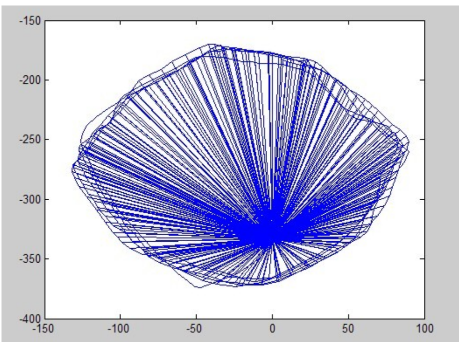


Fig. 3. With visual feedback, the training results of the circular motion

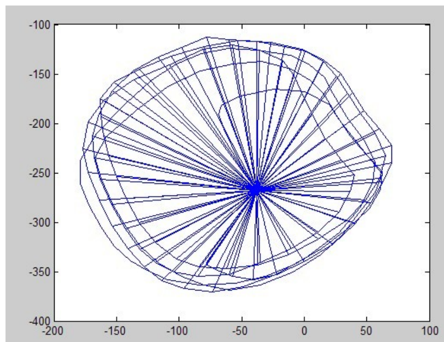


Fig. 4. Without visual feedback, the training results of the circular motion

also show that the trainees spend time to think of each movement, the size and direction of force was significantly different. With visual feedback, Trainee deliberately follows closely to the planning trajectory. Without visual feedback, the trainees apparently did not think after a long period, but the comfort of their own actions based on exercise, rather than close to the goal by overcoming difficulties.

According to the data, different trajectories of training have different effects to rehabilitation of patients' nerve system. As in example, human will avoid obstacles if they are in sight, which means the brains are stimulated during such progress. The complexity of path has different effects to trainees. Because different path conditions requires trainees make timely adjustments and change speed and magnitude of movement. Visual feedback reflects the patients' movement information at prompt, which encourages the patients to adjust their exercise constantly.

4 Summary and Discussion

This experiment indicate that the implementation of visual feedback of movement and nerve reconstruction. Trainees in the visual interaction of visual feedback based on the calculation of the brain and the corresponding nerve to muscle movements ordered the implementation, close to the target. Visual feedback can enhance the trainee's consciousness of balance .The whole process which from the brain to the Central nervous system to muscle is conducive to train the neural sensitivity and accuracy, and enhance muscle balance.

This is the preliminary results of our study. Further studies are needed on the angle of force and torque of trainees, and to identify an accurate assessment of the movement method.

References

[1] Cho, S., Ku, J., Han, K., Lee, H., Park, J., Kang, Y.J. In: Kim, Y., Kim, S.I (eds.) IEEE Virtual Reality Conference, VR 2009, pp. 283–284, Digital Object Identifier: 10.1109/VR, 4811056 (2009)

[2] Bagesteiro, L., Sarlegna, F., Sainburg, R.: Differential influence of vision and proprioception on control of movement distance. Experimental Brain Research 171, 358–370 (2006)

[3] Bridge, H., Thomas, O., Jbabdi, S., Cowey, A.: Changes in connectivity after visual cortical brain damage underlie altered visual function. Brain 131, 1433–1444 (2008)

[4] Danckert, J., Rossetti, Y.: Blindsight in action: what can the different sub-types of blindsight tell us about the control of visually guidedactions? Neurosci. Biobehav. Rev. 29, 1035–1046 (2005)

[5] Betker, A., Szturm, T., Moussavi, Z., Nett, C.: Video game-based exercises for balance rehabilitation: a sigle-subject design. Archives of Physical Medicine and Rehabilitation 87, 1141–1149 (2006)

[6] Sarlegna, F.R., Gauthier, G.M., Blouin, J.: Influence of feedback modality on sensorimotor adaptation: Contribution of visual, kinesthetic, and verbal cues. Journal of Motor Behavior 39, 247–258 (2007)

Can User Tagging Help Health Information Seekers?

Malika Mahoui[1], Josette Jones[1], Andrew Meyerhoff[1], and Syed Ahmed Toufeeq[2]

[1] Indiana University School of Informatics
535 West Michigan Street, Indianapolis, IN 46202, USA
[2] Department of Biomedical Informatics, Vanderbilt University Medical Center
2209 Garland Ave - Nashville, TN 37232, USA
{mmahoui,jofjones}@iupui.edu, ameyerho@umail.iu.edu,
s.toufeeq.ahmed@Vanderbilt.edu

Abstract. The Web is becoming a regular source of information for health information (HI) seekers. In the United States for example, several studies report that more than 80% of Internet users routinely utilize the Web to obtain medical and health information. While Google searching is by far the default point of access for HI seekers, using expert Web sites such as Medline+ and WebMD is also another alternative for users that are looking for quality content. In addition, health expert Web sites are supported with several advanced querying such as browsing using categories organized either as a hierarchy or in an A-Z search list. This mode of searching supported by expert generated taxonomies or metadata is particularly suitable for discovery purposes. In parallel social Web sites such as Delicious and StumbleUpon are providing Web users with a new opportunity to participate in the process of annotating Web resources. This is particularly true for health data as supported by the large amount of health resources annotated by Delicious for example. In this paper we investigate the relationship between expert health metadata and user annotations where the objective is to determine how they can be combined to provide HI seekers with better means to finding their HI needs. The results of the study show that while there is noticeable overlapping between these two types of data, user tags are clearly a new type of metadata that will play an important role in supporting HI seekers searching/browsing health resources.

Keywords: Information Retrieval, User Tags, Expert Metadata, Social Networks.

1 Introduction

The Web encompasses a wealth of information related to health; and health information (HI) consumers are leveraging this source of information for various reasons. In the United States for example, several studies report that more than 80% of Internet users routinely utilize the Web to obtain medical and health information [1, 2]. Search queries such as looking for the side effects of a drug that the doctor prescribed, looking for information on how to better manage the chronic disease of

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 389–397, 2011.

the spouse, or simply looking for more information about a new advertised weight loss technique are examples of such diversified user information needs. Broadly, they are two channels used by HI consumers for information searching: using expert Web sites such as Medline+ and WebMD and using Google search. Expert health Web sites provide HI consumers with a rich taxonomy supporting their browsing interface. This type of information searching has been reported to be very adequate for discovery purposes. On the other hand, searching using Google allows Web users to target a larger search information space and it is more suited for targeted queries. During the last decade, social networks are increasing in popularity among Web users. In particular, Web sites such as Delicious, StumbleUpon, and Digg are collecting a wealth of user annotations (i.e. tags) that users are voluntarily providing for resources they access. These tags are used to support social networks' Web sites users in their search in addition of using keyword search. The Health topic is one popular topic shared by social network Web sites' users. For example, a search of the tag "health" on Delicious social delivered 1,422,217 health-related documents in October 2010. This is an increase of more than one million resources since the same search performed in February 2010. This study investigates the relationship between metadata generated by experts and user tags where the long term objective is to propose a Web environment where both types of annotations are combined to provide better search capabilities for HI consumers. To conduct this comparative study we consider Meldine+ metadata because of its richness and simplicity to conduct studies with. We use the metadata as seed queries to search Delicious social Web site utilizing their data feeds APIs. A cleaning phase is deployed on the results tag sets before the comparative study is performed. We use PSA (Positive Scoring Agreement) approach [7] to compute the overlapping between metadata and user tags. The results we obtained provide us with more insights on how to better combine these two types of annotations to improve the search of health data.

The rest of the paper is organized as follows. Related research is described in section 2. In section 3 we explain the methodology we adopted. Results' presentation and analysis are detailed in section 4. In section 5 we further discuss the ramifications of the results of the study and outline some of the future work.

2 Related Work

Social network Web sites such as Delicious, Flickr and StumbleUpon are becoming very popular and include a wealth of tags provided by users in various domains. PatientsLikeMe is one of the few social networks in the domain of health. Patients are part of sub-communities centering about diagnoses such as Amyotrophic Lateral Sclerosis, Multiple Sclerosis and Parkinson's Diseases; and community members describe their symptoms using natural language and therefore can be considered as user tags. Few studies have investigated the impact of Web 2.0 for HI seeking. Exceptions include studies such as [3-6].

The authors in [3] studied the tags produced by PatientLikeMe users describing topics related to their health. They compared this type of user data against controlled vocabularies such as UMLS and SNOMED CT. The results of the study showed that 43% of the user terms are either exact terms, synonyms or terms from UMLS where

93% of the terms are from SNOMED CT. These results highlight the fact that even within a relatively health aware and connected community more than half of the terms they use to describe subjects such as symptoms and treatment do not overlap with the medical taxonomies such as UMLS.

The study described in [4] worked with two social bookmarking systems, Connotea and CiteUlike, where the tagging population tend to be from Academia. The study focused on a subset of pubmed documents tagged from these social sites and used metrics such as coverage of the document space, the number of metadata terms associated with each document, rates of inter-annotator agreement, and rates of agreement with MeSH terms. Some of the results reported indicate that users tend to provide few tags when tagging documents, and that there is a low agreement between users tagging the same resources (inter-annotator agreement), as well as between user tags and mesh terms.

Similar results are reported in the studies conducted in [5, 6] using general health documents. In [5], the authors focused on studying the overlapping between the content of an expert indexed health digital library in France (CISMeF) and Delicious social network; and used metrics such as the overlapping documents between the two collections and the overlapping of user tags with expert generated descriptors. First, the results show a very low overlapping (113 documents) of CISMeF and Delicious. To perform the manual qualitative study of the documents common to both datasets, they collected for each resource the descriptor and its thesauri relation (synonyms, related terms, and broader terms) as well as related terms that are not official, such as descriptor but still associated to the resource. The results of the qualitative study show that more than half of user tags are present in the thesauri category with more than 85% of them being with exact terms or broader terms. Among the 34% of terms that are not related to the expert indexed terms, the manual analysis shows that 68% are terms that are related to the content of the document and therefore can be used to index the documents. In summary, despite that the study was conducted on a small collection of documents, it reveals that user tags can compete with thesauri metadata.

In the study we conducted in [6], we adopted a more automatic approach. We started from a top 100 trusted health Web sites provided by the Consumer and Patient Health Information Section (CAPHIS)[1] of the Medical Library Association Web site[2]; and we extracted the topic-oriented categories these health Web sites provide for users to find their information needs. This is often presented to users as an A-Z index terms. We call these terms expert terms. We retrieved the health resources associated to these expert terms from the Web sites and we used the URLS to query Delicious. We also automatically extracted keywords from the resources using AlchemyAPI™ [2]. The metrics we used for the analysis include the overlapping of health oriented Web sites content and Delicious content as well as the overlapping between user terms and resource keywords in another hand. Similarly to the previous study, the overlapping between Delicious content and more traditional Web site is low - 93 resources for this case. The overlapping of words within this common collection was moderately small measured as 0.37 using the (0..1) specific positive agreement PSA

---

1 http://caphis.mlanet.org/

2 Only those sites who allow crawling allowed crawling, as indicated by each of the site's *robot.txt* file we used.

measure [7]. However, the PSA value was much smaller (0.1) when comparing the automatically extracted keywords from the content pointed by the URLs to user tags. This analysis while it needs refinement such as using the thesauri information, clearly reveal the complementary goal that users can play in annotating resources to be used for browsing/indexing especially when compared to automatically extracted keywords.

In this paper we address the main limitation of the last two studies that worked with health data, namely the small size of the collection of health resources used for the studies. The aim is to provide solid arguments when describing the relationship between user tags and expert generated metadata. The ultimate goal will be to find suitable methods to combine these two types of metadata. To be able to conduct the study on a large collection, we started from the unanimous fact that there is a little overlapping between resources available in health focused Web sites such as WebMed and resource's such as social networks. Therefore the approach we adopted is to use the expert metadata provided by one of the health focused Web sites as a seed to probe for resources on social Web sites such as Delicious. We choose to use the expert metadata provided by Medline+ because of its comprehensiveness and structure offering for each expert terms links to terms at different granularity levels (synonyms, boarder, etc.). Using the collected data, some of the research questions we aimed at addressing are:

- How much overlapping exists between user tags and expert metadata?
- What is the granularity of the user tags when compared to expert tags (exact, synonyms, broader, formal (mesh terms)?
- How much overlapping exists between user tags and at what granularity level?

In the remaining of the paper we describe the methodology we adopted to conduct our study and analyze the results we obtained.

3 Data Collection and Methods

3.1 Data Sets Collection

The study was conducted using two data sets: Medline+ expert metadata set and Delicious annotations set. Medline+ is a Web site that offers its users a wealth of health related resources. To help its users find their information needs the Web site offers browsing support using expert categories. The categories are organized in a taxonomy structure where the top level of the taxonomy is composed of terms called *groups (e.g. digestive systems)* to which standard health terms called medical health topics are linked. Each medical topic has a set of synonyms and mesh terms associated to it (if available). This metadata and its structural taxonomy are available in an xml format (http://www.nlm.nih.gov/medlineplus/xml/vocabulary/mplus_vocab_2011-02-26.xml). We used this source to extract only the metadata that is in English as the XML resource also includes metadata in Spanish. Next, we refer to each term of this data set as expert health topics (EHT) to contrast it to Delicious tags.

To collect Delicious user tags we deployed EHT as seed terms or keywords to query Delicious in order to secure a sufficient data set to conduct the study (see previous section for rationale). We used the data feed API (http://Delicious.com/help/json/) of Delicious to query the social Web site. We collected up to 400 annotations for each EHT. Note that an EHT can be a medical health topic (e.g. stomach ache), a mesh term, a synonym, or a group name.

An annotation is described mainly by the following features:

- A resource with its URL
- The id of the user annotating the resource
- The set of tags that he/she entered
- The time the resource was annotated

We also keep track of the time the resource was download for potential future use. The order in which the tags have been entered is also provided. Next we refer to user tags as user health tags (UHT).

3.2 Data Preprocessing

As described in previous sections, while user tags have the potential to be a resourceful source of keywords to be used for improving searching of health resources, the data is very noisy. Some of the reasons include the fact that (1) the annotations are performed at a voluntary basis which does impact the rigor aspect of the process, (2) the lack of a standard or a set of guidelines to help users provide useful tags for their own re-use but also to be shared by others, (3) the presence of spam activity, (4) and the limitations required by some social Web sites in terms of the format of entering a tag. In the context of the last factor for example, the definition of a tag in Delicious is restricted to one word as opposed to a key-phrase. As a consequence, users often find themselves concatenating several words in one tag (e.g. drugAddiction). Note that progressively, several Web sites supporting tagging activity are now allowing users to enter tags composed of more than one word.

We followed a multi-step process as follows to eliminate as much noise as we can from the data and to normalize the data before its analysis:

1. Detect spam activity and discard associated annotations
2. Perform a light manual cleaning of the data to correct obvious typo errors
3. Eliminate punctuation and stop words such as "the"
4. Eliminate duplicate tags
5. Perform stemming

To detect spam activity several heuristics can be employed (see [8, 9] for proposed approaches). For the purpose of this study we choose to employ a simple heuristic which consists of (1) eliminating annotations that are associated to users with a number of annotations above a threshold value and then (2) manually check the activity of the users detected in the first step to reduce the number of false positives.

In an effort to reduce the impact of typo errors and error's related to mistakes in entering or collecting the data, we performed a light manual correction of the data. This included replacing hyphen by space, and splitting merged words like "IrritableBowelSyndrome".

3.3 Data Analysis

We performed two types of analysis: descriptive analysis and term comparison. To compare the overlapping between expert health topics EHT and user health tags UHT we used the Positive Specific Agreement (PSA) measure. Given two data sets of terms S_1 and S_2, PSA(S_1,S_2) is given by the following formulas:

$$PSA(S_1,S_2) = 2a/(2a+b+c) \qquad (1)$$

Where:

a = Number of common terms between S_1 and S_2
b = Number of distinct elements of S_1
c = Number of distinct elements of S_2

PSA value ranges between 0 and 1; where a value “1” corresponds to a maximum overlapping.

To normalize the computation of the PSA we used the word as the unit of overlapping between data sets, as UHTs are composed of one word in the case of Delicious tags, and EHT may encompass more than one word. We compared UHTs and EHTs at different level of granularity to expand EHTs with, synonyms, mesh terms and group names. More precisely, in the first set of comparison we divided the EHTs into 4 groups depending on whether the term is the original medical term (hereafter referenced as medical topic), a group name, a mesh term or a synonym of the medical topic; and we computed the overlapping of each subset with the EHTs with UHTs.

In the second set of comparison, we replaced each EHT with related terms in three ways: with related synonyms, related group names, and related mesh terms. More precisely, if the EHT is a medical topic such as stomachache, then its replacement with each type of related term is straightforward. In the case where an EHT used for querying Delicious is not a medical topic but a related term itself such as digestive system, then the EHT term is first linked back to the original medical topic; and then using the medical topic the replacement by related terms is performed is a similar way as in the former case. Note that we include in the list of related terms the original medical topic used for the replacement but not the EHT used for querying.

4 Results and Analysis

Using the Medline+ XML file downloaded in fall 2010, we had a set of 2649 EHTs. This metadata was deployed as seed queries for searching Delicious to return up to 400 annotations per EHT. Only 2087 out of the initial 2649 EHTs returned results. On average 222.8 annotations per EHT were collected.

In total, we had 464944 annotations associated to 96700 unique resources provided by 44732 unique users. The maximum number of annotations associated to a user is 1728 with an average of 10.4 annotations per user.

The filtering of the spam activity resulted into discarding annotations from 47 users, adjusting the average annotations to 9.85 per users. This resulted into a drop to

440540 annotations to be considered for next steps associated to 91959 unique resources.

The average number of annotations per resource before and after filtering the spam activity is 4.80 and 4.79 respectively. Similarly the maximum number of annotations associated to a resource before and after the spam activity is 400.

Table 1 breaks down the number of EHTs and the number of annotations before and after the preprocessing (including the manual cleaning) per type of EHT: medical topic, group name, mesh term, and synonym. Note that an EHT can be in more than one type of EHT. This is the case for the medical topic "abdominal pain" also found as a mesh term.

Table 1. Distribution of the number of annotations before/ after preprocessing per type of EHT

	EHT all	EHT medical topic	EHT group term	EHT mesh term	EHT synonym
# EHTs before preprocessing	2649	859	44	1519	696
# EHTs after preprocessing	2049	751	33	1167	563
#annotations before preprocessing	464950	195913	7264	253737	129796
# annotations After preprocessing	440073	184846	6852	240373	121 983

The number of unique UHTs collected from the annotation data set was 65917 with an average of 6.3 UHTs per annotation, and a maximum of 494 UHTs per annotation. In table 2 we present the top frequent UHTs after removing the stop words like (of, for, and, etc.) Note that, if a UHT term such as disease has variations like its plural (diseases), it is not cumulated in the frequencies shown in table 2.

After the second step of the preprocessing that applies the light manual cleaning of the data, the number of UHTs was reduced to 63041. This pre-processing task did not affect the top 20 frequent UHT terms but did increase their frequency.

Table 2. Top twenty frequent UHTs

UHT	Frequency	UHT	Frequency
health	67590	food	9234
cancer	19865	pain	8877
treatment	19214	science	8866
medicine	17836	nutrition	8498
disease	16704	safety	8277
medical	14613	diet	7813
care	12134	heart	7752
surgery	11547	syndrome	7500
symptoms	10428	brain	7177
research	10121	disorders	6993

Table 3 summarizes the PSA results when computing the overlapping of UHTs with each EHT type (medical topic, group name, mesh term and synonym).

Table 3. PSA results based on the type of expert health terms (EHT)

	EHT all	Medical topic EHTs	Group name EHTs	Mesh term EHTs	Synonym EHTs
PSA	0.50	0.51	0.41	0.50	0.48

Table 4 summarizes the PSA results when comparing UHTs with each type of related terms.

Table 4. PSA results when expanding EHTs with related terms

	EHT all	EHT replaced with related medical topics	EHT replaced with related group names	EHT replaced with related mesh terms	EHT expanded with related synonyms
PSA	0.50	0.28	0.32	0.37	0.36

These results obtained using such a large dataset are in line with the results we obtained in the small study reported in [6], and with the results conducted in [5] with a manual computation of the overlapping between UHTs and EHTs. More precisely the results indicate that there is a noticeable overlapping between expert terms and user tags varying from 41% to 50%. Furthermore, the overlapping is more substantial when the expert terms are medical terms and mesh terms. The results summarized in table 4 suggest that replacing the term with related terms need to be carefully designed when used to improve the recall of a search system. More precisely, these results indicate that users tend to provide one type of tags to describe a resource as opposed to using several related terms in the annotation. In addition, the top 20 frequent user terms provide more support for using broad topics such as “disease”, “symptoms” and “treatment” as a mean to help users browse through a collection of health resources. Among the noticeable most frequent tags, the tag “research” suggests that the taggers population is perhaps more related to the scientific community.

5 Discussion

In this paper we described a study we conducted on a large dataset to assess the potential of utilizing user tags to support expert metadata when building systems for searching/browsing health resources. The results have several implications. First, there is a noticeable overlapping between user tags and expert metadata and this overlapping is more substantial with medical topics used by health expert websites such as Medline+ and with mesh terms. This observation suggests conducting more studies to assess the potential of deploying mesh terms more actively when building health related browsing taxonomies as users seem to identify more with this category of expert data. The potential of using user tags as a replacement of medical topics presents the advantage of having a “cheap” approach to determining these topics and these topics are more dynamic reflecting the continuous shifting of HI interests. Second, the fact that at least half of user terms are not overlapping with expert terms and given the observations made in [5], there is an evident need to leverage user tags when building health related browsing taxonomies; as users are describing health resources with more user specific terms not necessarily expressed by experts metadata. For future work we plan to conduct more analysis of the data to evaluate the specificity of user tags vis-a-vis expert metadata.

Acknowledgment. Thanks to Rania Nafiseh from the School of Informatics, Indiana University, for her help in the manual cleaning of user tags.

References

1. Interactive., H. The Harris Poll ®#76, July 31, 2007 Harris Poll Shows Number of Cyberchondriacs – Adults Who Have Ever Gone Online for Health Information– Increases to an Estimated 160 Million Nationwide (2007), http://www.harrisinteractive.com/harris_poll/index.asp?PID=792
2. Cybercitizen Health^{TM} v8.0. Over 60 Million U.S. Adults Engaging in Health 2.0. Manhattan Research reports that the number of Health 2.0 consumers has doubled since 2007 (2008), http://www.manhattanresearch.com/newsroom/Press_Releases/over-60-million-health-20-consumers.aspx
3. Smith, C.A., Wicks, P.J.: PatientsLikeMe: Consumer Health Vocabulary as a Folksonomy. In: Proceedings of AMIA Annual Symposium, pp. 682–686 (2008)
4. Good, B., Tennis, J., Wilkinson, M.: Social tagging in the life sciences: characterizing a new metadata resource for bioinformatics. BMC Bioinformatics 10(1), 313 (2009)
5. Durieuxa, V., Kerdelhué, G.: Looking for Health Information on the Internet: Can Social Bookmarking Systems Replace Expert Gateways. In: The European Association for Health Information and Libraries, EAHIL Workshop, 2009, Dublin Castle, Ireland (2009)
6. Mahoui, M., et al.: Can Collaborative Tagging Improve the Searching of Health Web Sites? In: Proceedings of the AMIA 2010 Annual Symposium (2010)
7. Hripcsak, G., Rothschild, A.S.: Agreement, the f-measure, and reliability in information retrieval. Journal of the American Medical Informatics Association: JAMIA 12(3), 296–298 (2005)
8. Liu, K., Fang, B., Zhang, Y.: Detecting tag spam in social tagging systems with collaborative knowledge. IEEE Press, Piscataway (2009)
9. Koutrika, G., et al.: Combating spam in tagging systems: An evaluation. ACM Transactions on the Web (TWEB) 2, 22:1–22:34-22:1–22:34 (2008)

Interactive Medical Volume Visualizations for Surgical Online Applications

Konrad Mühler, Mathias Neugebauer, and Bernhard Preim

Department of Simulation and Graphics, University of Magdeburg, Germany
muehler@gmail.com

Abstract. For surgical planning, the exploration of 3d visualizations and 2d slice views is essential. We introduce a new technique to bring visualizations of 3d scenes of any kind and 2d image stacks generated by CT or MRI into the web. We pre-render 3d scenes from different perspectives and show these images during user interaction in a Flash based framework. Therewith we can create a 3d spatial relation. We support different zoom levels and can provide different presets for visualizations. The new technique is used to provide complex and enhanced 3d visualizations for surgical therapy planning and educational purposes as well as for evaluation of new visualization techniques.

1 Introduction

The exploration of 3d visualizations plays a growing role in surgical planning and its training, since they provide a good spatial impression and a three-dimensional overview of complex organs and pathologies. 3d visualizations are used to assess infiltrations of anatomic structures, viable access paths and security margins. Especially for difficult cases, anatomical structures and pathologies are segmented to provide more quantitative information like distances and volumes and to separate different tissues that can be hardly distinguished due to similar values in the 2D slices.

For clinical routine there are many applications available to support the operation planning process, especially radiology workstations and applications running on the medical doctor's local computer. Due to the demand for higher flexibility and better integration in the hospital's IT infrastructure, web-based applications come more and more to the fore. Web-based applications can be used from nearly every computer. For education web-based lectures, e-books and multimedia material are widely used. Even if many applications can be transferred into the web by now, advanced 3d visualization and extended 2d slice views are still neglected. This is caused by the still very high demands of 3d visualizations on the underlying computer and graphics power.

To provide interactive 3d visualization to the user, a common approach is to adapt and re-implement the underlying visualization techniques for an online available browser plug-in like Java or VRML. A main drawback of these strategies is their limitation to single visualization techniques and the costs of the reimplementation for each new technique. Furthermore, volume visualization techniques, that are widely used by medical doctors, are mostly hardware accelerated and therewith not directly useable in an online environment (see Fig. 1 and 2).

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 398–405, 2011.

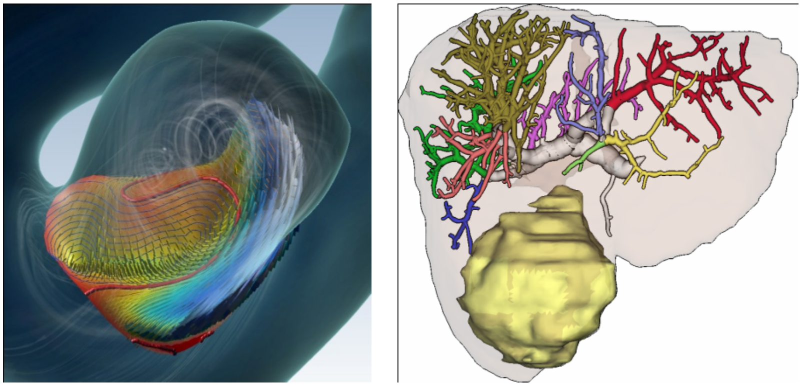


Fig. 1. Left: A 3d volume rendering visualization of an aneurysm with streamlines showing the inner vessel blood flow. Right: A 3d visualization of a liver with vascular structures and a tumor. The vascular structures are enhanced by illustrative rendering techniques.

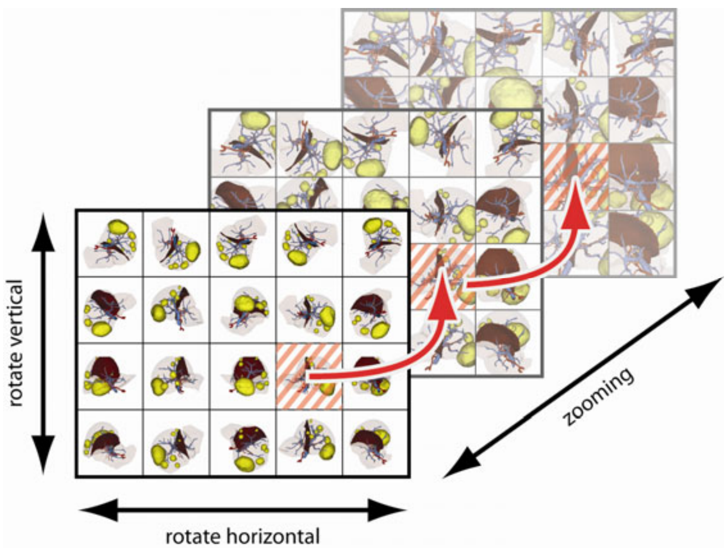


Fig. 2. The images are stored in different matrixes. Rotating the scene, the images from one distinct matrix are shown. For other interactions like zooming another matrix is used and the image that corresponds to the current viewpoint is loaded from that matrix.

We developed a new technique based on pre-rendered images of a 3d scene. Navigating through our images, the user gets a 3d impression of the scene and its objects. Our new format contains different extensions that are important for our medical users to get a better insight in the shown structures. We enable different levels of zooming and the change of the visualization within certain constraints: The user can show and hide important structures and load different states of a visualization. Since we use pre-rendered scenes, we can provide any imaginable visualization technique in our medical web applications. Therefore, we can also show 2d slice stacks for example from CT or MRI scans with colored overlays of segmented structures or textual annotations. The drawbacks of the technique are their lower flexibility with respect to visualization changes (like color changes or structure disabling) - this drawback can be overcome with careful and profound considerations during the pre-rendering process. We will discuss this in detail later.

2 Related Work

There are many formats and techniques to provide three dimensional scenes and objects interactively in the web. Among them are the popular VRML (Virtual Reality Modeling Language) and its successor X3D [1]. Other solutions, like Acrobat3D [2] or Google's O3D [3] are emerging. To show and interact with scenes based on these techniques, web browser plugins are needed. Even if, e.g., X3D is fully implemented in the newest browser versions as part of HTML5 (e.g., in Firefox 4 beta 9), the scope of current 3d implementation for the web is still very narrow.

Another big problem of common techniques is that they are limited to rendering polygonal scenes. Many applications like walk-through games or exploration of 3d mechanical models fits this constraints. But special rendering techniques like direct Volumerendering (e.g., [4]), streamline visualizations (e.g., [5]) or illustrative techniques (e.g., [6]) are not realized yet. There are some approaches, which tried to integrate volume rendering in to VRML [7]. This technique, however only provides a very low functionality with basic rendering algorithms.

To provide 3d impressions of products in online shops, the Quicktime VR Object Movie format was developed by Apple [8]. The 3d impression of single objects was imitated by a set of single photos that were taken from different perspectives. Navigating with the mouse in the images, the corresponding images to the current camera position are shown. Depending on the discretization of the images the user can get a real three dimensional impression of the object. The constraints of the approach are the limited degrees of freedom: The camera cannot be freely moved. Real zooming is not provided as well as the scene itself cannot be changed (e.g. a change of the color of the object or the blending of objects). A closely related technique is the Quicktime VR format. It is used to show full panoramic views based on a set of photos taken from within a virtual sphere.

3 New Datastructure for 3d Images and 2d Slices

Instead of adapting existing plugins, like VRML, we chose an approach that is independent from the used visualization technique and refer to it as Object Movies. Object Movies were primarily developed and used to provide 3d impressions from real objects like products in an online shop. Several images are captured from camera positions on a surrounding sphere around the object. These images are arranged in a 2d matrix, where rows represent a set of horizontal neighboring images. The user navigates in this image matrix by dragging the mouse and gets a 3d impression of the object. Current Object Movies are restricted to two degrees of freedom (DOF). The two DOF do not have to be necessarily the two rotation axes. Höhne et al. [9] used in their Voxel Man application one DOF to show different layers of anatomical structures (skin, muscles, vessels, bones). Those visualizations could only be rotated around one axis.

In our online environments, the user may explore several 3d visualizations of segmented anatomical structures. Those visualizations support the spatial impression and illustrate different surgical strategies, e.g., resection planes in the liver surgery or the blood flow in a vessel. Exploring those scenes, the user wants to zoom in/out, rotate the scene, enable/disable different structures and change the appearance of the visualization. Therefore, we enhanced the idea of Object Movies with more degrees of freedom. We use multiple 2d matrixes of images to provide additional facilities to the users. If the user rotates the scene, he navigates in one of these 2d matrixes (see Fig. 3). If he zooms in, another matrix is used and the corresponding image in this matrix is shown to the user.

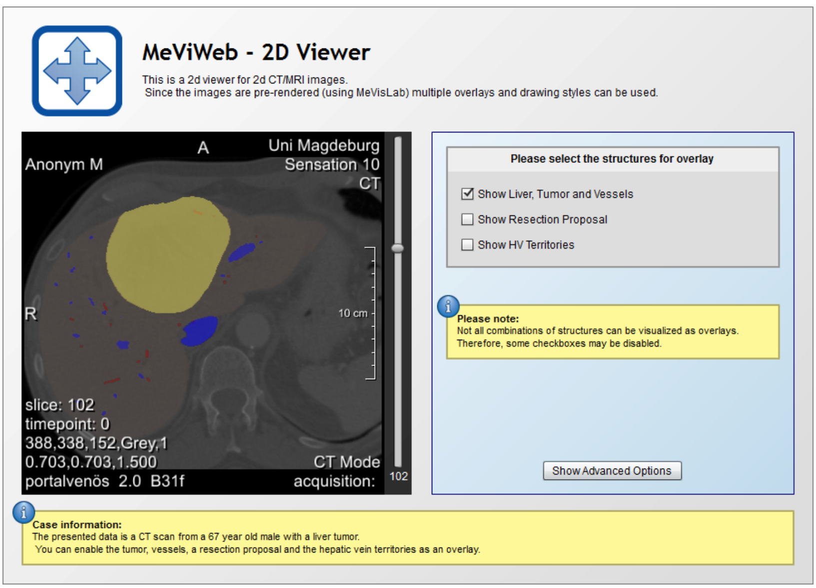


Fig. 3. This application presents a stack of slices to the user. The slices are enhanced with additional overlays. The user can enable the liver, tumor and vessels in combination with a resection proposal and the liver territories of the hepatic vein.

Depending on the angle between the different camera positions that were used for the images, the user gets a pretty seamless 3d impression of the presented scene. Our empiricism showed that an angle of about 18 degrees is appropriate. That equates 20 images horizontally and 10 images vertically. The images are pre-rendered and stored in a simple data format that enables other developers to create their own object movies.

We use FLASH to design and create the front ends of our application. FLASH is widely used plug-in and installed on many systems. Therefore, our FLASH applications can be embedded in larger websites or platforms, as we did, e.g., in an education platform for liver surgeons. Using FLASH, we can also provide our visualizations on mobile systems like Android tablets and smartphones.

3.1 2d Slice Visualizations

As we mentioned above, we are completely free in choosing a rendering or visualization technique. Hence we can use our technique also to provide advanced slice visualizations to the user. For educational purposes, our 2d slices are enhanced by colored overlay of segmented structures. These structures can be for example vital vessels, tumors or organs as context information. In addition, we add textual annotations to the slices, e.g., names of structures as well as additional information like volumes or distances between structures of risk. The user can scroll through the slices and interpret the presented information (see Fig. 4).

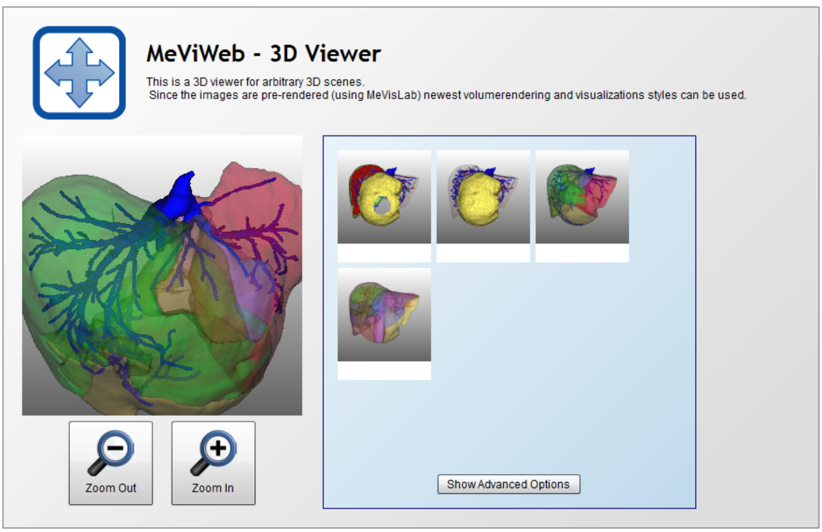


Fig. 4. In the exploration part of a larger surgical training system for liver surgeons, the user can explore the 3d scene. By loading a sequence of pre-rendered images from the underlying image matrix, an illusion of rotating and zooming the scene is created.

In comparison to the 3d version, we need one dimension less since we only scroll in one dimension. But we can use the second dimension for a specialty of tomography images: In normal radiologic workstations, the user can change the windowing of a slice. We can provide several different window settings as the second dimension. Those window settings (transfer function for gray values) can be automatically derived for CT scans, where the values for each type of structure (Hounsfield Units) are well known.

3.2 Capturing of Images

The still images of the scenes are captured in a pre-rendering step before the visualization is transferred in the online system. This process runs completely automatically and needs as an input only the visualization that should be shown. The images for different zoom levels are generated automatically. Since the user of the 3d visualization later can only see visualizations that were created before, it must be profoundly considered, what different visualizations are generated. In online systems based on polygonal meshes the rendering is performed in real-time online. The user can enable or disable different objects and gets the desired results immediately. Using pre-rendered images, this is not possible anymore. Theoretically, all possible permutations of objects can be pre-rendered and shown on demand - but this is very time- and memory consuming and thus not preferred. We found, that only a few sets of combinations of structures are of interest for the user of a special application. For example, liver surgeons often only inspect the liver in combination with one of the three main vascular systems of portal vein, hepatic vein or hepatic artery. Therefore, visualizations were, e.g., the hepatic vein is shown together with the portal vein, can be neglected and must not be pre-rendered. Taking many of these considerations into account results in a huge reduction of visualizations. Finally, e.g., for our learning environment for liver surgeons only 5-10 visualizations must be generated.

3.3 Performance

The basis of our technique are huge stacks of images. For a set of five visualizations with three zoom levels, 20 images per row and 10 per column as mentioned above, we have 3000 single images. Depending on the image resolution and compression results this in a data volume between 50MB and 90MB. Transferring this amount of data can be still a problem in the context of nowadays capacities. Therefore, we only transfer images which are currently needed. The user will never use every view in every zoom level. We pre-load those images in the background of the application that are in the nearer surrounding of the current view.

4 Application Areas

Currently, we use the interactive Object Movies for different tasks: In a surgical **training system**, the students can study different cases from the area of abdominal surgery. Each case consists of information like anamnesis, tomography image data, videos and several 3d visualizations. The students can explore those visualizations (see Fig. 5), zoom in and change between different views which illustrate important aspects (different vascular systems and segmented areas). In collaboration with experienced surgical educators, for each case the necessary visualizations (shown structures etc.) were determined.

In addition, we use the presented technique to **discuss new visualization techniques** with medical partners over long distances. Since our Object Movies are independent from the visualization technique, we can present newest visualizations based on currently developed volume visualization techniques. Since new techniques often reveal their potential during the 3d exploration, we overcome the presentation of

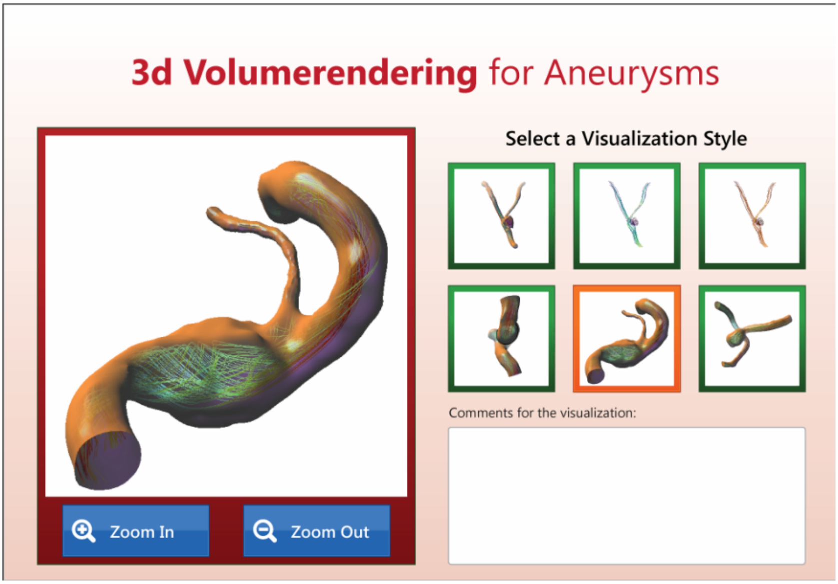


Fig. 5. This online application is used to discuss new visualization techniques. The user can select different visualization proposals from the right, explore them and give feedback for each technique.

single screenshots or playback of movies. In Figure 6 the interface of a FLASH application is shown, where the user can interactively explore 3d models of vascular aneurysms with simulated blood flow. At every viewpoint, he can change the current visualization technique or zoom in the visualization. For each visualization feedback can be given. Since the flexible FLASH frontend we can also provide complex evaluation questionnaires in combination with the 3d visualizations. This moves our evaluation processes for many new techniques forward since we can receive more feedback from our widespread users based on real 3d impressions.

5 Conclusion and Future Work

We presented a new technique to provide 3d and 2d visualizations in the web. The technique is independent from the used rendering techniques. This enables the user to show newest visualization techniques as well as specific techniques that are too expensive to re-implement in current 3d standards for the web. Future developments should address the performance and memory consumption of the generated image data. Since adjacent images differ only in a small amount, compression techniques adapted from video streams seem to be an adequate solution for a much better compression.

References

1. Web3D, http://www.web3d.org/x3d/
2. Acrobat3D, http://www.adobe.com/support/acrobat3d/
3. Google O3D, http://code.google.com/apis/o3d/
4. Bruckner, S., Gröller, M.E.: VolumeShop: An Interactive System for Direct Volume Illustration. In: IEEE Visualization, pp. 671–678 (2005)
5. Gasteiger, R., Neugebauer, M., Kubisch, C., Preim, B.: Adapted Surface Visualization of Cerebral Aneurysms with Embedded Blood Flow Information. In: Eurographics Workshop on Visual Computing for Biology and Medicine (EG VCBM), pp. 25–32 (2010)
6. Baer, A., Tietjen, C., Bade, R., Preim, B.: Hardware-Accelerated Stippling of Surfaces Derived from Medical Volume Data. In: IEEE/Eurographics Symposium on Visualization (EuroVis), pp. 235–242 (2007)
7. Behr, J., Alexa, M.: Volume visualization in VRML. In: 3D Web Technology, pp. 23–27 (2001)
8. Apple QTVR, http://en.wikipedia.org/wiki/QuickTime_VR
9. Höhne, K.H., Gehrmann, S., Nazar, T., Petersik, A., Pflesser, B., Pommert, A., Schumacher, U., Tiede, U.: VOXEL-MAN 3D Navigator: Upper Limb. Springer Electronic Media, Heidelberg (2008)

Age-Adapted Psychoacoustics: Target Group Oriented Sound Schemes for the Interaction with Telemedical Systems

Alexander Mertens, Philipp Przybysz, Alexander Groß, David Koch-Koerfges, Claudia Nick, Martin Kaethner, and Christopher M. Schlick

Institute of Industrial Engineering and Ergonomics of RWTH Aachen University, Aachen, Germany
{a.mertens,p.przybysz,a.gross,d.koch,c.nick, m.kaethner,c.schlick}@iaw.rwth-aachen.de

Abstract. For the interaction of elderly people with IT systems, an ergonomic and intuitive design as well as self-explanatory handling processes are particularly relevant. Herein adequate acoustic feedback, which accounts for the specific needs and experience of the target group, provides high efficacy and acceptance of technology with regard to Human-Computer Interaction. In this study, five different types of sound schemes are evaluated on their intuitive understanding and memorization by older users. The participants assign audible feedback to typical applications of telemedical monitoring and have to reminisce given classifications. This approach makes it possible to elicit the homogeneity of psychoacoustic models of elderly people and give recommendations for the design of acoustic feedback mechanisms for this audience. As a result, the use of familiar sounds from everyday situations has been found significantly better in terms of the consistency of the intuitive mapping and memorization for use cases in a telemedical context, in comparison to synthetic sounds that obtain their semantic denotation just by convention.

Keywords: psychoacoustic, telemedicine, acoustic feedback.

1 Introduction

The usage of complex IT systems can be greatly simplified by appropriate multimodal feedback mechanisms (such as visual, auditory, tactile). Here particular acoustic feedback helps to ensure a transparent process of interaction, when the deployed sounds can be associated with known situations, such as everyday sounds, nature sounds [1]. Especially for elderly users the appreciation of classic "beeps" is complicated due to the frequent lack of experience with computer systems and no adequately trained mental models for the incoming scenarios [2].

Within this study, an evaluation of typical applications in the telemedical context will be made, investigating what sound schemes support the most intuitive understanding and ensure the best memorization for senior users. The aim is to support an effective, efficient and satisfactory interaction with the computer system that runs error-robust even in critical situations. These criteria are known to influence

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 406–415, 2011.

the attitudes towards computers significantly and can help to bias acceptance of technology by older people positively [3].

2 Related Work

The application of acoustic signals in addition to a visual representation is one of the most meaningful methods for information transfer while interacting with electronic information systems. The Encyclopædia Britannica defines this in the article on sound design as "... feedback that is not only the need of consumers to interact with objects, but also relieves the cognitive system by distributing complex information to various perceptual channels".

Acoustic feedback provides the advantage that it is perceived regardless of the current focus of attention (subconsciously) and thus in critical situations significantly contributes to transparency and robustness of operating processes. What you hear comes from the ear via the auditory nerve to the limbic system, where it is linked by the auditory cortex with symbols and associations. The hearing is more sensitive, more accurate, faster, more efficient and less prone to deception than the eye. The perception of the human eye includes the range from 380 to 760 nanometers, which is one octave (an octave corresponds to doubling of the wavelength). At an average, healthy persons can hear a range of about ten octaves with remarkable speed (20 – 20000 Hz): the ear requires 3/1000 seconds to distinguish two successive stimuli, the eye with 20/1000, seven times as much [4].

The occurrence of presbyacusis (age-related hardness of hearing), caused by physiological aging processes, manifests itself increasingly from the fifth and sixth decade of life, especially in the high frequencies [5]. The empirical upper limit of hearing of elderly persons is around 5000 Hz and the lower at about 64 Hz. The understanding of whispered voice is about 4 meters, while values are at 20 years 15 m for the distance from ear to whisper.

Lemmens et al. were able to verify the influence of different feedback types on the reaction time experimentally, by measuring the processing speed for simple and parallel tasks. Furthermore, the subjects made less errors when acoustic feedback was provided compared to sole use of visual feedback [6]. Moreover it has been shown beforehand that subjects react significantly faster to positive, affirmative responses as to feedback that represents negative semantics [7].

Systems which have integrated consistent audible feedback in addition to the purely visual user interface, lead to a more efficient learning curve and better memorization of the system functionality. The subjective load of the user during the interaction with IT systems is substantially reduced since it is easier to keep track of the running and pending processes [8]. In this context, the felt utility value and the satisfaction of the user with a system with adequate audible feedback increased similar to the perceived complexity of the system.

As particularly suitable emerged so-called Audicons, which analog to their visual counterpart, the pictograms, use reduced, short acoustic elements to improve recognisability not only by convention, as so-called Earcons do [9].

However, too frequent use of acoustic signals may also lead to a negative effect in terms of performance and satisfaction of the user, as undesirable interruptions in the work processes may result [10]. At this, moderately complex sounds are less annoying than very simple or very complicated ones; however, perceived complexity decreases with familiarity.

3 Method

To support the independent and efficient handling of telemedicine systems and to allow intuitive operation, this study evaluates, what kind of acoustic feedback scheme is to prefer for an elderly target group [11]. Here, among the commonly used MS Windows sounds, everyday sounds (e.g. opening a door, starting a car engine, etc.), natural sounds (e.g. thunder, water splashing, etc.), jingles and spoken feedback are compared. The hypotheses are that sounds which may be associated with everyday situations, allow people, independent from particular computer literacy, a more effective and more intuitive comprehension. The efficiency for operating an IT system can thus be augmented significantly and increase the motivation to use it. In addition, it is hypothesized that the memorization of acoustic feedback mechanisms for intuitively identifiable sounds is better in comparison to digital sounds, whose semantics were adopted only conventional.

3.1 Procedure and Study Design

For the user study ten basic actions/events were identified that are relevant for the interaction of novice users with a telemedical monitoring system, such as warnings, reminder applications and status information [12]. These use cases have to be assigned by the participants to the played sounds/terms (see Figure 1). In a second part, the various sound schemes will be evaluated with regards to how reliable predefined mappings can be reminisced when they are memorized only for a short learning phase.

To guarantee comparability of the results a high consistency during each individual test was conducted. A standardized test protocol helped to achieve this. Part of this were short audio recordings, which explained the meaning of each scenario, that were played to the participants, before the experiment was started. To prevent learning effects, we counterbalanced the order of test patterns with even-size latin square.

Satisfaction about usability, cognitive and physical load as well as computer literacy were determined posterior with the help of an interview-administered questionnaire.

As the question of correct allocation of tones to situations obeys no objective criteria, the scale used for the ideal sound scheme is maximum possible consistency and uniformity in the assignments within the focused group of subjects and a low probability of errors in memorization. Through this a common understanding and best efficiency, effectiveness and satisfaction with regard to acoustic feedback mechanisms for the design of an age-friendly human-computer interaction can be provided.

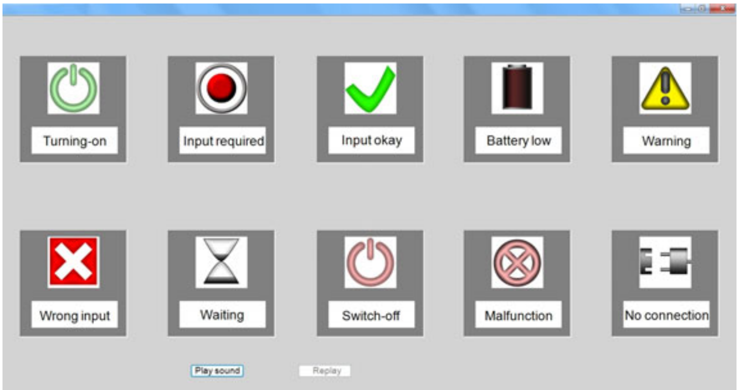


Fig. 1. Screen shot of the used test environment with the ten considered applications. The input will be carried out exclusively by the investigator to minimize non-quantifiable influences on the processing time for data entry.

3.2 Participants

A total of 85 probands (52 male, 33 female) participated in the user studies (age: min=32, q1=58, med=73, mean=71.45, q3=84, max=94). The participants were partitioned in five age groups: under 50 years (n = 9), 51-60 years (n = 21), 61-70 years (n = 26), 71-80 years (n = 19) and 81 - 95 years (n = 10). Most of the participants were recruited in the University Hospital of RWTH Aachen; the acquisition of other participants was accomplished in nursing homes, senior homes and outpatient facilities such as sports groups and cardiac care services.

With regard to health status, the participants reported to suffer, on average, from 3.2 of 14 (SD = 1.6) mentioned age-related symptoms.

The cognitive capacity for self-reliant processing of the task and the questionnaire was considered feasible for all subjects, so the results are transferable to unaffiliated interaction with equally designed telemedical systems. The social background of the participants was mixed, and thus provides representative indicators for the target group (see Fig. 2).

3.3 Pretests and Inclusion Criteria for Participation

Acceptance of Technology. The acceptance of technology as an independent variable influencing the performance of Human-Computer Interaction is determined a priori and thus enables a differentiated analysis of the results [13].

Based on the works of Beier [14] the attitudes of the elderly participants were evaluated with 15 strictly positive formulated statements that had to be rated according to a 4-point Likert Scale. These results represent the general attitudes for the seven ascertained dimensions (comfort, interest, efficacy, control, dehumanization, gender equality and value in use).

Fig. 2. Living situation, highest educational qualification and working range of participants

Computer Literacy. In order to measure computer literacy and evaluate the correlation with the test results a questionnaire from Sengpiel et al. [Sengpiel et al. 2008] with nineteen items was used. The survey is based on the symbol and term knowledge gained during Human-Computer Interaction and is generally seen as an indicator towards a person's competence in handling a computer system. Experiences such as duration, frequency and quality of usage and availability of a computer are asked in general demographic section.

Audiometry. A listening task was performed with all participants before the attendance, to ensure an adequate hearing in the relevant frequency ranges (if necessary with hearing aids) to prevent non-quantifiable effects due to incorrectly perceived sound items.

For this, the H-Man of the eccola company was used with digital MD4100 listen-only headphones of the Peltor company. The volume was calibrated to personal requirements to release between 20 - 70 dB and played alternately left and right tones of 500, 1000, 2000 and 4000 Hz. By an LED display it is possible for the investigator to follow the test and to check for each of the eight listening states. The subjective assessment of hearing was conducted in a quiet place without distraction factors. If the subjects could not hear more than one of the played tones properly, they were not included in the study.

4 Results

The study showed a significant distinction in the intuitive understanding of the different sound schemes (see Table 1). Here, the hypothesis was confirmed that for older computer users, the intuitive understanding of familiar sounds – especially everyday and nature sounds – is more homogeneous and therefore more effective in terms of human-computer interaction than the standard Windows sound schemes.

Homogeneity is determined from the track of the result matrix for each sound pattern. For this purpose, the items for each use case are arranged so that the result on the particular diagonal has the highest assignment rate. A high value of the trace thus represents a high consistency in understanding and allows a ranking of the different sound scheme alternatives.

The detection rate for spoken feedback as a reference value was, as expected, very high (track 934), but was rated "unpopular" by 86% of the participants. Reasons often mentioned were that they do not want to be dictated from a computer(voice) and that they felt more in a rush. The use of spoken feedback is therefore in terms of technology acceptance not recommended and should only be adopted for critical scenarios.

The results of the reminiscing task in the second part of the experiment covary with the results for the intuitive understanding in the first part. 91% of the acquired mappings of everyday sounds and 89% of the nature sounds were correctly reproduced. From the jingle scheme 77% were recognized correctly by the participants and from the Windows sounds 67% could be assigned to the predetermined use cases. The intuitive coupling of everyday/nature sounds to telemedical applications and functions can thus be affirmed to be an assistive tool to support the Human-Computer Interaction for elderly users.

The analysis of the average decision times for the assignment of the heard sound to one of the ten functions is another factor to support the hypotheses. The participants required on average 3.4 (SD = 0.7) seconds for everyday sounds to convey a mapping, while nature sounds took 3.2 (SD = 1.0) seconds, Windows sounds 4.4 (SD = 0.9) seconds and jingles took 5.2 (SD = 1.2) seconds to link the sound with a functionality. As for the last mentioned scheme the probands often considered where they knew the melody from, which explains the slow response times (Table 2).

Even if the shorter decision time does not correlate consistently with high homogeneity, it is still an indicator of increased efficiency in the cognitive processing of familiar acoustic stimuli by elderly users.

A Chi-Square test shows that the results are constant across age groups and the distribution is not significantly different ($\chi^2 = 11.76$, $p = 0.81$). The shown preference for everyday sounds is also independent of the frequency of use of computers ($\chi^2 = 14.3$, $p = 0.549$) and on the duration of use ($\chi^2 = 10.5$, $p = 0.69$).

4.1 Acceptance of Technology

The correlation between the user's attitude towards a system and his/her effectivity and satisfaction is unquestioned in the field of working environments [15]. Our results with the focus on personal assistance and aids for elderly people affirm the coherence in terms of relation between acceptance of assistive technology and the specific performance time. On average, the top third of participants with the most positive attitudes towards computers made 28% less errors in the reminiscing task in comparison to the third of participants with most negative attitudes towards computers. The medial performance time of participants with more than 3.5 points in average over all dimensions was 3.1 seconds while participants with less than 2.5 points in average over all dimension had a mean value of 4.9 seconds.

These parameters are of special importance when designing information systems for (tele)medical scenarios, as first contact is usually created within the conditions of a medical necessity, rather than voluntarily.

Table 1. Mappings of scenarios to played sounds by the participants in percent. The items are arranged so that the highest percentage of allocation is on the diagonal. Hereby the corresponding track of each matrix is a measure for the homogeneity of the mappings of the elderly participants.

Played sound → / Use Cases ↓	Everyday sounds										Natural sounds									
	Start a motor	Shut the car off	Brake action	Signal-horn	Ticking	Accident	Applause	Wait board	Engine stall	Random noise	Crow	Flapping Birds	Growling	Barking	Chirping	Elephant	Grunting	Purling water	Dropping water	Falling tree
Turning-on	**91**	0	0	0	2	0	7	4	5	3	**84**	11	0	0	11	4	1	0	3	6
Switch-off	0	**84**	0	0	3	0	0	2	3	4	0	**68**	0	0	3	0	3	2	3	9
Wrong input	0	4	**78**	10	0	12	0	0	4	8	2	3	**77**	17	3	7	12	0	0	14
Warning	0	0	11	**81**	0	7	0	5	3	3	3	4	21	**76**	0	10	11	0	0	21
Input required	4	7	0	2	**71**	0	4	9	7	5	9	4	0	0	**70**	2	2	13	9	0
Malfunction	0	5	4	7	0	**79**	0	0	6	2	2	5	2	7	4	**73**	9	0	0	13
Input okay	0	0	0	0	0	0	**76**	0	0	0	0	0	0	0	0	0	**61**	0	7	0
Waiting	2	0	0	0	17	0	13	**71**	2	8	0	4	0	0	6	0	0	**81**	11	0
Battery low	0	0	3	0	4	0	0	2	**67**	6	0	0	0	0	0	3	1	2	**66**	0
No connection	3	0	4	0	3	2	0	7	3	**61**	0	0	0	0	3	1	0	2	1	**37**
							Trace: 759										**Trace: 693**			

Played sound → / Use cases ↓	MS Windows sound-schemes										Jingles									
	Startup	Shutdown	Fault insertion	Ping	Incoming call	Critical Error	Ok	Notification	Battery dead	Ringout	Ode to Joy	Time to say GB	"Zonk"	Family Feud: No	Tetris	Beethoven	Family Feud: OK	Jeopardy	Trumpeting	# not available
Turning-on	**41**	9	3	0	0	4	7	9	5	2	**51**	6	0	11	11	7	12	9	8	0
Switch-off	9	**34**	4	5	0	2	3	7	5	16	11	**72**	0	7	8	5	3	5	6	5
Wrong input	3	8	**52**	17	6	13	9	12	11	7	2	9	**74**	3	8	10	2	2	11	8
Warning	3	3	15	**45**	9	11	11	9	9	5	5	3	9	**43**	7	7	4	0	12	5
Input required	9	15	6	12	**44**	5	6	14	5	7	8	0	3	6	**37**	6	11	16	9	8
Malfunction	2	6	9	11	9	**51**	11	4	10	8	6	4	8	11	8	**63**	5	0	8	4
Input okay	6	4	0	8	3	0	**49**	3	0	2	11	0	0	4	2	0	**41**	3	4	0
Waiting	9	12	3	0	11	2	0	**39**	5	3	5	5	3	8	6	0	9	**64**	0	3
Battery low	11	5	5	1	3	12	4	0	**37**	8	0	1	3	2	7	1	5	0	**35**	0
No connection	7	4	3	1	15	0	0	3	13	**42**	1	0	0	5	6	1	8	1	7	**67**
							Trace: 434										**Trace: 547**			

The mean values and standard deviations for the acceptance of technology are shown in Figure 3 for various subgroups. Over the entire sample the low interest (M = 2.24 SD = 0.8) and the fear of losing control (M = 1.98, SD = 0.83) is noticeable. A typical pattern in the sample is the extremely strong correlation between the dimensions of comfort and efficacy ($r = 0.88$, $p < 0.001$). In general there was a negative relationship between the seven dimensions of the questionnaire on technology acceptance and frequency of computer use and the fact that computers were used at work. The increasing duration of computer use was also associated with a high value of the dimension of comfort.

Table 2. Time intervals for the mapping of the played sound to one of the scenarios in seconds

Everyday sounds	Ø time	Nature sounds	Ø time	Windows sounds	Ø time	Jingles	Ø time
Start a motor	2.7	Crow	2.2	Startup	5.1	Ode to Joy	6.4
Shut the car off	3.1	Flapping Birds	3.9	Shutdown	5.3	Time to say GB	5.3
Brake action	3.4	Growling	2.7	Fault insertion	3.4	"Zonk"	3.5
Signal-horn	2.8	Barking	1.9	Ping	6.2	Family Feud: No	4.8
Ticking	3.8	Chirping	3.8	Incoming call	3.0	Tetris	7.3
Accident	3.5	Elephant	2.3	Critical Error	4.7	Beethoven	4.5
Applause	2.2	Grunting	4.5	Ok	3.8	Family Feud: OK	5.3
Wait board	4.2	Purling water	3.2	Notification	4.4	Jeopardy	3.8
Engine stall	4.3	Dropping water	2.6	Battery dead	4.3	Trumpeting	6.5
Random noise	4.1	Falling tree	4.6	Ringout	4.2	# not available	5.0
	3.4		**3.2**		**4,4**		**5,2**

A one-way ANOVA indicates that there are significant differences for the oldest participants in the study. A post-hoc analysis with Bonferroni correction showed that the age group of 80 to 95 year old participants is different on the four dimensions comfort, efficacy control and utility and on average scores more than one point less than the other groups. This shows that people who are older than 80 years often feel less comfortable and self-reliant in using computers, and appreciate the practical benefits of computer technology not as much as the other age groups do.

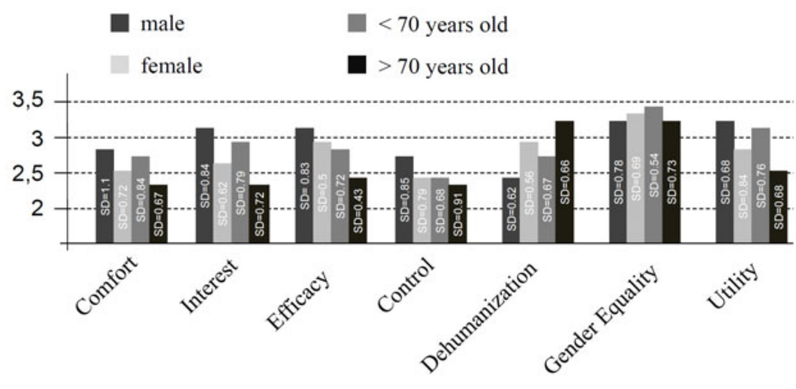


Fig. 3. Attitudes of elderly people towards computers; comparison between different subgroups

4.2 Computer Literacy

On average, the participants had about six years (SD = 4.9) experience in using computers. The use was primarily for private purpose (8% professional, 45% private, 47% in both commercial and free). 60% of the respondents said that they do not use a computer more frequently than three times a week. In the age groups over 70 years the average experience in using computers decreases noticeable and is only 2.6 (SD = 1.1) years, which is reflected in the measured results for computer literacy (see Table 3). Only for 16% of the participants over 70 years a computer is available in their own household, while in the subgroup of under 61 year old persons almost every second person has direct access. The evaluation of the experience with operation symbols and IT concepts showed that on average 60% of the 19 items were correctly assigned. Persons of the age groups over 70 years achieved almost always less correct results than the participants from younger control groups, which was to be expected due to lower experience and availability of computers both in the household and at work (Table 3). This tendency was reinforced by the fact that the elderly participants were not about to make a guess as required by the instructions if they were uncertain about the correct answer.

There was no association between the computer literacy of the participants and their preferences for everyday/nature sounds ($r = 0.23$, $p < 0.001$). However, a high success rate in the computer literacy test correlated positively with a better performance time for the mapping of sounds to scenarios ($r = 0.89$, $p < 0.001$) and correlated negatively with probability for errors in the memorization task ($r = -0{,}84$, $p < 0.001$). No significant correlation was found between the mappings of sound and the particular experience of the participants. Just for the MS Windows sounds a high homogeneity (trace: 732) for participants that correctly assigned more than 15 items in the computer literacy questionnaire could be ascertained.

Table 3. Percentage of correct responses while determining the computer literacy of the participants according to age group, education level and work area

Age group	< 50 years	51-60 years	61-70 years	71-80 years	81-95 years	
	86,2	77,3	56,12	41,9	21,8	
Education	General school	Secondary school	Diploma	Senior high school	Academic studies	Doctorate
	44,5	60,4	62,3	64,7	71,6	63,5
Work area	Tradesman	Administration	Engineering	Homemaker	Social	Other
	49,8	64,3	72,4	16,9	63,4	52,6

5 Conclusion

The importance of telemedicine for present and future medical care especially with regard to Demographic Change is undisputed. However, today's systems often do not adequately respect the specific requirements of elderly users because of missing specifications.

The results of this study allow to give design recommendations for the implementation of acoustic feedback mechanisms, taking into consideration the preferences and competences of senior users. Acoustic response as a guidance for the focus of attention especially for elderly users leads to a more robust and satisfactory interaction, for example with telemedical systems. The sole use of acoustic feedback, however, supplies no sufficient assistance and direction for abstract processes and therefore needs to be enriched with visual/textual items.

Everyday sounds in place of standard digital sound sequences and melodies (jingles) result in a significantly more intuitive comprehension regarding to the homogeneous technical understanding within the target group and to an increased performance in short-term learning/memorization.

The feedback from the participants about previous experiences with computers and IT systems straightens out that the success and sustainability depends to a large extent on how far individual barriers can be overcome and how the devices can be integrated into daily living without stigmatizing the user. The utilization of innovative digital and telemedical systems allows not only to increase the quality of life of elderly people significantly but is also strategically important for many companies to enter the "senior market".

The accomplished studies are basic research for the development of age-appropriate systems that meet the requirements and circumstances of the "Best Ager" and help to improve the acceptance of technology.

References

1. Marila, J.: Experimental Comparison of Complex and Simple Sounds in Menu and Hierarchy Sonification. In: Proceedings of the 2002 International Conference on Auditory Display, Kyoto, Japan, pp. 104–108 (2002)
2. Raskin, J.: The Humane Interface – New Directions for Designing Interactive Systems. Addison-Wesley, New York (2005)
3. Fastl, H., Zwicker, E.: Psychoaccoustics – Facts & Models. Springer, Berlin (2007)
4. Herrmann, M.: Psychoakustik und Sound-Engineering. Grin Verlag, München (2001)
5. Nieschalk, M., Hustert, B.: Der klinische Einsatz von Distorsionsprodukten otoakustischer Emissionen bei Presbyakusis. Laryngo-Rhino-Otolologie 75, 129–134 (1996)
6. Lemmens, P.M.C., Bussemakers, M.P., de Haan, A.: The effects of earcons on reaction times and error-rates in a dual task vs. a single task experiment. In: Cook, P.R. (ed.) Proceedings of the International Conference on Auditory Display, pp. 177–183 (2000)
7. Hevner, K.: The mood effects of the major and minor modes in music. American Journal of Psychology 47, 103–118 (1935)
8. Buxton, W., Gaver, W., Bly, S.: The Use of Non-Speech Audio at the Interface. In: CHI Tutorial Notes. ACM, New York (1990)
9. Garzonis, S., Jones, S., Jay, T., O'Neill, E.: Auditory icon and earcon mobile service notifications: intuitiveness, learnability, memorability and preference. In: Proceedings of the 27th International Conference on Human Factors in Computing Systems (CHI 2009), pp. 1513–1522. ACM, New York (2009)
10. Block Jr, F.E., Nuutinen, L., Ballast, B.: Optimization of Alarms: A Study on Alarm Limits, Alarm Sounds, and False Alarms, Intended to Reduce Annoyance. Journal of Clinical Monitoring and Computing 15(2), 75–83 (1999)
11. Moore, B.: An Introduction to the Psychology of Hearing. Elsevier Academic Press, Amsterdam (2004)
12. Mertens, A., Dünnebacke, D., Kausch, B., Laing, P., Schlick, C.M.: Innovation of homely rehab with help of telemedical services. In: IFMBE Proceedings of World Congress on Medical Physics and Biomedical Engineering. Springer, München (2009)
13. Gina, M.J., Sherry, L.W.: Influence of Direct Computer Experience on Older Adults' Attitudes Toward Computers. Journal of Gerontology: Psychological Science 47(4), 250–257 (1992)
14. Beier, G.: Kontrollüberzeugungen im Umgang mit Technik - Ein Persönlichkeitsmerkmal mit Relevanz für die Gestaltung technischer Systeme. Dissertation, Berlin (2004)
15. Dethloff, C.: Akzeptanz und Nicht-Akzeptanz von technischen Produktinnovationen. Pabst Science Publishers, Lengerich (2004)

Bringing the Home into the Hospital: Assisting the Pre-Discharge Home Visit Process Using 3D Home Visualization Software

Arthur G. Money[1], Anne McIntyre[2], Anita Atwal[2], Georgia Spiliotopoulou[2], Tony Elliman[2], and Tim French[1]

[1] University of Bedfordshire, Park Square, Luton, Bedfordshire, LU1 3JU, UK
{arthur.money,tim.french}@beds.ac.uk
[2] Brunel University, Uxbridge, Middlesex, UB8 3PH, UK
{anne.mcintyre,anita.atwal,georgia.spiliotopoulou, tony.elliman}@brunel.ac.uk

Abstract. The feasibility of using interactive 3D home visualization software (I3DHVS) as a tool to aid Occupational Therapists (OTs) in carrying out pre-discharge home visits (PDHV) is explored. Three focus groups involving 25 OTs from across the UK were carried out. Participants were asked to report their level of experience with Information Technology (IT) and gaming software. After a demonstration of the I3DHVS OTs were asked to discuss where, when, how and by whom this software may potentially be used, and to identify associated strengths, weaknesses, opportunities and threats (SWOT) of use within the specified contacts. A thematic template analysis was then carried out on the transcribed focus group data which focused on two key Technology Acceptance Model (TAM) criteria which mediate users' behavioral intention and actual use of new technologies: (1) the Perceived Usefulness (PU) of the software within the PDHV process; (2) the Perceived Ease of Use (PEoU) of the software. The results revealed that although a number of adaptations to the existing application may be necessary, OTs are optimistic about the use of I3DHVS with a range of patient groups. The tool was also seen to have the potential to improve communication and collaboration across inter-agency care teams.

1 Introduction

A key role of an occupational therapist (OT) is to facilitate the discharge process for patients who have been treated within the hospital-based acute care settings [1]. The hospital-home discharge process often requires a pre-discharge home visit (PDHV) to be carried out by an OT. The aim of the home-visit is to ensure that the discharge process is carried out in a timely and effective manner and allows the OT to develop a more comprehensive idea of the real-world environment and associated challenges the patient will be exposed to on discharge [2]. In some cases, this involves taking the patient to their home and carrying out an assessment of their ability to perform everyday tasks. PDHVs are typically required in a number of scenarios: when

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 416–426, 2011.

adaptations to the patients home are necessary in order to facilitate safe discharge and independent living; a patient returns home in a wheelchair; a patient has reported falling as a result of an environmental hazard; a patient is cognitively impaired and disorientated by the hospital environment; patients are not confident that they will be capable of living independently [3].

Although there appears to be policy in place to determine which scenarios require PDHVs, there is little research evidence to support the efficacy of such visits or indeed to support the notion that health and well-being is improved as a result of such activity. Questions over the efficacy focus on the lack of patient/carer involvement and consultation throughout the process [4]. In many cases, patients and carers have reported to feel excluded from the PDHV process and the decisions that are made with regards to adaptations to patient homes [5]. This in turn has been seen to impact negatively on levels of satisfaction and indeed levels of patient adherence [6]. Furthermore, some older adults perceive the process as demoralizing, daunting, and anxiety provoking [7]. Recent research findings indicate that there is a need evaluate the procedure currently adopted in the delivery of PDHV, particularly paying attention to the timing of the intervention within the discharge process and the length of time over which the discharge process is provided [2]. Due to the time and resource intensive nature of carrying out PDHVs, however, this may be challenging as OTs already find it difficult to prepare for, and carry out home visits adequately [7].

There is a need to develop new and innovative approaches to enable the efficient and effective provision of hospital-home discharge interventions, enhancing opportunities for patient engagement and reducing the likelihood of these interventions being perceived negatively by patients. Our research proposes utilizing an open source I3DHVS application [8] to aid the hospital-home discharge process. Although the precise application of such software is yet to be determined via this research, one potential application may be to ‘bring the home into the hospital’ by OTs and patients collaboratively developing 3D representations of the patient’s home whilst still in hospital. This could provide the potential to reduce the number of home-visits that are necessary and more actively involve the patient within the PDHV process. The SweetHome3D software enables users to rapidly create 3D representations of the interior of a patients’ home, which may be navigated through in real time. Fig. 1 provides an example of 3D home environments produced by the 3DSweetHome application.

Fig. 1. Examples of virtual home environments produced using 3D visualization software

A range research has already demonstrated the value of using I3DHVS within a therapeutic context. For example, in the treatment of public speaking anxiety [9]; the treatment of claustrophobia [10]; simulation of the effects of excessive alcohol consumption to patients [11]. Virtual 3D environments have also been used as an assessment tool for stroke patients [12] and for assistive memory tasks in the treatment of patients with amnesia [13]. To the best of our knowledge, the use I3DHVS has not yet been used within the PDHV process. Given the novel nature of this work, there is a need to scope the feasibility of exploiting this technology within this specific health care context.

Much research effort has been invested into understanding user end users' reactions and motivations to technology acceptance, adoption and use [14]. Although much of this research has been within the Information Systems research domain, there is increasing interest in gaining a better understanding of the factors that influence user acceptance, adoption and use of technology within the health care domain [15]. The Technology Acceptance Model (TAM) is perhaps the most notable theory applied in the explanation of user motivations, attitudes and responses to acceptance, adoption and use of new technology [16]. Despite its relative simplicity, the most basic form of TAM is typically seen to provide an explanation of approximately 40% of issues related to technology acceptance [14]. TAM proposes that when presented with a new technology, users' Behavioral Intention to use (BI) and their Actual Use (AU) of technology are typically mediated by two key factors: Perceived Usefulness (PU) - the extent to which the user perceives that the new technology will aid them in performing the task at hand, and Perceived Ease of Use (PEoU) - the extent to which the individual believes using the technology would be free of effort [17]. TAM is now increasingly being applied within the health care research domain [18]. Examples include exploring the acceptance of: telemedicine technology by nurses [19]; Personal Digital Assistants (PDAs) by physicians [20]; portable postural assessment technologies for use by physiotherapists [21]; mobile picture archiving technologies for dental care [22]; and a range of customizable and wearable health care devices for patients and practitioners [23]. Although the vast majority of TAM research to date has been quantitative, there is increasing recognition that qualitative enquiry, particularly in the early explorative stages, is well suited to scoping the potential feasibility of a given technology within a novel context [24].

In this explorative study, we recruit 25 Occupational Therapists (OTs) from across the United Kingdom to discuss, the potential of deploying this I3DHVS to aid the PDHV process. The focus group data is analyzed with a particular focus on the two key TAM criteria: PU and PEoU with the aim of gaining insights into the following research questions:

1. What are OTs beliefs about the Perceived Usefulness of this technology and with whom may this software be useful/not useful?
2. What are OTs beliefs about the Perceived Ease of Use of this technology and what are the perceived barriers to acceptance and use of this software?

The remainder of this paper is structured as follows. In Section 2 we describe a study carried out to explore the feasibility of incorporating I3DHVS into the PDHV process. Section 3 presents the results of the study, and Section four concludes the paper, reflecting on future research directions based on the outcomes of this study.

2 Method

A qualitative approach was taken to explore the above research questions. Three focus groups were conducted with OTs working within the National Health Service (NHS) within the United Kingdom. A good national level of participation was achieved (Scotland, Wales and England). The number of participants varied in each of the three groups (N= 5, 12, and 8 accordingly). Each focus group included: (a) a discussion to explain the session and establish level of experience with information and gaming technology, (b) demonstration of the I3DHVS, and (c) a semi-structured discussion to consider the potential utility of the software and areas of deployment. Over the course of the focus group sessions, participants were asked to discuss where, when, how and by whom this software may potentially be used, and to carry out a SWOT analysis (strengths, weaknesses, opportunities and threats) on the potential areas of deployment. SWOT analysis has served as a valuable tool in scoping the feasibility of deploying Virtual Reality (VR) technology within the health care context [25], and provided a valuable means of capturing a comprehensive representation of participants' views relating to this technology. All of the groups were audio recorded and transcribed verbatim. Ethical approval for the project was obtained from Brunel University's Research Ethics Committee. The data was transcribed and subsequently analyzed by the project team. Thematic template analysis was used to analyze the transcribed data. This is a form of thematic analysis, which involves development of a coding template that represents a summary of the themes that are seen by the researcher(s) as being of importance within the data set [26]. Analysis often begins with some *a priori* themes/codes that are of interest to the researcher, in this case, the focus group data was analyzed with the in order to identify themes representing OTs views relating to the TAM key criteria: Perceived Usefulness (PU) and Perceived Ease of Use (PEoU) of I3DHVS.

Initially, the textual dataset was perused in its entirety to conceptualize the overarching themes that existed. The textual dataset was then perused for a second time, to identify themes and sub-themes that corresponded with the interests of the researchers. This was carried out iteratively until a number of sub-themes within the two existing *a priori* themes were identified. Segments of text that corresponded with these themes were taken note of. The dataset was perused iteratively, and sub-themes were cut spliced and linked together until a clear consensus of the main themes and sub-themes were reached by the first and sixth authors. When no further refinement of the categorization could be derived, a final template representing the themes and sub-themes was produced, which was linked to the words of the respondents as they were presented within the focus groups. For a detailed description of the thematic template analysis process, see King [26].

3 Results

The findings of the analysis revealed a number of sub-themes relating to PU and PEoU. The results of this analysis are now presented.

3.1 Perceived Usefulness

This theme refers to views relating to whether the I3DHVS could be perceived as useful/not useful and in which scenarios the technology may be incorporated into the PDHV process. Two sub-themes of Perceived Usefulness are presented within this section, these are: Not a Replacement and Communication & Collaboration.

Overall, there was a feeling that this application could be useful in adding another dimension to the PDHV process. In particular, the notion of having a navigable 3D representation of a patient's home would provide a welcomed added layer of detail, as is articulated in the quote below.

> *"Well I think just seeing something in 3D as opposed to 2D is always going to be helpful anyway, because you've got that extra dimension to play around with, and I think that gives you much more information anyway just to make a kind of some sort of judgement."*

Not a replacement: There were, however, also some perceived drawbacks to the notion of utilizing this technology within the PDHV process. On one hand, there was a feeling of apprehension and concern that this technology may pose a threat to the role of the OT, and a perceived risk of job losses and adaptations to the role which may negatively impact upon patients. One scenario described was a vision of the future, in which patients would lose out on any form of personalized contact with OTs, resulting in the entire PDHV process being carried out remotely.

> *"In 10 years' time…this will be standard practice I'm sure…and you'll be doing your job sitting at home…"*

It was felt that in many cases, a significant reduction in patient/practitioner contact time, facilitated by the use of this technology, would impact negatively on the patient, particularly given the range of important real-world subtleties that must be taken into account when providing the PDHV intervention. In the quote below one participant suggest that technology for health care should be adopted with caution.

> *"I guess I'm just a little bit wary probably of a number of technologies being used, and the potential that it could be perceived as not personal. I think we've got to go this way, because I think technology…it's just everywhere… but then we're talking about using another bit of technology to take away"*

It was also felt that the I3DHVS could be potentially valuable, however, should not be considered as a potential substitute or replacement for face-to-face contact, but rather should be considered as a tool used to aid and assist the therapeutic process. In complex cases, and with patients that perhaps have difficulty travelling to their home prior to discharge, the use of the I3DHVS application could be particularly useful.

> *"Yeah I don't think it would replace the home visit. Because it's that patient's ability to engage within the environment that we'll always look at …Where we've got a real difficulty taking somebody difficult … like it can really help."*

Communication & collaboration: Use of the I3DHVS in order to improve communication and collaboration with a range of individuals involved with the PDHV process was seen as being potentially of notable benefit to a range of stakeholders in the therapeutic process. In particular, the function of being able to produce a tangible and navigable representation of the patient's home environment, which could be shared with a range of individuals, provides the opportunity for OTs to share representations of the patient's home with a wide range of members of a care team. This affords the potential to make joint decisions about potential adaptations to the home, and draw on skills and previous experience of other team members, that otherwise would not have been able to contribute to the decision making process. One particular scenario in which this was seen as useful was to facilitate more effective multi-agency working, decision making, and the communication of more explicit instructions to other care agencies in terms of provision of care. The quote below describes the potential communicative and collaborative benefits of using I3DHVS in the context of multi-agency working.

> *"I like the idea of being able to set up the room, because we've had cases where we're often saying 'The room's got to be like this'…And sometimes we'll find that the person sets up the home as we've described…and then the agency will go in and say 'You can't have that up against the wall'… and actually to have everyone together and say 'This is the space, we can provide the bed' it's a combined decision."*

The I3DHVS was seen as potentially providing a valuable tool through which family members and home carers could provide input into the process, as suggested in the quote below.

> *"I can see it potentially replacing an access visit more than a home visit. Especially if you had a very sensible next of kin that could do all the measurements for you, and then you could sit down with it…if you were au fait with that software you could sit down with them for 20 minutes – bang bang bang, and get it down, and then have the conversation with the patient – rather than you having to drive there, measure it all up, write it all up – I would see that being a time saver, but will not replace a home visit."*

There was a feeling that less senior staff such as OT assistants could be better utilized if this technology was made available to them. This in turn could lead to greater efficiencies in terms of use of these resources, and new opportunities to delegate certain activities to these team members.

> *"I think also in terms of an assistant…they're doing an increasing number of our access assessments, particularly where we're just wanting more feedback on the environment, furniture measurements, where furniture's located. And that can be something they'd quite struggle with communicating that to us and how to put that down as succinctly as possible…I think this could be a good use for that, because that then enables the qualified OT to analyze what that environment will be for the patient."*

Improving collaboration and engagement between the OT and the patient was also seen as a notable benefit of deploying I3DHVS. Particularly in a scenario which provided the patient with the opportunity to collaboratively construct 3D representations of the home with the assistance of the OT. It was felt that the OT would better appreciate the challenges that the patient could face when being discharged back to their home, but also such a scenario could improve the likelihood that any adaptations made to the home would be more readily accepted by the patient, since they contributed to the decision making process. The two quotes below, provide examples of OTs thoughts relating to this.

"…to picture pieces of equipment that we talk about in the hospital, you know whatever – this is what it will look like in your house. That would be nice for people to see that. Because despite how many times we practice with them in the hospital we still can't imagine 'Oh it's going to take up much more space in my house, my house is very different." …
"Also it might help sort of patients' feeling of empowerment in that kind of sense that you know you can come back and talk through it together rather than…at the minute you come back for an access visit and you say 'We think you should put this there and that there and change this around' but maybe you could sit down together at the computer and work through it and say well what about if we did this – gives them much more input and a feeling of you know being involved."

3.2 Perceived Ease of Use

This theme refers to the views presented by participants that related to the perceived level of effort that would be required to deploy the 3D visualization software and the perceived barriers to deployment. Two sub-themes were identified within this theme, these are: Attitudes Towards Information Technology and Required Adaptations.

Attitudes Towards Information Technology: A major consideration was the types of patient groups and health care practitioners that were likely to feel comfortable using this application. It was felt that many older adult patients may actually find this application useful, despite the commonly held 'misconception' that this group does not to feel confident with IT. OTs suggested that in many older adults are familiar with technology, often more so than health care practitioners, and hence this application could be useful for these patients.

"…we're finding increasingly…they [referring to older adults] are doing online shopping…a lot of them do email with their grandchildren and are in touch through computers. So I don't think they're all as unfamiliar as I am in some cases."

Cognitive ability and socio-economic background was felt to be a more useful indicator of familiarity and attitudes towards IT. Despite these being clearly perceived benefits to adoption of this software, it was felt that often there are organizational constraints that may make the use of this technology difficult.

"Even if the technology's there and the potential's there that you can see as a practitioner, sometimes it's difficult to actually, just because of the constraints of the organization, actually access it and use it in a way that you would like to."

Such constraints included a lack of IT hardware being available to OTs in order to actually access any software applications. It was also expressed that there often is no designated therapeutic area in which OTs can see patients and have access to appropriate IT equipment. This point is exemplified below.

So even for example like internet based tools … so if you want to educate somebody about I don't know the structure of their hip joint and what's happened, you know there's loads of stuff you could use on the internet to educate them about 3D sort of models and stuff, but where do you do that? Do you bring them into the office and ask everyone else not to make any phone calls at the same time? Or … do you know what I mean? There's some sort of practical frustrations from the NHS environment.

Required adaptations: Although the software was perceived as relatively easy to use in modeling patients' homes with a straightforward layout, it was felt that homes with more complex layouts may become more challenging to model. In particular, it was felt that some homes which had atypical items of furniture and varying floor levels may require considerable effort to model using the I3DHVS. This led some OTs to suggest that the use of this software could only realistically be justified in scenarios where there was significant therapeutic gain that outweighed the required effort more so than could be achieved via current practice.

"Well I think we were generally very positive about the software generally, but I think one of the issues was really balancing up the complexity of how you would use it…so there's always going to be this sort of balance, and a kind of tension between how kind of easy is it just to utilize on a day to day basis against having very complex functions which would be really helpful but would mitigate against kind of ease of use really."

Extending on from this point was a consideration of the usability and functionality of the I3DHVS. Some positive features of the software that are already offered include the facility to draw 3D models to scale, and to rapidly adapt the height and width of items of furniture within the model. However, there were a number of features that are not currently offered which OTs felt could significantly improve the application in order to make it more fit for purpose for the requirements of the PDHV process and in turn more appropriate for use on a day-to-day basis. In particular, variable lighting settings were felt to be necessary to model the potential complexity that poor lighting may add to mobility around the home. Varying floor textures and how these may affect mobility and health and safety considerations was also a required feature.

"If technology could be so open… you could start to think about the impact of different lighting on a setting. To me it's quite a logistics setting to look at this as a space…but you've not got the variables of different textures on the floor."

Other functionality that was seen as necessary included providing a suite of specific items of furniture and specialized OT equipment, which are typically introduced into a patient's home on discharge home. Some of these items included wheelchairs, hoists for offering access to the bath, hand rails and wheelchair ramps.

Furthermore, modeling ‘general clutter’ around a patient’s home was also noted a useful phenomenon to be modelled. Often items like newspapers, junk mail, laundry and other miscellaneous items are placed throughout the home and act as hazards for the patient. In the two comments below, OTs comment on how clutter is an important aspect that affects patients mobility within their homes and that the software as it is, could not easily model levels of clutter, although this would be desirable.

> *“A limited range of things and the objects. Sometimes clutter up and down the stairs…Stacks of newspapers…or just papers in general, or laundry”*
>
> *“I mean it’s really good, but you wouldn’t be able to get the clutter and things on it as it is at the moment.”*

4 Concluding Remarks

In this study, we have explored the feasibility of exploiting 3D home visualisation software (3DHVS) as a means to support OTs in carrying out pre-discharge home visits. Data from three focus groups, carried out with a total of 25 OTs from across the UK, was analysed using thematic template analysis focusing specifically on two key Technology Acceptance Model (TAM) criteria: Perceived Usefulness (PU) and Perceived Ease of Use (PEoU). The results revealed that in terms of PU, OTs in general had a positive attitude towards the potential value of this software within the PDHV process. It was anticipated that the 3DHVS would not replace a home visit, given the important real-world subtleties that must be taken into account in the process. However, particular value was seen in using this software in complex cases. There was also potential to use this software to improve communication and collaboration within multi-agency teams, make better use of less senior staff such as OT assistants, and to enhance the group decision making process. Furthermore, there was potential to improve the patient/practitioner relationship via collaborative use of this software, hence empowering the patient within the decision making process. These are promising results, given the need and potential benefits of better engagement with patients in the PDHV process [2]. Factors relating to PEoU revealed that there is potential scope for using this software with older adult patients. However, patients with some level of cognitive impairment or who are not familiar with IT may find it more challenging to gain a positive therapeutic outcome from a technology assisted intervention. Organisational constraints, in terms of lack of access to IT and designated space to use technology with patients may also pose a challenge to deploying the 3DHVS with patients. In order to gain maximum benefit and to make the software usable and applicable within the PDHV process, it was also noted that a number adaptations should be made to better suit the needs of the OT. Some suggested adaptations included a need for a wider range of items of furniture to be made available within the furniture library, and to offer the function of changing lighting and floor textures. In light of the positive findings of this study, there appears to be potential to pursue this research further. In principal, the software promises a number of benefits: to improve the multi-agency decision making process; improve levels of patient engagement; offer improved efficiencies by making better use of less senior staff’s time. Our future research will involve carrying out specific case study examples of deployment of this software, using Wizard of Oz techniques, to gain further insight into the extent to which these potential benefits may be realised within real world PDHV contexts. Our research will also explore the feasibility of using 3dHVS with a range of patient groups, in order to gain more insight into patient attitudes towards this technology, and to establish which of these groups are likely to benefit most from its use.

References

1. Cumming, R.G., et al.: Adherence to Occupational Therapist Recommendations for Home Modifications for Falls Prevention. The American Journal of Occupational Therapy 55(6), 641–648 (2002)
2. Atwal, A., et al.: Older adults and carers' perceptions of pre-discharge occupational therapy home visits in acute care. Age and Ageing 37(1), 72–76 (2008)
3. NHS. Occupational Therapy Home Visiting Policy (2010) [cited December 7, 2010], http://www.ekhuft.nhs.uk/home-page/search/?q=home+visit+policy
4. Nygard, N., Grahn, U., Rudenhammer, A.: Reflecting on practice: are home visits prior to discharge wothwhile? Scandanavian Journal of Caring Science 18, 183–203 (2000)
5. Mann, W.C., Gooshall, S., Justiss, M.D.: Dissatisfaction and nonuse of assistive devices among frail elders. Assistive Technology Journal 2, 130–139 (2002)
6. Walker, E., Dewar, J.B.: Dewar, How do we facilitate carers' involvement in decision making? Journal of Advanced Nursing 3, 329–337 (2001)
7. Atwal, A., et al.: Occupational Therapists' Perceptions of Predischarge Home Assessments with Older Adults in Acute Care. British Jour. of Occupational Therapy, 71(2), 52–58 (2008)
8. SweetHome3D (October 10, 2010), http://www.sweethome3d.com/index.jsp
9. Wallach, H., Safir, M., Bar-Zvi, M.: Virtual Reality Cognitive Behavior Therapy for Public Speaking Anxiety. Behavior Modification 33(3), 314–338 (2009)
10. Bruce, M., Regenbrecht, H.: A Virtual Reality Claustrophobia Therapy System – Implementation and Test. In: IEEE Virtual Reality Conf., Lafayette, LA, USA. IEEE Press, Los Alamitos (2009)
11. Fleming, M., et al.: Virtual Reality Skills Training for Health Care Professionals in Alcohol Screening and Brief Intervention. The Journal of the American Board of Family Medicine 22(4), 387–398 (2009)
12. Broeren, J., et al.: Virtual reality and haptics as an assessment device in the post acute phase after stroke. CyberPsychology & Behavior 5(3), 207–2011 (2002)
13. Brooks, B.M., et al.: Route learning in a case of amnesia: A preliminary investigation into the efficacy of training in a virtual environment. Neuropsychological Rehabilitation: An International Journal 9(1), 63–76 (1999)
14. Lee, Y., Kozar, K.A., Larsen, K.R.T.: The technology acceptance model: Past, present, and future. Communications of the Association for Information Systems 12, 752–780 (2003)
15. Holden, R.J., Karsh, B.: The technology acceptance model: its past and its future in health care. Journal of Biomedical Informatics 43(1), 159–172 (2009)
16. Bagozzi, R.P.: The legacy of the technology acceptance model and a proposal for a paradigm shift. Journal of the Association of Information Systems 8, 244–254 (2007)
17. Davis, F.D.: Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly 13(3), 319–340 (1989)

18. Yarborough, A.K., Smith, T.B.: Technology acceptance amongst physicians. Medical Care Research and Review 64, 650–672 (2007)
19. Kowitlawakul, Y.: The Technology Acceptance Model: Predicting Nurses' Intention to Use Telemedicine Technology (eICU). Computers, Informatics, Nursing (in press)
20. Yi, M., et al.: Understanding information technology acceptance by individual professionals: towards an integrative view. Info. & Management 43(3), 350–363 (2006)
21. Van Schaik, P., Bettany-Saltikov, J.A., Warren, J.G.: Clinical acceptance of a low-cost portable system for postural assessment. Behaviour & Infor. Tech. 21(1), 47–57 (2002)
22. Chang, H.: Application of the extended technology acceptance model to picture archiving and communication systems in dental hospitals. Journal of the Korean Society of Medical Informatics 15(3), 265–272 (2009)
23. Lishana, X., et al.: The perception and intention to adopt female-focused healthcare applications (FHA): A comparison between healthcare workers and non-healthcare workers. International Journal of Medical Informatics 78(4), 248–258 (2009)
24. Wu, P.F.: Opening the black boxes of TAM: towards a mixed methods approach. In: Thirteenth International Conference on Information Systems. Association of Information Systems, Phoenix (2009)
25. Rizzo, A., Kim, G.: A SWOT Analysis of the Field of Virtual Reality Rehabilitation and Therapy. Presence 14(2), 119–146 (2005)
26. King, N.: Using templates in the thematic analysis of text. In: Cassell, C., Symon, G. (eds.) Essential Guide to Qualitative Methods in Organizational Research. Sage, London (2004)

Design of a Paired Patient-Caregiver Support Application for People Coping with Cancer

Christine M. Newlon[1], Robert Skipworth Comer[1], Kim Wagler-Ziner[2], and Anna M. McDaniel[1,2]

[1] Indiana University School of Informatics, 535 W. Michigan, Indianapolis, IN 46202
[2] Indiana University School of Nursing, 1111 Middle Drive NU 340, Indianapolis, IN 46202
{cnewlon,rscomer,kwagler,amcdanie}@iupui.edu

Abstract. Smart medical interfaces have the potential to address the unique health needs of each individual user. However, they must be carefully designed and tested to provide a supportive and non-stressful experience for users who are ill or overloaded. This paper documents a design methodology used to develop such an application, describing issues that can be encountered during the development process. The paper also illustrates the use of patient-entered data to drive multiple processes, showing how a smart medical interface can mediate among multiple stakeholders.

Keywords: smart medical interface, human-computer interaction, iterative design, cancer, web-based intervention, symptom management.

1 Introduction – The Design of Smart Medical Interfaces

As the public's access to powerful computer resources grows, so does the effort to create computer-based tools that can intelligently customize themselves, based on the unique needs of their users. The design of smart medical interfaces is particularly promising for the area of consumer health applications because the health needs of individuals are affected by many different factors, and must be uniquely addressed [1, 2, 3]. However, computer-based applications for use by patients and their families must be carefully designed to provide a supportive and non-stressful experience. This is challenging because illness touches people from all walks of life, leaving them distressed and vulnerable. Therefore, it is important to apply iterative, user-centered development methodologies to the design and implementation of such interfaces [4].

This paper describes our *Pathways for Caring* project, which involves the creation of an application that serves as a mediator between patients suffering from cancer, their primary caregivers, their medical providers, and medical researchers looking for better ways to provide care. This system, known as the *Patient-Caregiver Management System* (PCMS), guides both cancer patients and their caregivers through an assessment of their symptoms and needs on a weekly basis. The information entered by the patients and caregivers is captured by the PCMS and used

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 427–436, 2011.

to produce customized guides for both the patient's symptom self-management and the caregiver's stress self-management. These guides are drawn from database repositories of self-help information, known as the *Symptom Management Toolkit* and the *Pathways for Caring Guide*, that have been developed in cooperation between Indiana University Purdue University at Indianapolis (IUPUI) and Michigan State University (MSU) [5, 6]. The PCMS also uses the gathered information to determine if the patient's provider should be notified about any developing problems, sending a HIPAA-compliant notification to the provider if this is the case. Finally, the PCMS captures non-identifiable patient and caregiver information, such as how much time these users spend looking at each page of their care guides, and how highly they rate the information in the guides. This growing repository of behavior and experience data will help medical researchers develop better care methods for the future.

Given the sensitive nature of these functions, it is important that the design team follows best practices in human-computer interaction [4]. In this paper, we describe the methodology we are using in hopes of providing guidance to developers of other patient-facing health interfaces, thus contributing to health practice in this area.

The paper begins with the user needs that drive development of the PCMS. This is followed by a description of the methodology used to design the application. Following this is a section detailing design and implementation of the pilot testing process. The paper then provides a walk-through of the most significant application interfaces resulting from our design process, and concludes with a discussion of the implications for the design of patient-facing smart medical interfaces.

2 Background – The Needs of Patients and Caregivers

Cancer is one of the most prevalent chronic illnesses in the United States affecting both patients and their family members. Cancer patients experience a number of distressing, life altering symptoms during treatment. Management of chemotherapy-related symptoms is a significant burden to both cancer patients and their caregivers. As they face these distressing symptoms during treatment, cancer patients and their families have significant needs for 1) help with symptom management, 2) communication and coordination of care, and 3) help with the burden of caregiving [7]. Provision of cancer-related information helps patients and their caregivers to improve symptom management and manage stress, thus increasing overall quality of life. It is this set of needs and opportunities that provides the impetus for the PCMS.

3 Design Methodology

One key for successful interface design is an iterative development process. Our design methodology begins with extensive information gathering and needs assessment, followed by a period of conceptualization and discussion within the design team. Many rounds of coding, internal testing, and heuristic inspection eventually yield a working prototype. At this point, we conduct several rounds of formal usability testing with external recruits, who are demographically similar to the intended application users, and who role-play being application users. Finally, once all identified usability problems have been corrected, the application is ready for a full-scale pilot test with real patients. At the conclusion of the pilot test the overall results are examined, desired improvements and elaborations are identified, and the design process begins again with a needs assessment for the new features. To date, we have successfully used our process for both the Phase I Symptom Self-Management interface and the Phase II Care-Giver Self-Management interface.

By including usability as well as pilot testing in our process, we follow the recommended best practice for software development [4]. This is necessary in order for the methodology to find and correct both software bugs and poorly designed features. Many of these design issues can only be found through testing.

3.1 Design of the Phase I Symptom Self-Management Interface

Design of the Phase I Symptom Self-Management interface began with the creation of a focus group of common citizens who shared an interest in cancer treatment. We also met with a pre-existing group of nurses, and with an oncologist, to get the provider perspective. During several meetings with the focus group and the providers, we gathered their ideas and also used a card-sort technique to understand how members of the public and the provider groups conceptualized cancer treatment information. Additional background was provided by evaluation of a version of the application, developed at MSU [5], that used telephone technology, rather than web technology.

The needs assessment and conceptualization stages of the Phase I development process are documented in Newlon, Hu, Stratton, and McDaniel [6], which also shows screen shots of the resulting interface. Implementation of the prototype was expedited through revision and reuse of self-help content from the telephone-based application.

Usability testing of this interface applied a simplified version of the best practices methodology. The testers were demographically suitable recruits role-playing pre-scripted scenarios as patients. Each participant's demographic information and prior knowledge were collected via a pre-test questionnaire, while overall reactions to the application were collected via a post-test questionnaire. Task successes and issues were collected by observer notes and by videotape.

3.2 Design of the Phase II Caregiver Self-Management Interface

The Phase II Caregiver Self-Management interface was created using a similar process. However, instead of a focus group, needs assessment and caregiver self-help content development were accomplished through consultation with domain experts.

Due to the variety of interfaces in the Phase II version, usability testing was conducted individually on each type of user (i.e., patient, family caregiver, health care professional), and also on a patient-caregiver pair, to make sure both the individual features and the linked functionality worked as required. The testers were again recruits role-playing pre-scripted scenarios as patients, caregivers, and providers.

4 Pilot Testing of the Application

A second key for success in developing a medical interface is rigorous field-testing. In addition to the usability testing that occurs during the standard development process, a full pilot test must be run on each version of the application using actual patients while they undergo actual treatment. This pilot study requires its own methods and strategies. In particular, the recruitment of recently diagnosed cancer patients is a challenge. It is a process that is impacted by stringent regulatory oversight, the limited resources of the providers, and the personal struggles of the patients and caregivers.

4.1 Pilot Test Design

The Phase I pilot study employed a single group, descriptive design to evaluate the system in terms of feasibility, acceptability, satisfaction and usability. The study recruited participants who were receiving chemotherapy at an NCI-designated cancer center. Based on each week's symptoms, the original interface generated evidence-based information tailored to each patient and sent notifications to providers when the symptoms reached a threshold level of severity. The gathered data was examined to identify patterns in user and system behaviors, providing a picture of how users interacted with the system and used the customized self-care guides to manage their healthcare. These research findings were used in the design of the Phase II PCMS.

As of this writing, the Phase II pilot study is just underway. It employs a single group, descriptive design, recruiting cancer patient-family member dyads. We hope for a sample that is large enough to correlate symptom levels with self-management behavior for at least the most common symptoms. This may be a difficult task, because the results of the Phase I pilot study demonstrated that the sample size of the symptoms is much smaller than the sample size of the participants. Even for the most common symptoms, not every patient has every symptom in any given week.

4.2 Methodology for Pilot Test Recruitment

For our Phase I pilot, gastrointestinal cancer patients who were seen in the clinics, and who were found to be both eligible and interested in the study by clinic staff, were introduced to the onsite study recruiter. After confirming eligibility, the recruiter consented and enrolled patients into the study. However, we found that identifying eligible patients and determining interest in the study quickly overburdened the clinic staff and resulted in low study accrual of 14 patients from 1-2009 to 11-2009. In addition, one participating nurse left during the study (7-2009). As a result, the one remaining nurse had difficulty facilitating recruitment due to an increased work load.

Consequently, during our Phase II pilot, we have sought to increase the effectiveness of our recruiting effort by allowing recruitment of patients with any solid tumor as long as participating nurses are caring for them. We also obtained a waiver of HIPAA authorization, for recruitment only, from the IUPUI Institutional Review Board (IRB). As a result, our clinic-based recruiters are now able to review the patient medical record for study screening purposes only, and identify eligible patients without burdening clinic staff. With permission from clinic staff to directly approach an eligible patient about the study, our clinic-based recruiters describe the study to patient/caregiver dyads and obtain written informed consent. Clinic nurses sign an informed consent for the study in order to receive patient symptom alerts from the system. To avoid confusion, one co-investigator for our study is responsible for explaining the study to the clinic nurses and obtaining their written informed consent.

This appears to be working. In 12 weeks of recruitment, 12 nurses and 12 of 50 planned patient/caregiver dyads have consented. We have had 1 patient not eligible due to visual impairment, 3 patients refuse due to feeling overwhelmed, 1 patient refuse due to no caregiver, and 3 caregivers refuse due to lack of interest.

However, Phase II has also had other complicating factors. Our Phase II pilot is a multi-site project with recruitment at one site and interviews at another. This multi-site approach required the development of a web-based encrypted administrative site for patient-identifiable demographic information (e.g., name, address, nurse) and a secure, web-based data capture system REDCap, allowing investigators at different sites immediate access to data. [8].

Because we have multi-site, yet distinct, study activities; we were required to have IRB approval at both IUPUI and MSU. These multiple IRB approvals slowed the start of recruitment by 5 months. Also, because recruitment is limited to patients of nurses who consent to the study, we cannot enroll patients if their nurse has not consented. Two nurses have refused the study because they are not interested in participating.

5 Description of the Patient and Caregiver Interfaces

Now that we have completed the next iteration of our design process, and begun our recruiting, we are currently engaged in the Phase II pilot study of the PCMS with the new caregiver interface included. This interface allows caregivers to receive updates on symptoms reported by the patient, and also to manage issues of their own. Some of the modules included in the new, PCMS are listed in Table 1.

Table 1. User interface modules for *Pathways for Caring* project's PCMS

The PCMS has several user interfaces:
1. Patient Symptom Self-Management Interface
a. Symptom Reporting Module
b. Custom Information Module
c. Full Information Module
d. Symptom History Module
e. Evaluation Module
2. Caregiver Self-Management Interface
a. Concern Reporting Module
b. Custom Information Module
c. Full Information Module
d. Patient Information Module
e. Evaluation Module
3. Provider Interface
4. Administrative Interface

5.1 General Design

The graphics and user-interface elements for the PCMS have been created with the awareness that the application will be used by people who are very sick and/or dealing with the stress of a family member's illness. The reporting/evaluation requirements for the application are complex, so much thought has been given to streamlining the user interaction to ease the cognitive load for people who are not feeling well.

The application's system architecture is designed to serve caregiver-patient dyads. The design assumes that the caregiver may need help in receiving current information about the patient's symptoms, either due to physical remoteness or to communication difficulties. Therefore, the application is designed to bridge this communication gap, but in such a way that it will still function if one partner or the other needs to drop out for a while. For example, the caregivers are shown the symptom severity ratings their partners have reported earlier in each week's sequence. If the patient did not report symptoms during the current week, the last available symptom report is used.

All log-ins to the system are done via a common gateway, and participants are sent to the appropriate module, based on their role and/or the task assigned by the Messaging/Scheduling Module. The Messaging/Scheduling Module uses a nightly script to determine when it is time for a participant to log in and do the weekly report of symptoms or concerns. Participants are sent an e-mail message, with the URL and instructions to log in to the *Pathways for Caring* website. In the initial week of the study, the study participants only report symptoms/concerns. In subsequent weeks the participants do an evaluation of the previous week's results before reporting the current week's symptoms/concerns.

5.2 Symptom Self-Management Interface

The Symptom Self-Management interface is for the chemotherapy patient in the caregiver-patient dyad. The patient receives a reporting reminder from the system a couple of days before the caregiver is notified. This gives the patient time to enter this week's symptom information before the caregiver logs into the system.

Symptom Reporting Module. This module has changed little since its Phase I implementation was described in Newlon, Hu, Stratton, and McDaniel [6]. Patients are asked to respond to questions about the severity of nine standard symptoms, with the option of adding more symptoms if needed.

The significance of this interface is the amount of coordinated action that it drives. The symptom types and levels reported determine what self-management information the system pulls from its repository and presents to the patient. In addition, however, the symptom levels determine whether the system notifies the provider that the patient is having problems; also, they are displayed during the caregiver's subsequent reporting session, allowing the caregiver to request appropriate symptom management information from the system; also, they eventually become dependent variables for researchers examining the effectiveness of the application as an intervention.

The symptom reporting interface has also been designed to help the providers with their reporting requirements. Symptom details are reported using a combination of numeric scales and multiple-choice descriptive statements. The descriptive statements in the multiple choice sections are intended to coincide with the National Cancer Institute Common Terminology Criteria for Adverse Events (CTCAE v3.0) [9] for each symptom, allowing the provider to receive the reportable toxicity level whenever the system sends a severe symptom alert to the provider (see Fig. 1).

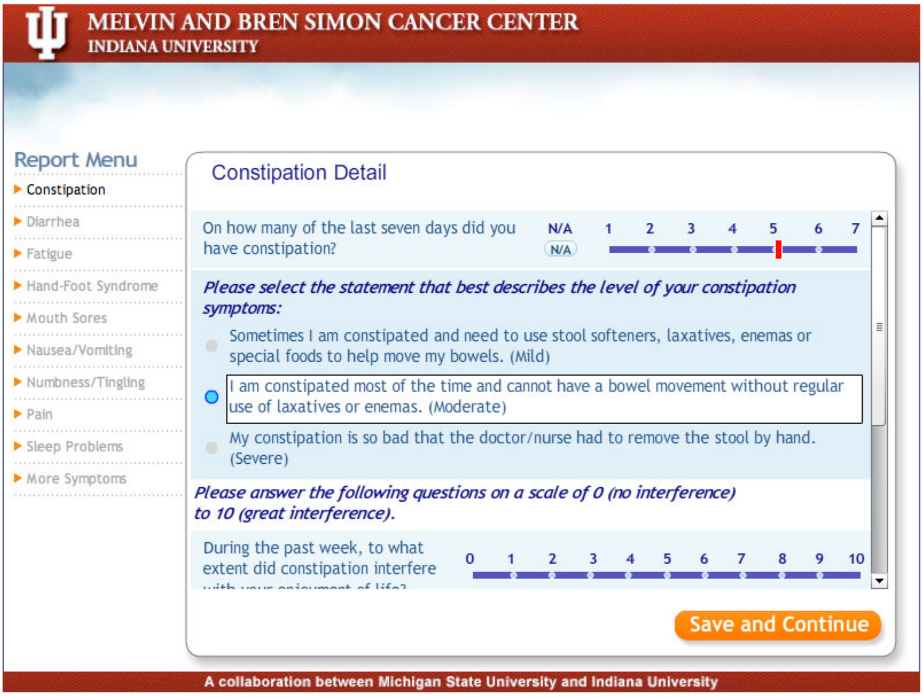


Fig. 1. The Symptom Self-Management interface's symptom reporting detail screen showing the toxicity criteria for constipation

5.3 Caregiver Self-Management Interface

The Caregiver Self-Management interface is for the caregiver in the caregiver-patient dyad. The caregiver receives a reporting reminder from the system a couple of days after the patient is notified. Again, this gives the patient time to enter the current week's symptom information before the caregiver logs into the system.

Concern Reporting Module. The Concern Reporting Module asks caregivers to report concerns about both patient symptoms and caregiver-centric issues.

The first reporting screen displays the most recent symptom severity report from the caregiver's patient (see Fig. 2) and asks the caregiver to indicate which symptom management information is needed. The date of the report is provided so the caregiver has a reference to the time and knows if the report is not from the current week.

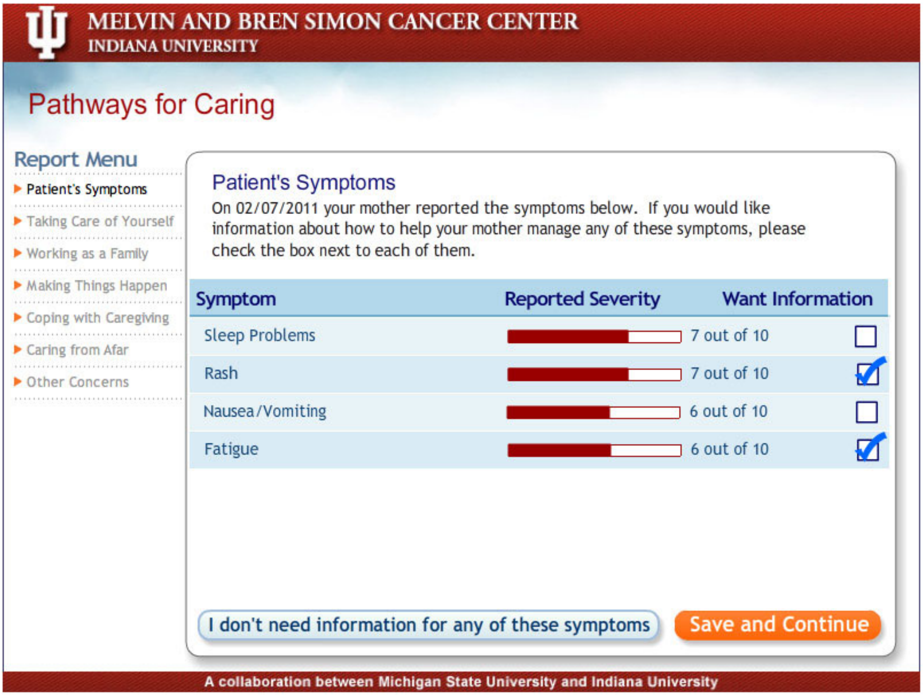


Fig. 2. The Caregiver Self-Management interface's Patient's Symptoms screen showing symptoms reported by the caregiver's hypothetical patient

On the next reporting screen, caregivers are asked to indicate whether they want or need information on various self-help topics. Caregivers living more than 50 miles from their partners also see a category for issues specific to long-distance care giving.

Custom Information Module. The PCMS records the caregiver's information requests and concerns and customizes information from the Symptom Management Toolkit and the Pathways for Caring Guide into a reference for the caregiver. The first section features any patient symptom information requests. In the following section, caregiver concerns that were ranked 3 or higher (out of 5) are displayed in order of highest rank to lowest. Each topic has practical advice and communication tips for friends, family, and health care providers. Fig. 3 shows an example of the Custom Guide.

5.4 The Provider Interface

The provider interface was designed to allow health care providers to respond to patient-generated alerts without the need to pass identifiable information through the alert e-mails. The alert tells the provider that a patient has generated an alert, but the patient is not identified. The project database keeps a list of each provider's patients, and patient alerts which have had no response. When the provider logs in, the patient list is displayed with outstanding alerts sorted to the top of the list. The provider can then click on the names to view the symptom reporting history for each patient.

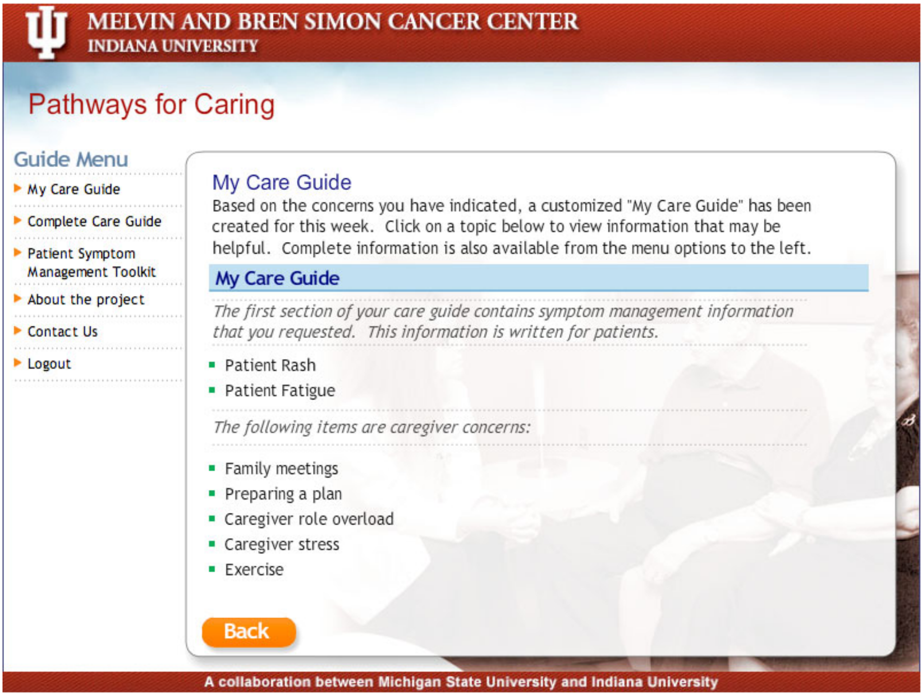


Fig. 3. Custom Guide screen showing links to self-help information about a hypothetical caregiver's concerns

6 Conclusion – Patient-Facing Smart Medical Interfaces

Given the broad demographic and the emotional frailty of those dealing with life limiting illness such as cancer, a well-tested interface is an imperative. This paper has covered the basic methodologies that we have used to develop and test patient-facing medical interfaces that are smart in the sense that they customize themselves based on the needs of the user. In particular, the *Pathways for Caring* project's *Patient-Caregiver Management System* provides an example of the potential for such interfaces to mediate interactions among such healthcare stakeholders as the patient, the caregiver, the provider, and the medical researcher. We would like to encourage further development of this sort of medical application. We also hold that a medical application design methodology such as the one herein described should be used as a general practice in the development of interfaces for health care applications.

Acknowledgments. We would like to acknowledge the contributions of Drs. Charles W. and Barbara Given of Michigan State University for their 30 years of research to help cancer patients and their families, upon which this research is based. Funding for this study is provided by a grant from the Walther Cancer Foundation, Indianapolis, IN.

References

1. Friedman, D.B., Hoffman-Goetz, L., Arocha, J.F.: Health literacy and the world wide web: Comparing the readability of leading incident cancers on the Internet. Med. Inform. Internet in Med. 31(1), 67–87 (2006)
2. Kreuter, M.W., Strecher, V.J., Glassman, B.: One size does not fit all: the case for tailoring print materials. Ann. Behav. Med. 21(4), 276–283 (1999)
3. Leydon, G.M., Boulton, M., Moynihan, C., Jones, A., Mossman, J., Boudioni, M., McPherson, K.: Cancer patients' information needs and information seeking behaviour: in depth interview study. BMJ 320(7239), 909–913 (2000)
4. Preece, J., Rogers, Y., Sharp, H.: Interaction Design: beyond human-computer interaction, 2nd edn. John Wiley & Sons Ltd, West Sussex (2007)
5. Sikorskii, A., Given, C.W., Given, B., Jeon, S., Decker, V., Decker, D., Champion, V., McCorkle, R.: Symptom management for cancer patients: a trial comparing two multimodal interventions. J. Pain Symptom. Manage 34(3), 253–264 (2007)
6. Newlon, C., Hu, C., Stratton, R., McDaniel, A.: Design of a Web-based Symptom Management Intervention for Cancer Patients. In: HCI International 2009, vol. 10, pp. 775–784. HCI, San Diego (2009)
7. Hewitt, M., Greenfield, S., Stovall, E. (eds.): From Cancer Patient to Cancer Survivor: Lost in Translation. National Academies Press, Washington (2005)
8. Harris, P.A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., Conde, J.G.: Research electronic data capture (REDCap) - A metadata-driven methodology and workflow process for providing translational research informatics support. J. Biomed. Inform. 42(2), 377–381 (2009)
9. Cancer Therapy Evaluation Program. Common Terminology Criteria for Adverse Events, Version 3.0, Bethesda, MD, National Cancer Institute (2009)

Importance of Physical Interaction between Human and Robot for Therapy

Takanori Shibata

National Institute of Advanced Industrial Science and Technology (AIST), Japan
shibata-takanori@aist.go.jp

Abstract. Mental health care of the elderly people is a common problem in advanced countries. Recently, high technology has developed robots for use not only in factories but also for our living environment. In particular, human interactive robots for psychological enrichment, which provide services by interacting with humans while stimulating their minds, is rapidly spreading. Such robots not only simply entertain but also render assistance, guide, provide therapy, educate, enable communication, and so on. Robot therapy, which uses robots as a substitution for animals in animal-assisted therapy and activity, is a new application of robots and is attracting the attention of many researchers and psychologists. The seal robot named Paro was developed especially for robot therapy and was used at hospitals and facilities for the elderly people in several countries. Recent research has revealed that robot therapy has the same effects on people as animal therapy.

Keywords: Mental Commitment Robot, Robot Therapy, Human–Robot Interaction, Elderly Care.

1 Introduction

Mental health care of elderly people is an important issue for caregivers at nursing homes [1]. Elderly people are depressed easily, as they experience difficult situations such as the loss of their family, friends, social roles, and physical functions. Depressive disorders are common among elderly people in nursing homes [2]. In addition, there are the people who suffer from other mental disease like dementia that causes various psychiatric and behavioral disturbances, such as hallucinations, aggression and wondering [3]. Such disorders have a negative impact on the quality of life of both elderly people and their caregivers. Therefore, while trying to communicate with the elderly people, the caregivers conduct several recreational activities, such as singing songs, coloring, drawing pictures, and origami. However, some people are too embarrassed to sing songs, and others because of their illness have difficulty moving their fingers when they try to draw. In addition, the caregivers might find communication with the elderly difficult because of lack of common topics.

Interaction with animals has long been known to be emotionally beneficial to people. In recent years, the effects of animals on humans have been researched and proved scientifically. Friedmann investigated the one-year survival rate of patients

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 437–447, 2011.

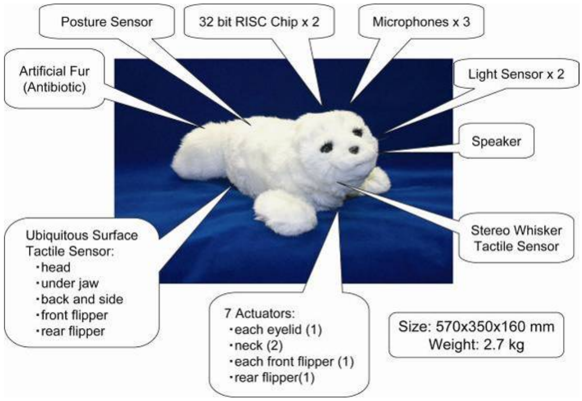


Fig. 1. Seal Robot, Paro

who were discharged from a coronary-care unit and found that survival among those who kept pets was higher than those who did not [4]. Baun et al. reported that the patient's blood pressure dropped when they petted their dog [5]. Garrity et al. studied elderly people who were socially isolated and had lost their partner within the previous year, and they found that the intensity of depression among those who had no pets was higher than those who did [6]. Lago et al. investigated through telephonic interviews, the influences of owning pets on elderly people. They revealed that mortality and attrition were higher for former owners than current owners [7]. Hart et al. studied the social influences of animals on people, and found that the number of friendly approaches by strangers to people with dogs were greater than to people without dogs [8].

In medical applications, especially in the United States, animal-assisted therapy and activities (AAT&AAA) are becoming widely used in hospitals and nursing homes [9, 10]. AAT has clear goals set out in therapy programs designed by doctors, nurses, or social workers, in cooperation with volunteers. On the other hand, AAA refers to patients interacting with animals without particular therapeutic goals and depends on volunteers. AAT and AAA are expected to have three effects:

1. *Psychological effect (e.g., relaxation, motivation)*
2. *Physiological effect (e.g., improvement of vital signs)*
3. *Social effect (e.g., stimulation of communication among inpatients and caregivers)*

For example, a hospitalized child, who was in significant pain because of his disease, was afraid to get up and walk around. However, when he was asked to take a therapy dog for a walk, he immediately agreed and walked off happily, as if all his pain had diminished. Moreover, the dog acted as a medium for interaction between him and the other children [11]. In another case, a boy who was born as a crack-exposed baby could not speak and walk. However, through interaction with therapy dogs and birds, both his linguistic and motor abilities improved [12].

For AIDS patients, it is important to reduce their stress as it is strongly related to the complications of immune deficiency. AAT helps them relax and stay connected with the world [13].

In addition to these effects, AAT and AAA at nursing homes provide rehabilitation to the elderly people, and offer laughter and joy to a patient who has a short time to live [14]. AAT reduces loneliness in residents of long-term care facilities [15]. The presence of therapy animals particularly has been useful in reducing agitated behavior, decreasing episodes of verbal aggression and anxiety, and increasing social interaction in institutionalized elderly people suffering from dementia [16-18].

However, most hospitals and nursing homes, especially in Japan, do not allow animals, although they admit the positive effects of AAT and AAA. They are afraid of the negative impact of animals on human beings, such as allergic reactions, infection, bites, and scratches.

This mini review introduces robot therapy and discusses its potential to care for the elderly people. The next section, describes the new robotics area, human-interactive robots for psychological enrichment. We then explain the required functions for therapeutic robots and the seal robot, Paro. Finally, we provide examples of robot therapy and the conclusions.

2 Human-Interactive Robots for Psychological Enrichment

Industrial robots have been used widely in manufacturing industries since the early 1960s. The industrial robots typically perform welding, assembling, painting, packaging, and palletizing in the automotive manufacturing and other industries. Such robots work very fast and with accuracy, although initially they need to be taught by a human operator and their environment need to be specially designed for them to accomplish their tasks. Most industrial robots are considered as potential danger to humans, and therefore are kept isolated from people.

Meanwhile, the rapid development in high technology has produced robots not only for factories but also for our living environment, such as homes, offices, and hospitals. For example, wheelchair robots enable elderly people to easily move outside [19]. Robot suits, which can expand physical capability of humans, are expected to reduce the workload of caregivers [20]. A horseback-riding robot promotes patient's physical strength [21]. Human-interactive robots for psychological enrichment, in particular, are expected to be a new application of robotics and are attracting many researchers and companies [22]. Human-interactive robots are designed for entertainment, communication (social activity), guidance, education, welfare, mental therapy, and other purposes. Various types of robots, such as humanoid, animal, and robots with unique appearance, have been developed.

These robots offer more interaction with humans than industrial robots. They are evaluated not only in terms of objective measures, such as speed and accuracy, but also in terms of subjective measures for interacting with humans, such as providing comfort and bringing joy. Robots for entertainment are good examples of the importance of a subjective evaluation of their value.

There are four categories of human interactive robots for psychological enrichment, in terms of their relationship with humans: (1) performance robots; (2) tele-operated performance robots; (3) operation, building, programming, and control robots; and (4) interactive autonomous robots.

1. Performance robots

Performance robots have a long history and they execute movements that express meanings to humans, mostly for fun. Mechanical puppets that could play an organ, draw pictures, and write letters were developed in Switzerland in the 18th century. Karakuri dolls were developed in Japan during the same era to perform dance, magic, and so on. Recently, many performance robots have been used at exhibitions, museums, movies, and amusement parks such as Disneyland and Universal Studios. Recent humanoid robots, such as Honda's ASIMO and Sony's QRIO, can be included in this category [23, 24]. A performance robot can amuse a sizable audience at any time. However, their movements will probably be preprogrammed and mostly repetitive; therefore, they are not usually very interactive with humans. A high degree of complexity in the performance robots is required to keep humans amused.

2. Tele-operated performance robots

Tele-operated performance robots are controlled remotely by a hidden operator. Their movements can appear reactive to the humans who interact with them because the operator, based on the audience's actions, sends commands to the robots to simulate reactive behavior. At exhibitions or amusement parks, for example, human-type robots are used as tele-operated performance robots.

3. Operating, building, programming, and controlling robots

Humans derive a lot of fun and joy from operating, building, programming, and controlling robots. Moreover, one can watch the performance of the robot that he or she is operating. A simple example of this is the "UFO catcher," a stuffed-animal game machine at amusement centers. Building and programming a robot is also included in this category. Contests between robots such as Micro-mouse, RoboCup (robot football), and RoboOne (robot wrestling) are popular examples [25, 26], as are LEGO-Mindstorms and I-Blocks. Because building and programming robots can stimulate children's creativity, this activity combines entertainment with education, and is often referred to as "edutainment" [27, 28].

4. Interactive autonomous robots

Interactive autonomous robots connect with humans in the physical world. They use verbal and nonverbal communications, depending on the functions of the robots. Contrary to the robots in the other categories, the human-robot interactions are mostly personal. For example, Sony's dog robot, AIBO, which is designed for entertainment, has a mechanical appearance and attracts people's interest using nonverbal communication [29]. The communication robot, ifbot, produces conversation using facial expressions and a large number of prepared conversation scenes [30]. The human friendly information terminal, PaPeRo, can control home electric appliances and collect information via the Internet by voice command, and even entertain people by dancing and playing games [31]. Guide robots in museums and exhibitions [32], and mental commitment robots described in this paper also belong to this category.

In the area of welfare and mental therapy, Shibata et al. studied and developed a mental commitment robot, which aims to engender mental effects, such as pleasure and relaxation, in its role as a personal robot [22, 33-49]. They also propose a robot therapy, which uses robots as substitutes for animals in animal-assisted therapy and activity. Robot therapy targets the people in medical and welfare institutions where animals are not allowed. A seal-type mental commitment robot, named Paro (Fig. 1), was developed especially for robot therapy and is used at pediatric hospitals and facilities for the elderly people in several countries. Recent research has revealed that robot therapy has the same effects on people as animal therapy. Robot therapy, in particular, is recognized as a new method of mental health care for the elderly people (including dementia patients).

3 Therapeutic Robot

3.1 Required Functions

In robot therapy, it is important to stimulate people's knowledge and experiences of animals through interaction with the robots and bring out their feelings when they are interacting with animals. Therefore, the shapes, feelings of touch, autonomous behaviors, and responses that mimic animals are the features that are required to be present in the robots.

In addition, the devices are used not only at people's homes but also at hospitals and nursing homes. Many people who have lost their physical strength and healing capability due to aging and illness are expected to interact with the robots. Therefore, the robots should be easily accepted by people and also be harmless and hygienic. A concern with such robots is that individuals are expected to physically interact by touching and hugging them, and therefore there exists a possibility of people being harmed. Furthermore, in case of the robot being used by people with deteriorated immune systems such as leukemia patients in hospitals, the robot may possibly transmit germs. Furthermore, some people visit medical welfare facilities for a few hours at a time for day care and ambulant treatment, but some stay or are hospitalized for years (e.g., nursing homes and long-term care). Therefore, the robots over the long-term have to sustain interaction with people in their daily lives.

These robots would be used by doctors, nurses, therapists, caregivers, and volunteers during a certain period of time. In addition, the users would play with them whenever they want. Therefore, it is important that these robots be designed in such a manner that anyone can operate them, and no specialized knowledge is required to do so.

3.2 Mental Commitment Robot, Paro

Mental commitment robots are not intended to offer people physical work or service [22, 33–49]. Their function is to engender mental effects, such as pleasure and relaxation, in their role as personal robots. These robots act independently with a purpose and with motives while receiving stimulation from the environment, as with living organisms. Actions that manifest themselves during interactions with people can be interpreted as though the robots have hearts and feelings.

Mental commitment robots can stimulate the different senses of human beings through physical interaction. Therefore, the primary characteristic of mental commitment robots is nonverbal communication. A basic psychological experiment was conducted on the subjective interpretation and evaluation of robot behavior following human-robot interactions. This study showed the importance of appropriately stimulating the human senses and extracting associations. Sensor systems, such as visual, aural, and tactile senses for robots, were studied and developed. A plane soft tactile sensor was developed to cover the robot to enhance bodily contact between people and the robot [42]. This sensor can detect position and force when people touch the robot, and at the same time, it is soft to touch.

The shapes of the animal robots are classified into three categories:

i. Familiar animals (e.g., dog, cat)
ii. Non-familiar animals (e.g., seal)
iii. Imaginary animals or characters

The dog, cat, and seal robots were developed as models. The robot operates using the three elements of its internal states, sensory information from its sensors, and its own diurnal rhythm (morning, daytime, and night) to perform various activities during its interaction with people.

Subjective evaluations of the cat and seal robots were conducted using questionnaire [35, 36]. People evaluated both robots high. However, the subjects complained about the softness and the reactions of the cat robot in comparison with their knowledge of real cats. On the other hand, most people do not have much knowledge about seals, and hence were unable to compare the seal robot with their knowledge of seals. Therefore, the seal robot's evaluation was higher after the interaction. These results revealed that more people accepted the unfamiliar animal shape. Cross-cultural studies on the subjective evaluation of the seal robot were conducted in seven different countries: Japan, the U.K., Sweden, Italy, Korea, Brunei, and the U.S. [40, 49]. The data were obtained from over 1,800 respondents. The subjective evaluation provided overall high scores, and further revealed that the seal robot could be widely accepted despite cultural and religious differences.

The seal-type mental commitment robot, Paro, was designed for therapy. For this purpose, it was functionally designed to be soft and evoke a feeling of warmth. Each seal robot was trimmed in artificial fur and eyelashes were sewn on it by craft workers to achieve high quality. In addition, its artificial fur was antibacterial and, dirt resistant, and would not fall out. An electromagnetic shield was provided to the internal circuit to prevent any ill effect on heart pacemakers. The withstand voltage test, drop test, one hundred thousand times stroking test, and a long-term clinical test over the years confirmed that Paro is highly safe and durable. It was designed to be simple enough for anyone to be able to operate. Paro has only one on/off switch for power, and a pacifier-type charger. Learning the functions of its name and behaviors allows its users to gradually build a relationship with it, thus preventing them from losing interest and in turn encouraging them to show their affection for Paro. In addition, the baby harp seal was ecologically investigated to model its liveliness and cuteness in the robot; actual baby seal calls were sampled and used.

3.3 Robot Therapy

Robot therapy using the seal robot is conducted at hospitals and nursing homes in many countries: Japan, Sweden, Denmark, Italy, the U.S., and etc. Robot therapy consists of robot-assisted therapy programs designed by doctors, nurses, and social workers, and also robot-assisted activity, which allows patients to interact with robots without any particular therapeutic goals. Such activities do not depend on volunteers, but are conducted by facility staff. Robot-assisted therapy is mainly conducted at medical facilities, such as hospitals and clinics, while robot-assisted activity is conducted at welfare facilities, such as nursing homes (Figs. 2 and 3).

(a) Speaking to and stroking Paro

(b) Hugging and kissing Paro

Fig. 2. Interaction between elderly people and Paro

Fig. 3. Social mediator, Paro, between elderly people and caregiver

3.4 Robot Therapy for Elderly People

As an example for robot therapy, Paro was used to assist the elderly people at a day service center [38, 39, 43]. To investigate the effects of seal robots on the elderly people, their mood was evaluated using face scales [50] and questionnaires. Changes in their reactions to stress were measured by the hormones in urine: 17-ketosteroid sulfate (17-KS-S) and 17 hydroxycorticosteroids (17-OHCS) [51, 52]. In addition, the stress that the nursing staff experienced was investigated by questionnaires, i.e., the burnout scale [53]. The day service center was provided with seal robots for five weeks, as a result the feelings of the elderly people improved by their interaction with the robots. Urine samples showed that their ability to overcome stress also improved. Moreover, the stress levels of the nursing staff decreased because the elderly people required less supervision while interacting with the robots.

A long-term experiment was conducted at a health service facility since August 2003 [45]. Approximately 10 people joined the interaction with Paro for 1 hour, twice a week. One or two caregivers managed the interaction with Paro. To investigate the effects of Paro, face scale and geriatric depression scales [54] were used. The results showed that the feelings of the elderly people improved over the year, and depression in the participants also was reported to have reduced. The caregivers commented that interaction with Paro made the elderly people laugh and become more active. Their facial expressions changed, softened, and brightened. On the day of the activity, they looked forward to interacting with Paro, sitting down in their seats even before the interaction session began. Some people who usually stayed in their rooms came out and willingly joined the activity. In addition, Paro encouraged people to communicate with each other as well as with the caregivers by becoming their common topic of conversation. Thus, the general atmosphere became brighter. Even now, these elderly people enjoy playing with Paro.

In another example, Paro was introduced in the public area of a care house, a type of communal housing for the elderly people, and was activated for over nine hours each day for the researchers to investigate the effects of free interaction with it [47]. To examine the psychological and social effects, each subject was interviewed and his or her social network analyzed. In addition, the activities of the residents in public areas were video recorded. For physiological analysis, residents' hormones in urine; 17-KS-S and 17-OHCS, were analyzed. The results indicate that interaction with Paro increased their social interaction. Furthermore, the urine tests showed that the reactions of the subjects' vital organs to stress improved after interacting with Paro.

Meanwhile, a dementia care center in Copenhagen, Denmark, investigated the effects of robot therapy on dementia patients as part of the national project “Be-Safe.” Twelve Paros were introduced to 10 different places. The results obtained from the seven-month clinical trial showed that Paro had positive effects on these patients. Based on the results, the Danish government decided to introduce 1000 units of Paros to nearly all of the elderly care facilities in Denmark.

4 Conclusion

Various robots have been developed and are being introduced in our lives as commercial products. Each robot is designed for a specific purpose. The seal-type mental commitment robot, Paro, whose goal is to enrich daily life and heal human minds, is designed to maintain long-term interaction with people and provide them with psychological, physiological, and social benefits.

Robot therapy, mental health care using animal-type robots, is an emerging field. The results of exploratory experiments showed that Paro has a great potential to provide mental health care for elderly people. However, more subjects and a control group are necessary to scientifically verify its effects. Further experiments will be conducted in this regard.

References

1. Daies, B., Knapp, M. (eds.): Old People's Homes and the Production of Welfare. Routledge & Kegan Paul PLC (1981)
2. WPA/PTD Educational Program on Depressive Disorders, Module 3: Depressive Disorders in Older Persons, `http://www.wpanet.org/education/ed-program-guidelines.shtml`

3. Alzheimer's Disease Interanational: Psychiatric and Behavioural Disturbances in Dementia, p. 7. ADI Factsheet (1999)
4. Friedmann, E., et al.: Animal Companions and One-year Survival of Patients after Discharge from a Coronary Care Unit. Public Health Rep. 95(4), 307–312 (1980)
5. Baun, M., Bergstrom, N., Langston, N., Thoma, L.: Physiological Effects of Human/Companion Animal Bonding. Nurs. Res. 33(3), 126–129 (1984)
6. Garrity, T., Stallones, L., Marx, M., Johnson, T.: Pet Ownership And Attachment As Supportive Factors In The Health of The Elderly. Anthrozoos 3(1), 35–44 (1989)
7. Lago, D., et al.: Companion Animals, Attitudes Toward Pets, and Health Outcomes Among the Elderly: A Long-Term Follow-up. Anthrozoos 3(1), 25–34 (1989)
8. Hart, L., Hart, B., Bergin, B.: Socializing Effects of Service Dogs for People with Disabilities. Anthrozoos 1(1), 41–44 (1987)
9. Standards of Practice for Animal-Assisted Activities and Therapy. Delta Society (1996)
10. Fine, A.H.: Handbook on Animal Assisted Therapy: Theoretical Foundations and Guidelines for Practice, 2nd edn. Academic Press, London (2006)
11. Kale, M.: Kids & Animals. Inter Actions 10(3), 17–21 (1992)
12. Animal-Assisted Therapy and Crack Babies: a New Frontier. Delta Society Newsletter 1(2) (1991)
13. Haladay, J.: Animal Assisted Therapy for PWAs – Bringing a Sense of Connection. AIDS Patient Care, 38–39 (1989)
14. Gammonley, J., Yates, J.: Pet Projects Animal Assisted Therapy in Nursing Homes. J. Gerontol. Nurs. 17(1), 12–15 (1991)
15. Marian, R., et al.: The Effects of Animal-Assisted Therapy on Loneliness in an Elderly Population in Long-Term Care Facilities. J. Gerontol. A: Biol. Sci. Med. Sci. 57, 428–432 (2002)
16. Richeson, N.E.: Effects of AAT on agitated behaviors and social interactions of older adults with dementia. Am. J. Alzheim Dis. Other Dement 18(6), 353–358 (2003)
17. Fick, K.M.: The influence of an animal on social interactions of nursing home residents in a group setting. Am. J. Occupat. Ther. 47(6), 529–534 (1993)
18. Fritz, C.L., et al.: Association with companion animals and the expression of noncognitive symptoms in Alzheimer's patients. J. Nerv. Ment. Dis. 183(7), 459–463 (1995)
19. Matsumoto, O., Komoriya, K., Hatase, T., Nishimura, H.: Autonomous Traveling Control of the "TAO" Aicle Intelligent Wheelchair. In: Proc IEEE/RSJ Int. Conf. IROS 2006, pp. 4322–4327 (2006)
20. Web site of robotic suit, HAL, `http://www.cyberdynejp`
21. Web site of horseback riding robot, JOBA, `http://www.panasonicjp/fitness`
22. Shibata, T.: An Overview of Human Interactive Robots for Psychological Enrichment. Proceed. IEEE 92(11), 1749–1758 (2004)
23. Hirai, K.: Humamoid Robot and Its Applications. In: Proc. IARP Int. Conf. Humanoid and Human Friend Robot, vol. 1, pp. 1–4 (1998)
24. Kuroki, Y., et al.: A Small Biped Entertainment Robot. J. Robot. Mechatro. 14(1), 7–11 (2002)
25. Kitano, H. (ed.): RoboCup 1997. LNCS, vol. 1395, pp. 62–73. Springer, Heidelberg (1998)
26. Robo-One Offical Home Page, `http://www.robo-one.com`
27. Druin, A., Hendler, J. (eds.): Robots for Kids; Exploring New Technologies for Learning. Morgan Kaufmann Pub., San Francisco (2000)

28. Lund, H.: Modern Artificial Intelligence for Human-Robot Interaction. Proceed. IEEE 92(11), 1821–1838 (2004)
29. Fujita, M.: On Activating Human Communications With Pet-Type Robot AIBO. Proceed. IEEE 92(11), 1804–1813 (2004)
30. Web site of ifbot, `http://www.business-designcojp`
31. Web site of PaPeRo, `http://www.incxneccojp/robot/robotcenterhtml`
32. Bischoff, R., Graefe, V.: HERMES – a Versatile Personal Robotic Assistant. Proceed. IEEE 92(11), 1759–1779 (2004)
33. Shibata, T., Inoue, K., Irie, R.: Emotional Robot for Intelligent System - Artificial Emotional Creature Project. In: Proc. IEEE RO-MAN, pp. 466–471 (1996)
34. Shibata, T., Irie, R.: Artificial Emotional Creature for Human-Robot Interaction - A New Direction for Intelligent System. In: Proc. IEEE/ASME Int. Conf. AIM 1997, #47 (1997)
35. Shibata, T., et al.: Subjective Interpretation of Emotional Behavior through Physical Interaction between Human and Robot. Proc. IEEE Int. Conf. SMC, 1024–1029 (1999)
36. Shibata, T., Tanie, K.: Influence of A-Priori Knowledge in Subjective Interpretation and Evaluation by Short-Term Interaction with Mental Commit Robot. In: Proc. IEEE/RSJ Int. Conf. IROS, vol. 1, pp. 169–174 (2000)
37. Shibata, T., et al.: Mental Commit Robot and its Application to Therapy of Children. In: Proc. IEEE/ASME Int. Conf. AIM, pp. 1053–1058 (2001)
38. Shibata, T., Wada, K., Saito, T., Tanie, K.: Robot Assisted Activity for Senior People at Day Service Center. In: Proc. Int. Conf. ITM, pp. 71–76 (2001)
39. Saito, T., Shibata, T., Wada, K., Tanie, K.: Examination of Change of Stress Reaction by Urinary Tests of Elderly before and after Introduction of Mental Commit Robot to an Elderly Institution. In: Proc. Int. Symp. AROB, vol. 1, pp. 316–319 (2002)
40. Shibata, T., Mitsui, T., Wada, K., Tanie, K.: Subjective Evaluation of Seal Robot: Paro - Tabulation and Analysis of Questionnaire Results. J. Robot. Mechatron. 14(1), 13–19 (2002)
41. Wada, K., Shibata, T., Saito, T., Tanie, K.: Psychological and Social Effects in Long-Term Experiment of Robot Assisted Activity to Elderly People at a Health Service Facility for the Aged. In: Proc. IEEE/RSJ Int. Conf. IROS, pp. 3068–3073 (2004)
42. Shibata, T.: Ubiquitous Surface Tactile Sensor. TExCRA 56 (2004)
43. Wada, K., Shibata, T., Saito, T., Tanie, K.: Effects of Robot Assisted Activity for Elderly People and Nurses at a Day Service Center. Proceed. IEEE 92(11), 1780–1788 (2004)
44. Marti, P., Palma, V., Pollini, A., Rullo, A., Shibata, T.: My Gym Robot. In: Proc. Symp. Robot Companions, pp. 64–73 (2005)
45. Wada K., et al.: Psychological and Social Effects of One Year Robot Assisted Activity on Elderly People at a Health Service Facility for the Aged. In: Proc. IEEE ICRA, pp. 2796–2801 (2005)
46. Marti, P., et al.: Socially Assistive Robotics in the Treatment of Behavioural and Psychological Symptoms of Dementia. In: Proc. Int. Conf. Bio. Rob., pp. 483–488 (2006)
47. Wada, K., Shibata, T.: Living with Seal Robots— Its Socio- psychological and Physiological Influences on the Elderly in a Care House. IEEE Trans. Robot. 23(5), 972–980 (2007)
48. Wada, K., Shibata, T., Musha, T., Kimura, S.: Robot Therapy for Elders Affected by Dementia. IEEE Eng. Med. Biol. Mag. 27(4), 53–60 (2008)
49. Shibata, T., Wada, K., Ikeda, Y., Sabanovic, S.: Cross-cultural Studies on Subjective Evaluation of Seal Robot. Advanced Robotics (2009) (in press)

50. Lorish, C.D., Maisiak, R.: The Face Scale: A Brief, Nonverbal Method for Assessing Patient Mood. Arthritis and Rheumatism 29(7), 906–909 (1986)
51. Selye, H.: Stress and aging. Journal of American Geriatric Society 18, 669–676 (1970)
52. Nishikaze, O., et al.: Distortion of Adaptation (Wear & Tear and Repair & Recovery)-Urine 17- KS-Sulfates and Psychosocial Atressin. Humans-Job Stress Res. 3, 55–64 (1995)
53. Pines: The Burnout measure. Paper Presented at the National Conference on Burnout in the Human Services, Philadelphia (November 1981)
54. Yesavage, J.A.: Geriatric Depression Scale. J. Psychopharmacology Bulletin 24(4), 709–711 (1988)

The Relationships between Morphology and Work for the Nursing Performance of Hand Controls in Emergency Surgery

Chao-Yuan Tseng and Fong-Gong Wu

Department of Industrial Design, National Cheng Kung University, Tainan, Taiwan
{P38991034,fonggong}@mail.ncku.edu.tw

Abstract. Coming from the point of emergency room environment, this research studies the motion behavior, the control panel conformation, and the virtual operating movements of surgical nursing staff in the emergency room, in order to provide a reference for instrument design. The research began with professional interviews, followed by experiments on hand motion of surgical nursing staff in the emergency room. Test participants consist of surgical nursing staff from emergency rooms and the surgical department. A total of 20 participants, 10 females from each department were chosen to take the tests. The research is divided into three categories. Results of this study presented a difference in operation height range between different control panel directions amongst emergency room and non-emergency room surgical nursing staff. From the results we can see that participants with the same height do not necessary present the same wrist angles.

Keywords: emergency surgery, nursing, hand controller, motion analysis, virtual command.

1 Introduction

The Emergency Medical Services System in Taiwan, abbreviated EMSS, was found since 1989. From the results of the Average Daily Medical Services of Taiwan Hospitals – by Ownership, Department of Health, Executive Yuan, we can see that the daily average of emergency room patients in hospitals around Taiwan is close to 13,677 people. Li[19] in terms of emergency room facilities, close contacts are setup between the examination counter contact systems, such as telephones and intercoms, and the emergency vehicle, in order to be ready for varies emergency operations. Armstrong[1] presented results that proved different hand postures causing different levels of injury; Kroemer[14] also claimed a varies hand motions that cause long-term injuries. Amongst which, long-term fatigue usually causes excessive work of hand muscles during facility operations, including excessive strain or contraction; long-term pressure around the hands can lead to Carpal Tullnel Syndrome (Armstrong[1]), Tendinitis or Tenosynovitis (Kroemer[14]). This research can be explained in the following aspects: (1) In terms of research procedures: a) Constructing static hand

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 448–457, 2011.

testing data of emergency room surgical nursing staff and other related staff for the data base. b) Recognizing hand movement relationships during facility operation by emergency surgical nursing staff, further understand and analyze the relative hand relationship using motion analysis equipment. c) Understanding the distribution of hand operation movement groups of emergency room nursing staff. (2) In the aspect of future research and development: a) Constructing database for the static hand data of nursing staff in Taiwan. The procedures and methods of this research may be helpful to the prototype construction of hand data of nursing staff in Taiwan. b) Promoting local and independent medical equipment design.

Barnes, R. M.[2] mentioned that the range of wrist movement is roughly 80°~90°. When the back is bent sideways, extensor carpi radialis longus muscles, brevis, and ulnaris muscle have a movement range of about 50°. During ulnar flexion, the ulnar muscle group effect creates a movement range of about 60°. During radialis flexion, the radialis muscle group effect creates a movement range of about 30°.

McCormick & Sanders[15] categorized the descriptions of movements during activities into the following categories: Positioning Movements, Continuous Movements, Manipulative Movements, Repetitive Movements, Sequentia1 Movements, and Static Adjustment. Hence, according to the hand movement features described by the above scholars, this research thoroughly examine the hand control conditions through observation. At this point, the research group categorizes the results of McCormick and Sanders, and Hsu, Sheng-Hsiung *et al.*[18] studies, and generalizes into Operation Motion Analysis of Emergency Surgical Nursing Staff, including movements of extension, pressing, rubbin, pinching, and twisting.

Sperling *et al.*[16] found in their research that the longer the operation time, the higher the probability of injury occurrence. Operating a tool in the same motion for more than 4 hours a day or 30 minutes continuously is likely to cause accumulative injury to the hands; while operating for less than an hour per day or 10 minutes continuously will reduce the chance of accumulative injuries. Also, according to the results of our observations, most of the controllers in the emergency surgical department consist of rotation and button controllers. Hence, it is vital to study and observe the relationship between hand operation motions and controller formations, and the effects on possible injuries caused by movements.

The adduction and abduction of shoulders should be under 45°, and angle of more than 90° should be avoided. Barnes, R. M.[2] claimed in his research, that hand movements can be categorized into five levels according to the hand movement range and energy consumption: level 1 includes finger movements; level 2 includes finger and wrist movements; level 3, 4, and 5 includes forearm, upper arm, and other body part movements consecutively. Categorizing emergency surgical nurses according to these levels, it is found that emergency surgical nurses fall into level 5, which is large movement range and energy consumptions. Being in this level also meant that other body parts are required to move on top of hand movements. Many other scholars and organizations have done similar controller appearance recognition studies, analyzing in detail the relationship between hand and human-machine interfaces.

2 Method

The entire experiment procedure will be carried out in the human engineering laboratory of our faculty, with the air of 3D motion analysis system to derive data. In the aspect of research investigation, the information regarding emergency surgical nursing members will be collected in two parts. The first being the professional interview; interviews are conducted with emergency surgical staff in major medical facilities around the Tainan area (Tainan, hospital, National Cheng Kung University hospital, Chi Mei medical center), to investigate the types of instruments and the possible instruments employed during emergency surgical situations. The second part being the hand movement experiment of emergency surgical nursing staff; the experiment instruments are chosen in respect of the instrument and control categories investigated in the first part of the information collection.

2.1 Participant Numbers

The participants were chosen to be 10 female emergency surgical nurses and 10 female non-emergency surgical nurses from the major medical facilities in the Tainan area. A total of 20 female nursing staff with no handicap in their hands and their right hands being the preferred hand will be participating in the experiment.

2.2 Record and Measurement Items

Principle independent variables: (1) Controller formations. (2) Body measurement values; randomly selected without any limitations. (Measurements include height, palm length and width, shoulder width, elbow height, and hand length). (3) Controlling directions (categorized by left, center, and right with a 45cm panel width). (4) Actual and virtual controls. (5) Nursing staff categories (Emergency surgery and non-emergency surgery). (6) Relative height during instrument operation. Secondary independent variables include (1) Procedure related variable. Experiment procedures and steps are explained prior to the experiment; participants are given a chance to adapt to controller formations and heights; a one-minute practice is given to participants for each controller formations, plus a three-minute recess before the formal experiment. (2) Environment variable. In order to avoid other light sources affecting the image deriving, all lights in the laboratory were turned off and the window covered with black curtains during the experiment. Room temperature was maintained at around 25℃ in order to ensure the most comfortable experiment condition for the participants. (3) Participant related variable. a) Participants were chosen from female, right-handed emergency and non-emergency surgical nursing staff, able to endure 30 minutes of experiment process. b) Comfortable and dark colored clothing was required for the experiment (to avoid reflections caused by light colors, creating error in data deriving). Independent variables: (1) The diameter range of fingertips grabbing the controller. (2) The difference between operating angle and panel angle. (3) Operating height during virtual operations. (4) Wrist joint angles formed (flexion and extension angles).

2.3 Experiment Hypothesis

Hypothesis 1: The wrist joint angle when operating instruments standing (flexion angle) is dependent on the height and elbow height. This means when the hand comes in touch with the instrument under normal standing positions (fixed instrument height position), the taller the person is, the larger the wrist angle and vice versa.

Hypothesis 2: In the emergency and non-emergency surgical nursing staff operating under actual and virtual situations from three directions (left, central, right), the angles formed under virtual situations would appear to have bigger differences than angles formed under actual situations (panel angle of 66°).

Hypothesis 3: Under virtual operation situations, the relative hand height would be taller under actual operation situations.

2.4 Data Analysis Method

For the convenience of data analysis, hereby describe the joint angle calculation process. (1) The space dot product mode can be used to calculate the angle presented by the wrist and the control panel (*Figure* 1、Formula 1.) (2) The space between two points can be used to calculate the diameter between thumb and finger pinching the controller, in order to obtain the controller diameter of each emergency surgical nursing staff during actual and virtual controlling. (*Figure* 2、Formula 2) (3) To calculate the consistency between hand and panel angles during control, the included angle θ could be obtained by assuming finger control as the XY plane created by the experiment panel top. (Formula 3).

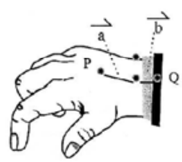
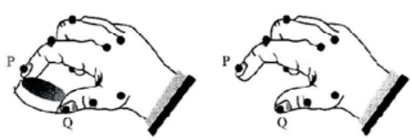


Fig. 1. Wrist angle calculation hypothesis

Assuming the space vector formed by P and Q are $\vec{a}$、$\vec{b}$ respectively, then

$$\alpha = \cos^{-1}\left(\frac{\vec{a}\ \vec{b}}{\overrightarrow{|a|}\ \ \overrightarrow{|b|}}\right) \tag{1}$$



Actual control Virtual control

Fig. 2. Actual control and virtual control hypothesis

Assuming P(a_1、b_1、c_1)、Q(a_2、b_2、c_2) then

$$\overline{PQ} = \sqrt{(a_2 - a_1)^2 + (b_2 - b_1)^2 + (c_2 - c_1)^2} \tag{2}$$

Assuming the included angle between the XY plane equation $a_1X + b_1Y + c_1Z = d_1$ and the plane equation created by the thumb, middle finger, and the index finger $a_2X + b_2Y + c_2Z = d_2$ is θ, then

$$\theta = \cos^{-1}\frac{a_1a_2 + b_1b_2 + c_1c_2}{\sqrt{a_1{}^2 + b_1{}^2 + c_1{}^2}\sqrt{a_2{}^2 + b_2{}^2 + c_2{}^2}} \tag{3}$$

3 Results

3.1 Static Measurements of Nursing Staff

Amongst the emergency surgical nursing staff, the average age was 31, average height was 159.1cm, the average elbow height was 99.1cm, the average hand length was 16.85cm, the average shoulder width was 35.84cm, the average palm width was 7.31cm, the average palm length was 9.66cm, and the average year of experience was 8.2 years. Amongst the non-surgical nursing staff, the average age was 26.7, average height was 161cm, the average elbow height was 100.48cm, the average hand length was 17.03cm, the average shoulder width was 35.95cm, the average palm width was 7.04cm, the average palm length was 9.6cm, and the average year of experience was 5.1 years.

3.2 Regression Analysis of Elbow and Wrist Joint Angle of Emergency and Non-emergency Surgical Nursing Staff during Operation from the Central Position

The R value represents the relationship between independent and dependent variables, and the descriptive relationship range between the two is as follow: 0~0.2 represents a low relationship, 0.2~0.5 represents a minimal relationship, 0.5~0.7 represents a sure relationship, while 0.7~1.0 represents a deep relationship. Hence, the R^2 value of 0.414 shown in table 1 is within the range of a sure relationship. From Figure 5(a), we can see that the elbow height of between 102~108cm is closer to the range of the second regression linearity range, meaning that the wrist joint angle value (flexion) could be obtained for nursing staff with an elbow height within this range. Amongst the non-emergency surgical nursing staff, the value R^2 shown in the figure is 0.516, which falls in the deep relationship range, representing a deep relationship between elbow heights and wrist joint angles. The ideal regression equation for non-emergency surgical nursing members shows that elbow heights of between 102cm and 104 cm obtains a bigger wrist joint angle (about 155°), while in reality, non-emergency surgical nurses presented bigger wrist joint angles when their elbow height is between 98 and 104 cm. However, from Figure 3(a), we can see that participants with elbow height of 98cm are outside of the confidence level.

$$Y = F_1X^2 + F_2X + F_3 \tag{4}$$

Amongst which, Y : wrist joint angle ; R^2 : differentiation coefficient ; X : elbow height (Table 1) / height(Table 2) ; F_1, F_2, F_3 : regression coefficient.

Table 1. Emergency and non emergency nursing staff elbow height and wrist joint angle regression equation statistic summary chart

Regression equation	R^2	df	F
$Y = -0.2702X^2+56.7252X-2818.6$	0.414	8	2.47
$Y = -0.256X^2+52.544X-2545.4$	0.516	7	3.72

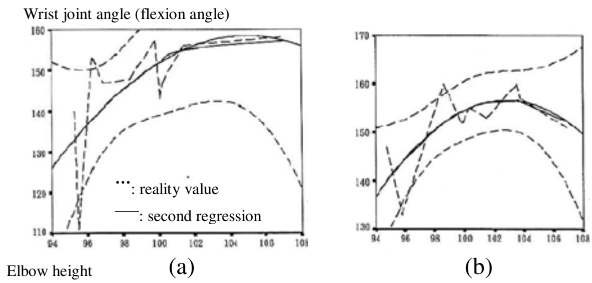


Fig. 3. Elbow height and wrist joint angles during operation from the central position. Second regression linearity chart (a) emergency surgical nursing staff (b) non emergency surgical nursing staff

Table 2. Emergency and non emergency nursing staff height and wrist joint angle regression equation statistic summary chart

Regression equation	R^2	df	F
$Y=-0.405X^2+133.023X-10753$	0.611	7	5.50
$Y=-0.1351X^2+ 45.1697X-3616.8$	0.477	7	3.19

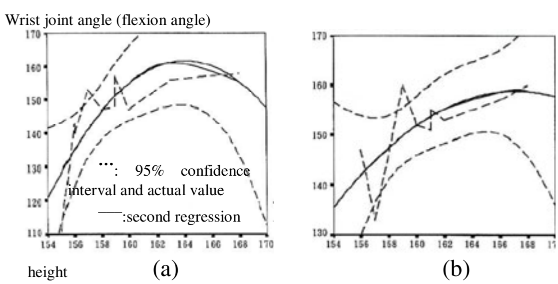


Fig. 4. Height and wrist joint angles during operation from the central position. Second regression linearity chart (a) emergency surgical nursing staff (b) non emergency surgical nursing staff.

3.3 Regression Analysis of Height and Wrist Joint Angles for Emergency and Non-emergency Surgical Nursing Staff during Operation from the Central Direction

The differentiation coefficient R^2 shown in Table 2 is 0.611, which falls in the deep relationship range. This means significant relationship is present between height and wrist joint angles. From Figure 4 (a) we can see that ideal second regression presents the largest wrist angle at participant height 164cm. Participants with height 164cm and above, on the contrary, shows a smaller wrist angle than 160°.

3.4 Variance Analysis of Actual and Virtual Operation on Controllers

From Table 5 we can see that the correlation between actual and virtual operations in three different positions was significant ($F=13.840$, $P<0.05$). The results were also significant amongst participants ($F=10.408$, $P<0.05$). In the aspect of the correlation between actual and virtual operations with emergency and non-emergency surgical nursing members, the results were significant ($F=19.775$, $P<0.05$), however the chief effect amongst participants were insignificant.

Table 3. Comparison chart of virtual and actual operation relation analysis data (emergency / non-emergency surgical nursing members)

	Group number	Relational coefficient	Probability
Emergency surgical nursing members			
Relational analysis of virtual and actual left positio n average	10	0.775	0.009*
Relational analysis of virtual and actual central pos ition average	10	0.484	0.157
Relational analysis of virtual and actual right position average	10	-0.294	0.410
Non-emergency surgical nursing members			
Relational analysis of virtual and actual left positio n average	10	-0.092	0.801
Relational analysis of virtual and actual central pos ition average	10	0.134	0.712
Relational analysis of virtual and actual right position average	10	0.069	0.850

3.5 Nursing Staff Operational Panel Height Analysis

In the data of the emergency surgical nursing members, the actual operation height from the left position ranged from 92.97 to 95.64cm; height from the central position ranged from 92.1 to 98.72cm; height from the right position ranged from 91.62 to 99.14cm. The virtual operation height from the left position ranged from 94.19 to 96.92cm; height from the central position ranged from 93.84 to 97.78cm; height from the right position ranged from 89.73 to 94.6cm. While in the data of non-emergency surgical nursing members, the actual operation height from the left position ranged from 91.79 to 96.64cm; height from the central position ranged from 93.01 to 98.27cm; height from the right position ranged from 91.26 to 98.59cm. The virtual operation height from the left position ranged from 89.54 to 100.73cm; height from the central position ranged from 89.21 to 94.89cm; height from the right position ranged from 92.97 to 95.64cm. From the above data, we can summarize the data and information into the following: amongst all the participants, the virtual button height ranged from 89.21 to 100.73cm. This means, in both emergency surgical and non-emergency surgical nursing staff groups, the objective control height range is larger than the present instrument height range of 92 to 95cm.

Table 4. Pair test comparison chart of actual and virtual operation controls (emergency / non-emergency surgical nursing staff)

	Pair test			t value	freedom	probability
	Average difference	Standard error	Average standard error			
Emergency surgical nursing members						
Virtual and actual left position average pair test	-2.68	11.935	3.774	-0.710	9	0.496
Virtual and actual central position average pair test	6.06	12.918	4.085	1.483	9	0.172
Virtual and actual right position average pair test	-25.24	17.947	5.675	-4.447	9	0.002*
Non-emergency surgical nursing members						
Virtual and actual left position average pair test	23.07	22.388	7.080	3.259	9	0.005*
Virtual and actual central position average pair test	21.86	12.328	3.899	5.607	9	0.479
Virtual and actual right position average pair test	-4.29	18.359	5.806	-0.739	9	0.000*

Table 5. Angle variance analysis chart for actual and virtual operations

			ANOVAs[a,b]				
			Sum of Squares	df	Mean Square	F	Sig.
Angles	Main Effects	**Between participants**	**1573.386**	**4**	393.347	2.338	0.060
		Three-position recognition	595.395	2	297.698	1.769	0.175
		Actual and virtual control recognition	540.601	1	540.601	3.213	0.076
		Emergency surgical recognition	437.390	1	437.390	2.599	0.110
	2-Way Interactions	**Amongst participants**	**8756.464**	**5**	1751.293	10.408	0.000*
		Three-position recognition × actual and virtual control recognition	4557.483	2	2328.742	13.840	0.000*
		Three-position × emergency surgical recognition	771.500	2	385.750	2.292	0.105
		Actual and virtual control recognition ×emergency surgical recognition	3327.480	1	3327.480	19.775	0.000*
	Model		10329.9	9	1147.761	6.821	0.000*
	Residual		18509.4	110	168.267	-	-
	Total		28839.2	119	242.346	-	-

a. Angles by three-position recognition, actual and virtual control recognition, emergency surgical recognition
b. *:P<0.05

4 Discussion

According to the results of this research, we can discuss the following: (1) When emergency surgical nurses operate buttons from the central direction, the range of wrist joint angles (dorsal flexion) range from 111° for participants that are 155cm tall and 158° for participants that are 168cm tall. As for non-emergency surgical nurses operation buttons from the central direction, the wrist joint angle ranges from 133° for 157cm tall participants and 160° for 168cm participants. We can see from the results that participants with the same heights do not necessary obtain the same wrist angles; while the smallest 111° wrist angle sample does belong to the 95% confidence interval. The factor that caused this result could be due to the behavior differences produced by the operating habit of participants. (2) In the aspect of emergency surgical nurses operating actual and virtual buttons, the only movement formations that presented significant differences were the right-hand direction of virtual and actual control; no significant differences were found for the control from the central direction or the left-hand direction. This result shows that emergency surgical nurses adapt better to operating buttons from the central and the left directions. In the aspect of right-hand direction control, the virtual right-hand side difference has a great effect on operation angles. This tells us that buttons operated from the central direction is more acceptable for emergency surgical nurses. (3) In the aspect of non-emergency surgical nurses operating actual and virtual buttons, the most significant differences occurred in the movement formation of virtual and actual controls from the left and the center; while no significant differences were presented for the virtual and actual control from the right-hand direction. This shows that non-emergency surgical nurses adapt better to controls from the right-hand direction.

5 Conclusion

Due to the insufficiency of sample numbers, differences between the above experiment results and the hypothesis could not be avoided. During general dynamic experiments, the better setting for video data deriving is six or more cameras, meaning the larger the camera numbers the easier it is to obtain reactions from the reflection balls, hence the easier it is to be derived. As for the calculation of angles and the judgment of linear distance between two fingertips, the concept of space coordinates was employed. The formula for space vectors were transferred for the program, future studies should try to induce the concept of database into the program in order to obtain the complete data result for hand angle calculation and the relative value. The experiment process of this research is to stick reflection balls on the gloves; reflection balls may shift slightly during the operation process, resulting in error in the coordination derivation by the movement analysis system. Hence, suggestions are made to future researches of hand movement analysis, to consider sticking reflection balls directly on hands to avoid errors.

Acknowledgements. Our deepest appreciations to the emergency room staff of the Tainan Hospital and National Cheng Kung University Hospital for the staff arrangements for experiment participation.

References

1. Armstrong, T., Foulke, J., Joseph, B., Goldstein, S.: Investigation of cumulative trauma disorders in a poultry processing plant. AIHAJ 43(2), 103–116 (1982)
2. Barnes, R.M.: Motion and Time Study: Design and Measurement of Work, vol. 235, pp. 116–128, 366-389. John Wiley, New York (1980)
3. Barnes, R.M.: Motion and time study, 5th edn. Wiley, New York (1963)
4. Bobjer, O.: Screwdriver handle design for power and precision. In: Proceedings of the 1984, International Conference on Occupational Ergonomics, Human Factors Association of Canada, pp. 443–446 (1984)
5. Bradley, J.V.: Tactual coding of cylindrical knobs. Human Factors 9(5) (1967)
6. Chiu, H.Y., Su, F.C.: The motion analysis system and the maximal area of fingertip motion. Journal of Hand Surgery 21(B), 604–608 (1996)
7. Fleiss, J.L.: The design and analysis of clinical experiment. John Yiley & Sons, New York (1986)
8. Young, F.W.: Multidimensional Scaling-history, theory, and applications. L. Erlbaum Associates, Hilisdale (1987)
9. Green, P.E., Carmone Jr., F.J., Smith, S.M.: Multidimensional Scaling, concepts and applications. Allyn and Bacon, Massachusetts (1989)
10. Hunt, D.P.: The coding of aircraft controls, Tech. Rept, U. S. Air Force, Wright Air Development Center, pp. 53–221 (1953)
11. Jobson, J.D.: Applied Multivariate Data Analysis. Springer, Heidelberg (1992)
12. Hartigan, J.A.: Clustering algorithms. Wiley, New York (1975)
13. Kantowitz, B.H., Sorkin, R.D.: Human Factors, Understanding people-system relationship. John Wiley & Sons, New York (1983)
14. Kroemer, K.H.E.: The state of ergonomic knowledge about cumulative trauma disorders. In: Marras, Karwowski, Smith, Pacholski (eds.) The Ergonomics of Manual Work, pp. 19–24. Taylor and Francis, London (1993)
15. McCormick, E.J., Sanders, M.S.: Human Factors in Engineering and Design, p. 181. McGraw-Hill, New York (1982)
16. Sperling, L., et al.: A cube model for the classification of work with hand tools and the formulation of functional requirements. Applied Ergonomics 24(3), 212–220 (1993)
17. Shi, Y.-J.: Virtual Reality Technology. Computer and Communication 28, 3–15 (1994)
18. Hsu, S.-H., Shieh, K.-K., Yang, W.-D.: Classification of Hand Movements. Journa1 of the Chinese Institute of Industrial Engineers 13(2), 145–155 (1996)
19. Li, C.-H.: Job characteristics of emergency department and process design of the basic equipment and staffing, department of design planning hospitals and clinics, pp. 114–119 (1988)
20. Lin, G.-M.: Emergency Nursing. Wey Far Books Co., Ltd, Taipei City (1993)

Upper Limb Contralateral Physiological Characteristic Evaluation for Robot-Assisted Post Stroke Hemiplegic Rehabilitation

Lap-Nam Wong, Qun Xie, and Linhong Ji

Division of Intelligent and Biomechanical System, State Key Laboratory of Tribology, Tsinghua University, 100084 Beijing, PRC
Lap-Nam Wong-wong1055@hotmail.com,
Qun Xie-quentinx2002@gmail.com,
Linhong Ji-jilh@tsinghua.edu.cn

Abstract. An innovative robot-aided quantitative evaluation method was established in order to evaluate disability level of post hemiplegic stroke patients. A sEMG-driven musculoskeletal model utilized the physiological characters of impaired limb and normalized by healthy limbs' physiological characters can be applied to calculate muscle strength and other dynamics data (such as elbow joint torque). By comparing physiological characters of impaired upper limb to contralateral healthy limb of same subject, researchers are able to investigate the difference of motor function between impaired and healthy. Comparing the comparability of both sides' with patients and healthy person may assist researchers and doctors to create more appropriate training or treatment plan to clinical needs.

Keywords: Rehabilitation robot, quantitative evaluation, stroke, hemiplegic.

1 Introduction

According to World Health Organization (WHO), stroke is considered as the most common cause of disability and the second-most common cause of death worldwide.[1-2] Post-stroke survivors will have various levels of neurological dysfunction in motor, sensory, speech, cognitive, emotional and other aspects. Hemiplegia, showing half-length or unilateral paralysis, is the common form of motor dysfunction after stroke. [3] Because of causing inconvenience to patients and their families, the recovery of motor function is regard as an essential section of rehabilitation. Over the past two decades, many rehabilitation robots for limb motor recovery were developed, yet most of these robots are still in the experimental stage. Due to the absence of effective evaluation method, it will still take some times to achieve widespread clinical application, even though some robots have been commercialized.

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 458–463, 2011.

The contemporary development trend of rehabilitation engineering no longer pays attentions to mechanism design or system upgrade of rehabilitation robot. Clinical application of such robot became a focal point of the area. Quantitative evaluation is a prerequisite for clinical use, not only presents credible evidences of clinical efficiency, but also provides necessary parameters for intelligent robot system. In order to provide effective rehabilitation for different individual, researchers should establish targeted training task based on patient's physiological character. Yet previous study show that current assessment method, such as Fugl-Meyer (FM) assessment, may not be effective to robot-assisted rehabilitation.

Post stroke motor recovery is a movement re-learning process based on brain plasticity. By stimulate particular muscles, the residual voluntary control capabilities of patients over those muscles will be aroused and strengthened, thus its function can be restored. Some commercial rehabilitation robots are able to record movement pattern of the healthy limbs of patient, and lead the impaired limbs of patient to imitate the movement pattern. However, a superficial imitation may not satisfy the clinical needs. In our previous work, researchers confirmed that tilting operation plane of shoulder and elbow coordination training task by different angle will influence surface electromyography (sEMG) values of upper limb muscles. Changing to a particular operation plane could strengthen certain muscle groups. [4] To motivate impaired muscles, the impaired limbs require a targeted training task based on patient's physiological character which may be completely different from the movement pattern of the healthy limbs.

In the current study, we establish an innovative robot-aided quantitative evaluation method, which is able to investigate the difference of motor function between impaired limb and healthy limb during upper limb hemiplegic rehabilitation. In this method, the contrast between impaired limbs and contralateral healthy limb physiological characters, such as sEMG, kinematics data or muscle strength, will be showed. Comparing the comparability of both sides' physiological characters with patients and healthy person may become basis for researcher to evaluate disability level of patients and to create training plan.

2 Contralateral Physiological Characteristic Evaluation

We proposed a quantitative evaluation method which is based on contralateral healthy limb physiological characteristic called "Upper Limb Contralateral Physiological Character Evaluation". This evaluation method is based on symmetry analysis of human body. A research in healthy people, conducted by Van der Loos et al, indicate that muscle contraction and sEMG activity on both sides of upper limbs have similarity and consistency, when both sides receive similar stimulation or training task.[5-6] We also confirmed this phenomena in our study[7]. Obviously, there will be fewer similarity and consistency between impaired and healthy extremity on hemiplegic patients. Therefore, setting contralateral healthy limbs' physiological characteristics as basis to evaluate impaired limb function may be available.

By the method, we use sEMG of impaired limb to establish a sEMG-driven musculoskeletal model. Importing sEMG and other physiological parameters of healthy limb into the musculoskeletal model, a strength model of impaired muscle can be estimated. We utilize Motion Analysis Corporation's SIMM software to construct sEMG-driven musculoskeletal model. Because of using a theoretical model from the Hill's muscle model, only sEMG and kinematics data need measured by actual movement. In this study, kinematics data were captured by Motion Analysis Corporation's Motion-Capture System. The training task possesses similar trajectory and equivalent velocity when evaluating both sides of one patient. In our study, patients holding different FM scores achieved different motor performance when making a voluntary movement in trajectory tracking tasks (Fig. 1.). According to Terry K.K. Koo et al [8], the calibration of a model by certain movement task can only be suitable for quantitative prediction of joint torque and muscle force in similar movement task. Hence, we employ rehabilitation robot to provide an assistant force during trajectory tracking tasks so as to ensure consistency of each task.

When estimate sEMG of the impaired limb obtained by a certain training task, MVC-sEMG of contralateral healthy limb is introduced as the normalized basis. Using a simple trajectory (such as reciprocating linear trajectory) during data collection so as to avoid confounding factors interferes with the musculoskeletal model. In order to prevent compensatory movement, the limb of subjects should be fixed on rehabilitation robot and their body should be fixed with seat belt (Fig. 2.).

The participations of each muscle in same joint are different, when patients do different trajectories or movement task. We assign appropriate weights to each muscle base on the participations in the calculation so as to make the results more realistic value. In order to verify validity this musculoskeletal model, we will compare the results computed by the model with the results calculated by inverse dynamics.

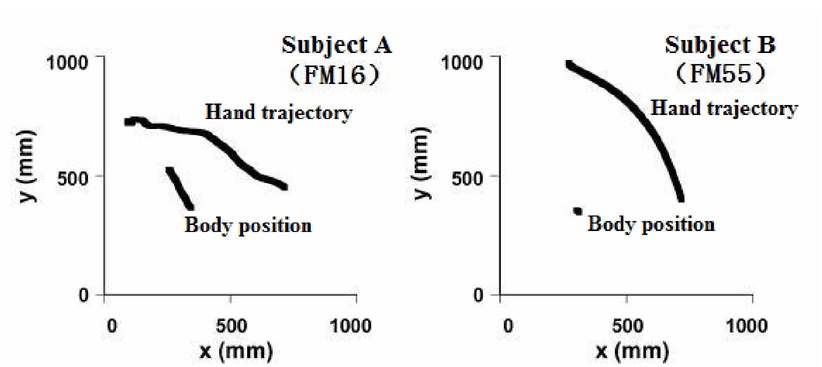


Fig. 1. Patients with different Fugl-Meyer scores achieved different motor performance when tracking counterclockwise circular trajectory. This figure shows the trajectory in the fourth quadrant.

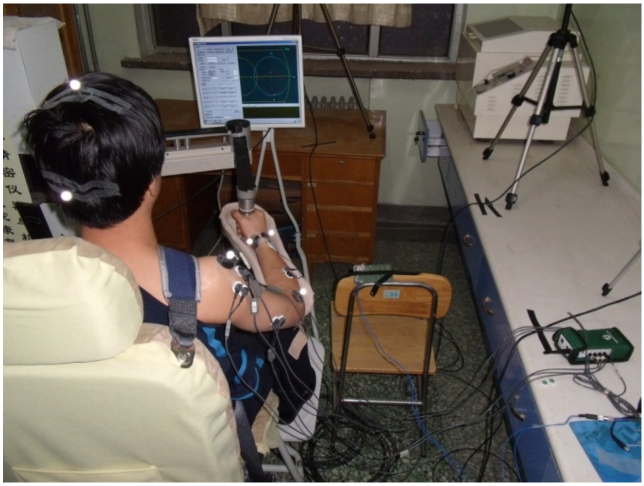


Fig. 2. Robot-aided movement data capture workstation

3 Relationship between Conventional Assessment and Movement Trajectories

According to a preliminary experiment we conducted, there is comparability between both sides of patients and limbs of healthy person. Subjects were asked to make voluntary protrusive movement by their upper limbs and kinematics data was measured. We observed that trajectories of protrusive movement made by the patients' impaired limb during are different from trajectories of protrusive movement made by patients' healthy limb and healthy persons' limbs. Both of these two differences are comparable and the comparability also applies to FM assessment scores. Accordingly, cognizing the relationship between clinical assessment and the training track is a prerequisite of quantitative evaluation.

The sEMG-driven musculoskeletal modeling based on the evaluation method is able to analyze the relationship between conventional assessment method and trajectories which make by patients. In previous work, we confirmed that the sEMG contribution values of muscles are different during diverse movement period of trajectory tracking tasks. In the current study, researchers confirmed the correlation between both sides' sEMG contribution values of muscles in healthy persons' limb. Using one side of limb making circular trajectory in voluntary trajectory tracking tasks and making a reversed circular trajectory tracking tasks implement by the contralateral limb will provide similar characteristic of muscles' sEMG contribution values. When subjects making clockwise circular trajectory follow 3rd-2nd-1st-4th quadrantal order by right hand, the similar sEMG contribution values of left hand are provided by making counterclockwise circular trajectory follow 4th-1st-2rd-3nd quadrantal order. Motion sequences of circular trajectory tracking task possess similar characteristic of muscles' sEMG contribution values is shown in Table 1.

Table 1. Motion sequences of circular trajectory tracking task possess similar characteristic of muscles' sEMG contribution values. The sequences are expressed in quadrantal order. These sequences apply to healthy people.

Upper limb of subject	Clockwise	Counterclockwise
Right hand	3rd-2nd-1st-4th	4th-1st-2nd-3rd
Left hand	4th-1st-2nd-3rd	3rd-2nd-1st-4th

By comparing performance of muscle activity in every movement quadrant of selected tasks, researchers may able to understand muscle coordination of subjects' limbs. Using this method to analysis coordination of patients, we can recognize the period of movement possess abnormal muscle activity. According to the performance of patients in each motion period, we can determine what kind of trajectory may cause obstruction to patients and what kind of trajectory may cause more motivation to specific muscle. This information can be used to estimate with muscle strength and joint torque which is calculated by the musculoskeletal model so as to acquire a quantitative evaluation of muscles activation during varied training period and know about status of residual motor function of patients. Comparing the trajectory with conventional assessment result, we can estimate the relationship between them. Thus, we are able to select the most appropriate trajectory for patient who retains certain disability level.

4 Discussion and Conclusion

Combining the method with conventional assessment method, we can get a further understanding of relationship between upper extremity tracking data, muscle strength-trajectory evaluation and conventional clinical assessment. The data can be used as technical reference for doctor or physiatrist to make a prescription training task for rehabilitation robot. It will be convenient for us to design a targeted training task closing to the clinical needs.

Using these quantitative data, automatic rehabilitation robot is able to provide corresponding training task to different patients. Because of modeling based on one's personal physiological characters, the quantitative relationship is unique. Thus, the adaptability of robot to different individual can be enhanced. On the other hand, researchers of rehabilitation engineering are annoyed by the complexity of post stroke hemiplegia for a long time. It is difficult to rule out the effect of individual differences and recruit a larger amount of similar patients in clinical experiments. This self-comparison method is competent for avoiding the effect of individual differences and reduces required sample size.

In summary, we developed a novel quantitative evaluation method for robot-assisted rehabilitation which is able to assist researchers and doctor to make targeted training program based on personal clinical needs of post stroke patient. The evaluation can provide parametric data for robot-assisted trajectory tracking training task. It will improve the adaptability of robot to different individual. We using rehabilitation robot, sEMG, motion capture system, force sensor in this method. The future work will be involved in application of feedback control base on the quantitative evaluation.

Acknowledgments. This study was a cooperation program with PLA General Hospital of China, which participated in the development of the rehabilitation robot. Patients and experimental site was provided by the hospital.

References

1. Strong, K., Mathers, C., Bonita, R.: Preventing stroke: saving lives around the world. Lancet Neurology 6, 182–187 (2007)
2. Donnan, G.A., Fisher, M., Macleod, M., Davis, S.M.: Stroke. Lancet 371, 1612–1623 (2008)
3. Johnstone, M.: Home Care for the Stroke Patient: Living in a pattern, 3rd edn. Churchill Livingstone, New York (1996)
4. Wang, Z.X., Chen, L.N., Yao, C.Y., Xie, Q., Ji, L.H.: The influence of transforming operation plane on compound movement training in upper limb robot aided rehabilitation. Chinese Journal of Rehabilitation Medicine 24, 65–67 (2009) (in Chinese)
5. Johnson, M.J., Van der Loos, H.F.M., Burgar, C.G., Shor, P., Leifer, L.J.: Principle for designing in motivation into a robotic stroke therapy device. In: Second Joint EMBS/BMES Conference, pp. 2439–2440. IEEE Press, Houston (2002)
6. Lum, P.S., Burgar, C.G., Kenney, D.E., Van der Loos, H.F.M.: Quantification of force abnormalities during passive and active-assisted upper-limb reaching movements in post-stroke hemiparesis. IEEE Trans. Biomed. Eng. 46, 652–662 (1999)
7. Zhang, Y.B., Wang, Z.X., Ji, L.H., Bi, S.: The Clinical Application of the Upper Extremity Compound Movements Rehabilitation Training Robot. In: The 2005 IEEE 9th International Conference on Rehabilitation Robotics, pp. 91–94. IEEE Press, Chicago (2005)
8. Koo, T.K.K., Mak, A.F.T.: Feasibility of using EMG driven neuromusculoskeletal model for prediction of dynamic movement of the elbow. Journal of Electromyography and Kinesiology 15, 12–26 (2005)

A Case Study of the Design and Evaluation of a Persuasive Healthy Lifestyle Assistance Technology: Challenges and Design Guidelines

Jie Xu, Ping-yu Chen, Scott Uglow, Alison Scott, and Enid Montague

University of Wisconsin-Madison, Mechanical Engineering Building
1513 University Avenue, Madison, Wisconsin, USA
{jxu29,pchen34,suglow,ascott2,emontague}@wisc.edu

Abstract. Technology can be used as an intervention of unhealthy lifestyles, but designing such a technology is challenging – usability as well as the ability of changing the user's behavior needs to be considered. The design and evaluation process of a future generation persuasive healthy lifestyle assistance technology which involves physiology, environment monitoring, and automation was studied in this paper.Several challenges were identified and design guidelines were developed for designing such a technology which is used as an intervention of the user's unhealthy lifestyle.

Keywords: persuasive technology, case study, lifestyle.

1 Introduction

1.1 Persuasive Technology and Lifestyle

Since an unhealthy lifestyle (such as an unhealthy diet and physical inactivity) might lead to many diseases and health risks and is becoming a threat to the general population [1], intervention is needed. Well-designed technology can be a good intervention, because appropriate user interface and information design have shown to be able to cause behavior changes [2]. Literature review also indicated that healthy living persuasive technologies showed that it can change a user's attitude and behavior [3].

According to Chatterjee and Price's [4] categorization of persuasive technologies in healthy living, current generation technologies have the following characteristics: wearable sensors to collect information from the user; awareness the user's current status; persuasive technique and real time information exchange within the system. The future generation technologies should introduce an automation component in the system to make human intervention minimal.

In this case study, the design team developed a future generation persuasive healthy lifestyle assistance technology called HealthyEdge. This technology is able to monitor a user's physiological status (through physiology monitoring accessories worn by the user), the environment, and devices which are closely related to lifestyle

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 464–471, 2011.

(e.g. physical exercise facilities and television), and helps the user plan his/her healthy lifestyle by offering health related information as well as a set of automatic planning tools.

1.2 Apersuasive Healthy Lifestyle Assistance Technology

The aim of the project "HealthyEdge" was to develop a persuasive healthy lifestyle assistance technology that is able to support important aspects of user's healthy lifestyle, including diet and physical exercise, while being embedded into the user's daily life.

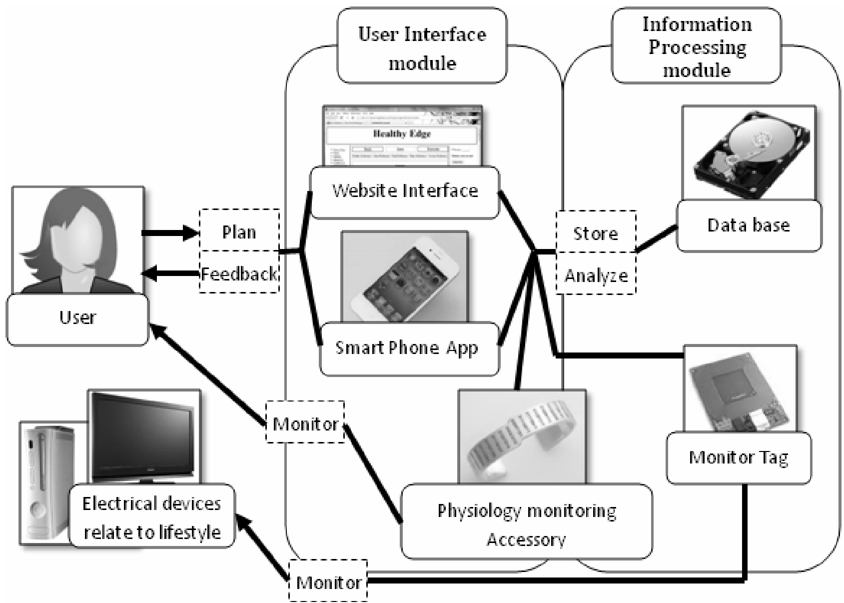


Fig. 1. The structure of the persuasive healthy lifestyle assistance system – HealthyEdge

The final outcome of this project was the HealthyEdge system. The system includes two main modules (see Fig. 1 for visualization of the structure of the system): the information processing module and the user interface module. The information processing module included a data base of user's health information, monitor tags and a physiology monitoring accessory. The monitor tags are used for monitoring how long the user has been watching TV, playing video games, or being inactive with other electrical devices. The tags can be attached to electrical devices. The information gathered by the tag will be transferred to the data base for further analysis. The physiology monitoring accessory is a wearable device which is used for gathering physiological information about the user. It gathers information related to health, such as heart rate and blood pressure. The information stored in it will further be transferred to the data base.

The system's website, which connects the user to the data base, has two main functions: to plan the diet and exercise schedule and to offer feedback from the information provided by the physiology monitoring accessory and the monitor tags. The fully functional website is able to create a healthy meal, generate a grocery list, make a physical exercise schedule, and to keep track and update health goals (e.g. weight control, body building) for the user. The website can automate lifestyle planning for the users; however, users who are not satisfied with the automatic planning can customize the plan under the instruction of the system as well. Also, various healthy lifestyles related information will be provided. The smart phone application will have similar functionality as the website.

As a case study of the design and evaluation process of HealthyEdge, the objective of this paper is to identify the challenges and develop guidelines for designing a persuasive healthy lifestyle assistance technology which is used as an intervention of the user's unhealthy lifestyle.

2 The Design Process

This case study followed six user-centered design and evaluation stages towards designing the technology (shown in Fig. 2). In the first stage - Contextual Inquiry [5], each member of the design team observed one participant interacting with technologies to explore opportunities for possible design. During this stage, each member in the design group did their own observation with the participant. The tasks for the participants were stemmed from the group members' own product concepts. Different participants' performing various tasks on the internet or some other interfaces was observed. This was very important because the user may interact with the interface of the system at anytime, so that good interfaces that could be fitted into different contexts are critical.

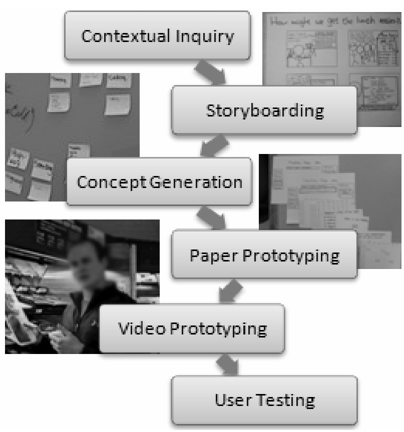


Fig. 2. The design process of the technology

In the second stage – storyboarding [6], each member created and evaluated cartoon scenarios which displayed users performing different tasks. Then the storyboards were shown within the group to get feedback of the ideas. Participants were also recruited to make comments on the storyboards. Through storyboarding, a more adaptable healthy lifestyle assistance system was created conceptually.

In the third stage - concept generation, the design team came to a consensus of the final system by brainstorming and combining ideas and concepts from previous design stages. The design team finally landed on apersuasive healthy lifestyle assistance system.

In the fourth stage - paper prototyping [7], the design team created a rough prototype and conducted an user testing with it to facilitate the interface design. According to Snyder [7], paper prototyping is good for testing the terminology, navigation, content, page layout and functionality; and is helpful for finding problems in the early design stage of the interface to allow for rapid changes. A low fidelity model of HealthyEdge website using paper and markers was developed by the design team. User testing of the prototype provided front-end results based on usability and difficulty of use of the system. The participant engaged in the think aloud method while interacting with the system, was observed, and participated in an interview. Think aloud method eliminated the time consuming activity of designer-user communication and proved to be effective [8]. According to the feedback from the user, an efficient website was developed with only a few errors and confusing points in the tasks assigned during the user testing process. This was very important for a healthy lifestyle assistance system of which the stakeholders will have various ages, educational and income backgrounds to have a universally designed website interface.

In the fifth stage - video prototyping, a video prototype was created to simulate how the system will be used in the real-world. Video prototyping can help the designer include contextual elements in the design [9]. The video created by the design team simulated what life would be like using a finished HealthyEdge system. The video mainly focused on the interaction between the user, environment, physiology monitoring accessory and monitor tags. In the process of making the prototype, the design team got a better idea of how the system is able to monitor the activities of the user without disrupting his/her daily life and how to help the user build up a healthier way of living through common devices such as a computer and a cell phone.

In the sixth stage - user testing, an interactive prototype was developed for a multi-stage user testing. The user testing consisted of two components: (1) using of the website, (2) simulating a day using HealthyEdge combined with daily life. In addition to observations, interviews were administered to collect feedback on learnability, efficiency, memorability, errors, satisfactions, and further suggestions for the design.

3 Challenges and Guidelines

3.1 User's Acceptance and Actual Use of Technology

In the early stages of the design process, the biggest challenge was designing the user interface and integrating technologies to increase user's acceptance and technology usage, which was one of the overall goals of the system as well. One important aspect of this problem is that the technology might not be well integrated into the user's normal daily life so additional work (setting up the program, looking for internet access while updating the information is needed, etc.) will be needed for the user to interact with the technology. Under this situation, the user won't work with the technology in the way as the designer expected [10]. The design team also tested how the system could be embedded in the users' daily life through video prototyping. A challenge was addressed: how to change the user's behavior without interfering with the user's normal life. One possible solution could be designing a smart interface to facilitate the communication between the system and the user. The guidelines for designing a compatible system are also applicable here.

Guideline 1: Compatible with other technologies. In HealthyEdge, the communication between the user and the system needs various channels, for example, email or cell phone text messages .For example, if the system is sending message through cell phone texts, in order to avoid disturbing the user, the message should be sent when the user is available. Therefore the system will work with a calendar program to check the availability of the user. Another issue is that the smart phone application should have different version that works with different operating systems. The monitor tags also need to be compatible with the electrical devices. In the design of a system that is to be embedded into user's daily life, it is necessary that the system beingcompatible with the devices or programs which are currently used by the user (for example, television or online calendar) so that the user does not need to exert extra efforts to start using the new system.

Guideline 2: Create a "player experience" for the user. The system should have a high level of usability as well as player experience [11] to make sure the user can have fun while interacting with the system, so that technology usage will be increased.A literature review showed that video games are used for changing health related behaviors [12]. "People fun" refers to the emotions created by enjoyment of social interaction [11], and this kind of play experience maybe used by healthy lifestyle persuasive technologies; for example, social support from an online community could have an impact on health related attitude and behavior change [4]. Some kind of multi-user games (for example, games involved in competing or collaborating with other users in losing weight) can be designed for using in the online community to motivate the user. Other than games, "player experience" can also be created in product designs [11]. A goal setting module of health status (for instance, set a weight goal and visualize how the user's weight changed over time and finally reached the goal) could offer the user "serious fun", which is related to the enjoyment of real-world benefits from interacting with the system [11].

3.2 Adapting the System to the User

Through paper prototyping, the design team mainly investigated an efficient way for the user to understand and control a system. In the video prototyping, how to make the system adaptable to a user's need and daily life was considered. The main challenge that the design team encountered was how to make the interface adaptable to a user who is not aware of healthier lifestyles and a user who knows more and would like a customized health planning.

Guideline 3: Design an understandable interface. A survey based research showed that about one third of the respondents did not understand calorie labeling and about half of the respondents would not use the calorie information in restaurants [13]. This implied that information should not be merely presented in the way that the user is "supposed to understand". In the design process, the design team made efforts to group the concepts related to lifestyle in an understandable manner to help the user learn, navigate and explore the system more efficiently. Health information should be presented in an easy way that is understandable to most people while more detailed information is also made available for advanced users. For example, for novice users, nutrition information could be presented using food categories like "grain", "oil", "vegetable", etc.; for more advanced users, the same information could be presented as "calorie", "protein", "fiber", etc.

Guideline 4: Support special needs of the individual user. The needs of individual users may vary. For example, normal daily diet may not work well for the users with diabetes. So the individual information concerning the user's health condition should be gathered by the system. This will be critical if the system is automated. In HealthyEdge, there was an option for automatically doing diet, physical activity monitoring and planning. The information offered by the automated system should not conflict with physician's advice or mislead the user. In most cases, this information could be provided by the user; but in some cases, the system will need to work with other information sources to acquire more detailed and accurate information. One possible way to accomplish this is to work with personal health records [4]. However, more research should be done on this because of the involvement of political and ethical issues. Another option may be to work with the caregiver of the individual user, such as nurses, to acquire information needed.

Guideline 5: Automation made adaptable to "novice users" and "expert users". In the design of HealthyEdge, a prototype with an automatic meal and exercise planning feature was developed to help the novice user get started; meanwhile, different levels of automated life planning and information offering were also provided to the users who have a better idea of health related concepts and knowledge. Offering more information to advanced users can lead to better understanding of the underlying mechanism of the automation and may increase their trust towards the system. This was also served for the purpose of educating the novice users. As the usage by users increased, the system would offer more detailed information for the user (for example, more detailed nutrition intake recommendation for a day). In this way a novice user will eventually become an advanced user.

3.3 Usefulness and User Satisfaction

In the user testing stage, the design team explored factors attributed to perceived usefulness and user satisfaction of the system. Main design challenges included the design of the functionality of the system, physiology monitoring accessory, and the information visualization method.

Guideline 6: Capture the full scope of user's lifestyle. HCI design guidelines were well developed for how to maintain ease of use with rich functionalities [14]. In the design of a healthy lifestyle assistance system, enough functions should be included into the system to capture the full scope of the user's lifestyle. There are four health lifestyle characteristics (HLCs) frequently cited in the literature – nonsmoking, healthy weight, fruit and vegetable consumption, and leisure time physical activity [15]. The basic functions of the system should include smoking cessation, weight control, health diet, and physical activity assistance. Previous research showed that persuasive technologies could be effective in helping the user with those HLCs (for a review please see [4]). A system that could offer assistance in all four aspects would be helpful.

Guideline 7: Well designed and customization of wearable physiology monitoring accessories. The physiology monitoring accessories should be in various shapes and customizable, since it is the device that the user will be wearing. However, there are still other issues. In the user testing stage of this experiment where a user's one day life with HealthyEdge was simulated, the participant chose a metal bracelet when multiple choices regarding material and types of accessories were offered. In the interview after the simulation, when being asked if the bracelet was comfortable to wear, our participant answered "*No. It was too heavy and I always noticed it. I took it off to sleep even though I knew I was supposed to keep it on at all times. I also took it off to shower.*" Therefore, it is necessary to find a material that is comfortable to wear at all times, including sleeping time. More importantly, a return/exchange policy should be adopted for physiology monitoring accessory, since even when offered the choice for different accessories, the user may choose the "wrong one" that they then find it uncomfortable to wear.

4 Conclusion

A future generation persuasive technology like HealthyEdge will soon become possible given the recent advances in internet, mobile computing, and physiology monitoring technologies. This paper studied the design and evaluation process of such a system, discussed some of the challenges, and presented seven design guidelines for overcoming the obstacles. However, there are still many other challenges other than the design of the technology itself. For example, the ethic and policy related problems with monitoring the user's behavior and using the user's health records. These problems should be studied by future research.

References

1. World Health Organization, `http://www.who.int/dietphysicalactivity/publications/facts/cvd/en`
2. Marcus, A., Jean, J.: Going Green at Home: The Green Machine. Information Design Journal 17, 235–245 (2009)

3. Kroeze, W., Werkman, A., Brug, J.: A systematic review of randomized trials on the effectiveness of computer-tailored education on physical activity and dietary behaviors. Annals of Behavioral Medicine 31, 205–223 (2006)
4. Chatterjee, S., Price, A.: Healthy living with persuasive technologies: Framework, Issues, and Challenges. Journal of the American Medical Informatics Association 16, 171 (2009)
5. Holtzblatt, K., Jones, S.: Contextual Inquiry: A Participatory Technique for System Design. In: Schuler, D., Namioka, A. (eds.) Participatory Design: Principles and Practices, pp. 177–210. Lawrence Erlbaum, N.J. (1993)
6. Van der Lelie, C.: The value of storyboards in the product design process. Personal and Ubiquitous Computing 10, 159–162 (2006)
7. Snyder, C.: Paper prototyping. Morgan Kaufmann, San Francisco (2003)
8. Wright, P., Monk, A.: The use of think-aloud evaluation methods in design. ACM SIGCHI Bulletin 23, 55–57 (1991)
9. Bardram, J., Bossen, C., Lykke-Olesen, A., Nielsen, R., Madsen, K.H.: Virtual video prototyping of pervasive healthcare systems. In: Proceedings of the 4th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques, pp. 167–177. ACM, London (2002)
10. Grudin, J.: Why CSCW applications fail: problems in the design and evaluationof organizational interfaces. In: Conference Why CSCW Applications Fail: Problems in the Design and Evaluationof Organizational Interfaces, pp. 85–93. ACM, New York (1988)
11. Lazzaro, N.: Why We Play: Affect and the Fun of Games: Designing Emotions for Games, Entertainment Interface and Interactive Products. In: Sears, A., Jacko, J.A. (eds.) The Human-Computer Interaction Handbook, pp. 679–700. Lawrence Erlbaum, N.Y. (2006)
12. Baranowski, T., Buday, R., Thompson, D., Baranowski, J.: Playing for real: video games and stories for health-related behavior change. American Journal of Preventive Medicine 34, 74 (2008)
13. Krukowski, R., Harvey-Berino, J., Kolodinsky, J., Narsana, R., DeSisto, T.: Consumers not use or understand calorie labeling in restaurants. Journal of the American Dietetic Association 106, 917–920 (2006)
14. Moallem, A.: Excellence in ease of use with rich functionality how enterprise software applications with rich functionality can be built to excel in ease of use. In: Conference Excellence in ease of use with rich functionality how enterprise software applications with rich functionality can be built to excel in ease of use, pp. 672–681. Springer, Heidelberg (2007)
15. Reeves, M., Rafferty, A.: Healthy lifestyle characteristics among adults in the United States, 2000. Archives of Internal Medicine 165, 854 (2005)

Novel Human-Centered Rehabilitation Robot with Biofeedback for Training and Assessment

Runze Yang, Linhong Ji, and Hongwei Chen

Department of Precision Instruments and Mechanology , Tsinghua University
10084 Beijing,P.R.China
runzeyang@gmail.com, {jilh,hwchen}@tsinghua.edu.cn

Abstract. We present the novel human-centered rehabilitation methods from the research as well as literature to provide the robot assisted rehabilitation control strategies and motor function assessment methods. The research is based on the upper extremity compound movements (UECM) rehabilitation training robot [1], which is applied to the rehabilitation of upper extremity functions in patients with movement disorders. So called "human-centered" [2]or "patient-cooperative" strategies can take into account the patient's individual situations, intentions and efforts rather than imposing predefined instructions. It is considered that such robot-assisted methods can improve the therapeutic outcome compared to classical rehabilitation methods.

Keywords: Human-centered; Rehabilitation robot; Control strategy; Biofeedback.

1 Human-Centered Rehablilitation Methods

Many clinical studies support the effectiveness of the robot-assisted rehabilitation training, particularly patients after stroke [3]. Repetitive movements can improve muscular function and neural activity in patients, as well as prevent complications such as muscle atrophy, osteoporosis, etc.

Manually training by therapist has several limitations, such as high labor intensity, lacks of experience, training duration, assessment and so on. In contrast, the rehablilitation robot is consistent and customized to the patient, thus the appropriate afferent sensory input, which is believed to be important in improving the status of limb muscular activity and nervous activity [4], can be activated in the normal way.

In recently years, there is an increasing emphasis on "human-centered" robotic system for human rehablilitation after neurological injury[2]. The "human-centered" strategy means close interaction between the rehablilitation robot and the human, meanwhile the human is always in the center of the therapy. As the abnormal muscular tone may cause strain or dislocation of patient's joint[5] and other problems, rehabilitation robot should be safe, flexible, interactive and gentle toward the patient and therapist[6]. In consideration of the synergies of joints and muscles, the robot should also assist the patient's movemont adaptability and behave in a patient-cooperative way.

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 472–478, 2011.

The development of novel rehabilitation strategies such as virtual reality[7] and automatic adaptation control method can improve the interaction between the rehabilitation robot and the patient with hemiplegia or aphasia. Optimal training effects also depend on appropriate feedback and assessment about patient's individual performance.

Several researchers and groups have developed different rehabilitation robots, for example, MIT-MANUS[8] by MIT, LOKOMAT[9] & ARMin[10] by ETH Zurich, Intelligent Robotic Arm[11] by Chicago Rehabilitation Inst, etc.

In our study, the novel human-centered robot mentioned above has been developed. The robot provide different training modes(such as passive movement, impedance movement[12], patient-driven motion reinforcement movement[13],etc). The robot arm move the patient's limb in a defined trajectory , can also recognize and follow the patient's movement intention. Or allow the patient to deviate a certain range from a predefined reference trajectory in impedance control strategy. Allow the patient to reinforce the normal patterns of muscular abilities and give them feedback just like the therapist do.

We will present the human-centered principles that applied in the robot in this paper.

2 The Control Strategies

The robot consist of two-joint robotic arm that is used along with a limb-support display panel. The limb is actuated by a fixing bracket with bi-axes diver system. The panel can display the limb trajectory real-time and instructions ahead. The panel can also train patient in higher dimensions by changing its tilt angle.

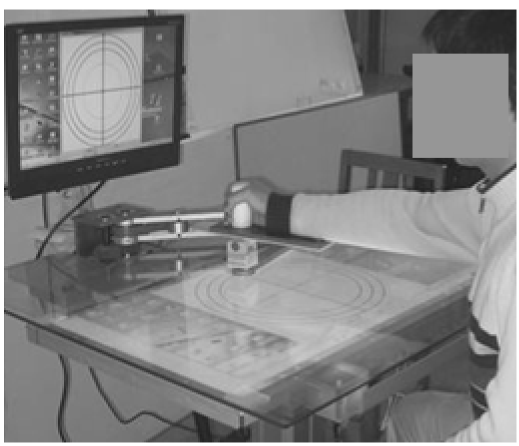


Fig. 1. The UECM rehabilitation robot[1]

The robot detect the patient's movement efforts and provide complex spacial trajectory training modes for the patients. Different control strategies are described and compared in this process: (1)The passive motion strategy, robot support patient's movements completely; (2) The motion reinforcement control strategy: robot assist patient's movements induced by patient ; (3) The impedance control strategy; (4)The adaptive control strategy[14] will be preformed in the future. The motion reinforcemen strategy and impedance control strategy are described in the following subsections.

Subjects were instructed to do exercise in circular trajectory with the robot arms in different training modes. The biofeedback and motivation method is applied in the process. Several training parameters, such as panel postures and movement speed, are varied to assess the effectiveness. They were also asked to change their performances and activity levels by observing visual feedback.

2.1 Motion Reinforcement Control Strategy

The actual movement induced by the patient is feedback into an inverse dynamic model to determine the robot contribution. So patient need to do some voluntary efforts to maintains the movement supported by the robot, and a scaling factor is used to change the supporting force.

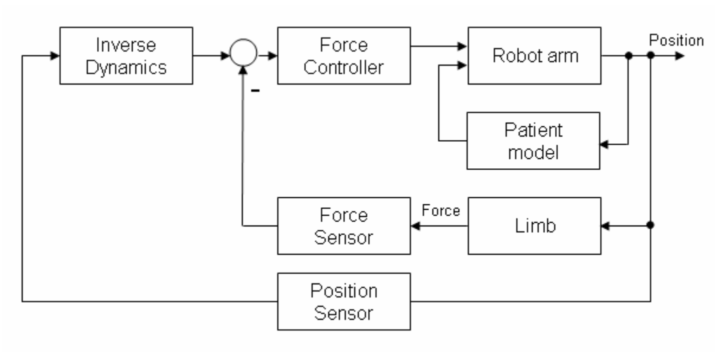


Fig. 2. Block diagram of motion reinforcement control strategy[15]

2.2 Impedance Control Strategy

Commonly, the impedance control strategy in rehabilitation robot allow deviation from a given trajectory and makes the robot compliantly[12]. Under such condition, the patient need to apply force to achieve the deviation trajectory. The deviation magnitude depends on the patient's muscular contribution, and keep the end of limb stay in a defined range along the trajectory by the impedance force .

It's considered that patients with paraplegia may be restricted and get less possibilities to change the muscular activation pattern bacause of the spasms[16].

2.3 Adaptive Control

In order to suit for the patient in severe spasm ,the adaptive control algorithms will be used and compared in the future. In this control strategy, patient will contribute voluntary muscular efforts to do his preferred trajectory under a reference [16]. Robot minimizes interaction forces between the patient and the robot, adapt the different impedance force to the individual patient's desired motion.

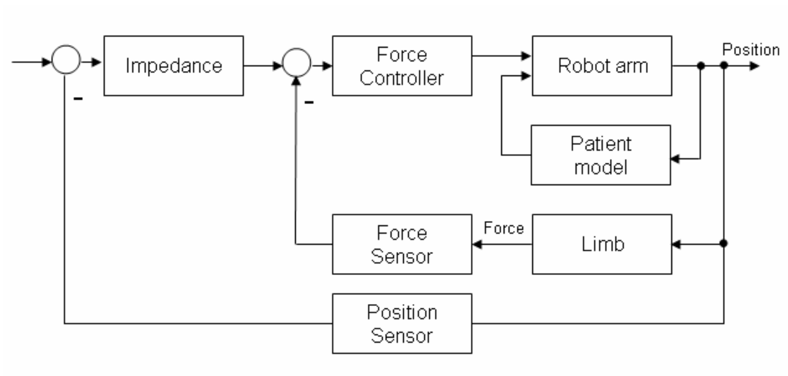


Fig. 3. Block diagram of impedance control strategy[15]

3 Biofeedback and Assessment

3.1 Overview

To promote the patient's active participation, several biofeedback elements are activated: stimulating patient's interest and neural activity; providing training status for the patient through real-time display of limb trajectory, speed, interactive force and tactile input; muscular activity and cardiopulmonary activity recordings; different control strategy for the individual patient.

The robot measures the track of upper extremity, movemont duration, range of motion by position sensors, as well as the reaction forces by force sensors on the fixing bracket. The muscular activity are also recorded by electromyographic (EMG). The recorded parameters of biomechanical and neurological states are processed and feed back to the patients via both the displays in front and the displays below the limbs [1].

Other parameters, such as heart rate, blood pressure, the spasticity evaluation, can also be measured and used as biofeedback values. These values are coonsider to be adequately important for presenting training condition and making rehabilitation plans for the therapist.

Robot-assisted assessment and feedback can extend and improve robot-assisted training therapy. The therapists can adapt the therapy and give further instructions to the patients. Combining robot-assisted training and assessment with biofeedback will make future therapy easier and more efficient.

3.2 Results

Results about the effects using different visual feedback method have been also obtained via experiment by healthy subjects. Variance analysis of the trajectory as follows (Heavy line: Desired trajectory; thin line: Actual trajectory; dotted line: Fitting average of actual trajectory).

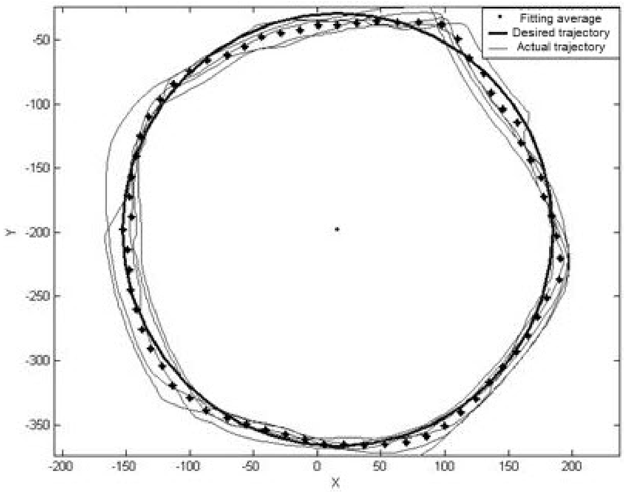


Fig. 4. Motion trajectory using visual feedback below limb (Visual feedback method 1)

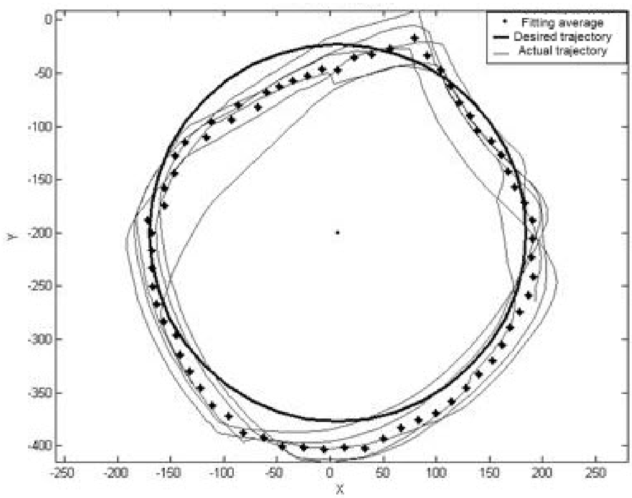


Fig. 5. Motion trajectory using visual feedback in front (Visual feedback method 2)

Table 1. Statistical deviation of trajectory[1]

Tilt angle of support panel	Horizontal (0°)	Lean to human side(35°)	Lean to other side (-10°)
Visual feedback method 1	0.7893±8.2095mm	-0.9773±13.735mm	0.7443±14.9341mm
Visual feedback method 2	4.4482±22.159mm	3.7021±21.7491mm	-4.8251±22.4571mm

We can see that the error band using visual feedback below limb (method 1) along with impedance control strategy is less as comparing the data using visual feedback in front (method 2) along with impedance control strategy. Meanwhile, the influence of the support panel tilting angle is more distinct in method 1 than method 2. In method 2, error is mainly composed of visual feedback factor rather than tilting angle. In the tilting panel condition, shoulder and elbow joints will stretch more, and deltoid muscle could get more training because of gravity verified by Motion Analysis systems[17] and SIMM module[18] in our previous experiments [19] .

4 Conclusions and Outlook

The common feature of control strategies mentioned above is the cooperatively interactive performance of the robot, including the compliant behaviour in impedance control strategy, supporting behaviour in motion reinforcement control strategy, adaptive behaviour in adaptive control strategy. All these strategies try to activate natural and muscular activities induced by the patient.

More function of the rehabilitation robot should be developed by imitating human therapist: for example, the rehabilitation robot can take appropriate treatment even when patient is passive and weak in the early stage of the rehabilitation.

Future clinical evaluations should be performed with a large population of different patient groups, including patients with hemiplegia after stroke, and others such as patients with Parkinson's disease.

Acknowledgments. We thank Lining Chen and Yupeng Ren for their contributions to this paper.

References

1. Chen, L.: Development of A Multi-posture Hemiplegia Rehabilitation Training Robot for Upper Extremity Compound-Motion. Tsinghua University, Beijing (2008)
2. Khatib, O., Brock, O., Chang, K.S., Ruspini, D., Sentis, L., Sriram, V.: Human-centered robotics and interactive haptic simulation. The Int. Journal of Robotics Research 23, 167–178 (2004)
3. Barbeau, H., Rossignol, S.: Enhancement of locomotor recovery following spinal cord injury. Current Opinion in Neurology 7, 517–524 (1994)
4. Hogan, N., Krebs, H.I., et al.: Robot-aided neurorehabilitation. IEEE Trans. Rehab. Eng. 6, 75–87 (1998)
5. O'Dwyer, N.J., Ada, L., Neilson, P.D.: Spasticity and muscle contracture following stroke. Brain 119, 1737–1749 (1996)
6. Stefanov, D., Bien, Z.Z.: Advances in Rehabilitation Robotics. LNCIS, vol. 306, pp. 25–44. Springer, Heidelberg (2004)
7. Riener, R., Wellner, M., Nef, T., et al.: A view on VR-enhanced rehabilitation robotics. In: 2006 International Workshop on Virtual Rehabilitation, New York, pp. 149–154 (2006)
8. Hogan, N., Krebs, H.I., et al.: MIT-MANUS: A Workstation for Manual and Training. In: IEEE International Workshop on Robot and Human Communication, Tokyo, pp. 161–165 (1992)
9. Colombo, G., Joerg, M., Schreier, R., Dietz, V.: Treadmill training of paraplegic patients using a robotic orthosis. J. Rehabil. Res. Dev. 37(6), 693–700 (2000)

10. Nef, T., Riener, R.: ARMin-Design of a novel arm rehabilitation robot. In: ICORR 2005, 9th International Conference on Rehabilitation Robotics, pp. 57–60 (2005)
11. Zhang, L.-Q., Park, H.-S., Ren, Y.: Developing an intelligent robotic arm for stroke rehabilitation. In: ICORR 2007, 10th International Conference on Rehabilitation Robotics, pp. 984–993 (2007)
12. Hogan, N.: Impedance control: An Approach to Manipulation. Journal of Dynamic Systems, Measurement, and Control 107, 1–23 (1985)
13. Riener, R., Fuhr, T.: Patient-driven control of FES-supported standing-up: A simulation study. IEEE Transactions on Rehabilitation Engineering 6, 113–124 (1998)
14. Jezernik, S., Colombo, G., Morari, M.: Automatic gait-pattern adaptation algorithms for rehabilitation with a 4-DOF robotic orthosis. IEEE Trans Robotics and Automation 20(3), 574–582 (2004)
15. Ren, Y.: Robotic Assist Control and Evaluation System for Upper Limb Rehabilitation. Tsinghua University, Beijing (2004)
16. Riener, R., Lünenburger, L., Colombo, G.: Human-centered robotics applied to gait training and assessment. JRRD 43, 679–694 (2006)
17. Motion Analysis Corporation, `http://www.motionanalysis.com/`
18. MusculoGraphics, Inc., `http://www.musculographics.com/`
19. Wang, Z., Chen, L., Yao, C., et al.: The influence of transforming operation plane on compound movement training in upper limb robot aided rehabilitation. Chinese Journal of Rehabilitation Medicine 24, 65–67 (2009)

Intensity Analysis of Surface Myoelectric Signals from Lower Limbs during Key Gait Phases by Wavelets in Time-Frequency

Jiangang Yang[1], Xuan Gao[1], Baikun Wan[1], Dong Ming[1,*], Xiaoman Cheng[2], Hongzhi Qi[1], Xingwei An[1], Long Chen[1], Shuang Qiu[1], and Weijie Wang[3]

[1] Department of Biomedical Engineering, College of Precision Instruments and Optoelectronics Engineering, Tianjin University, Tianjin, China
richardming@tju.edu.cn
[2] School of Science, Tianjin University of Technology
[3] Department of Orthopaedics and Traumatology, Ninewells Hospital, University of Dundee, UK

Abstract. This paper presented a time-frequency intensity analysis feature extraction approach of lower limb sEMG (Surface Electromyogram) to identify the key gait phases during walking. The proposed feature extraction method used a filter bank of non-linearly scaled wavelets with specified time-resolution to extract time-frequency aspects of the signal.The intensity analysis algorithm was tested on sEMG data collected from ten healthy young volunteers during 30 walking circles for each. Each walking cycle was made up of four key gait phases:L-DS(Left Double Stance), L-SS(Left Single Stance),R-DS(Right Double Stance),R-SS(Right Single Stance).The identification accuracy of 7 subjects using intensity analysis reached 97%, even up to 99.42%.The others were about 95%. The algorithm obviously achieved a higher accuracy of sEMG recognition than the other algorithms such as root mean square and AR Coefficient. In the future, the feature of sEMG signal under different key gait phases may be used in the control of Functional Electrical Stimulation (FES) and other intelligent artificial limbs.

1 Introduction

The surface myoelectric signal, recorded as surface electromyogram, is the recording of the myoelectric activity from the surface of skeletal muscles [1]. It is closely related to the size and activity of the muscle and a measure of the functional state of muscle fibers [2-3], and is useful to estimate the strength of contraction of the muscle. sEMG recognition and classification has been widely applied to many applications such as rehabilitation, design of mechatronic systems for prosthesis, and human-robot interaction communication. The focus of the ongoing research in this field is on how to enhance the sEMG recognition accuracy. Like other pattern recognition problems, sEMG classification challenges researchers with the same two issues: feature selection and classifier design. Previous studies [4] have shown that the representation

* Correspondence Author.

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 479–488, 2011.

of the sEMG signal plays an important role in myoelectric pattern recognition systems[5]. Time and frequency analysis methods have been widely used for feature extraction due to their obvious physical interpretation and convenient computation. For time domain analysis, the methods used for feature extraction include mean absolute value (MAV), mean absolute value slope (MAVSLP), zero crossing (ZC), slope sign changes (SSC), waveform length (WL), wave complexity (WC)[6,7,8] , root mean square (RMS)[9], autoregressive (AR) coefficients[10,11,12], and multivariate AR model[13]. For frequency analysis, the methods include Fourier transform coefficients [14] and cepstral coefficients [15]. More recently, instead of analyzing the sEMG signal independently in the time or frequency domain, time-frequency methods, such as short time Fourier transform, wavelet, and wavelet package, have been introduced for feature extraction[16].

The sEMG data collected in our study was analyzed to distinguish the four key gait phase of 30 periods. A decomposition of the signal into events requires a time resolution sufficiently. A time-frequency analysis based on wavelets is probably a more appropriate tool for such an approach. The concept of intensity represents a quantitative analysis that approximates the power of the EMG signal at time t. The intensity yields the timing and the strength of the event of muscle activation at various frequencies. Events with intensities that occur at one frequency can then be compared with those occurring at other ones. Once such events were measured, the timing of muscle activation in biomechanical studies can greatly be improved. The filter-bank of non-linearly scaled wavelets will be used to achieve the intensity measurements.

2 Theory of Intensity Analysis

2.1 Definition of the Wavelets

The wavelets used for the intensity analysis were defined in frequency space and wavelets in time space were obtained applying the inverse Fourier transform. The wavelets were functions of the frequency defined by the parameters, center frequency cf and scale.

$$F\psi(f, cf, scale) := \left(\frac{f}{cf}\right)^{cf \cdot scale} \cdot \exp\left(\frac{-f}{cf} + 1\right) \cdot cf \cdot scale \tag{1}$$

f frequency in Fourier respectively frequency space (index k), cf center frequency, scale scaling factor. The center frequencies indicate the position of the maximum of the wavelet in frequency space. A whole set of wavelets indexed by j 0 to J was generated with center frequencies that were calculated as follows:

$$cf_j = \frac{1}{scale} \cdot (j + q)^{r} \tag{2}$$

All wavelets have amplitude 1 at the center frequency and the $(F\psi)^2$ is again a wavelet of the same type. The scale defines the frequency range covered by the wavelets. For EMG signals a scale of 0.3 was selected and the parameters describing the wavelets are shown in Table 1. Changing the scale is the preferred way of changing the frequency range covered by the wavelet. A different scale has also to be selected to obtain a different time-resolution. The number of wavelets J defines the extent of the frequency range covered. The parameters q and r were used to optimize the spacing of the wavelets. The criterion used to determine q and r was that the sum of all wavelets should be as close as possible constant in a restricted frequency range of 20 to 200 Hz.

Table 1. Parameters of wavelets calculated for scale 0.3,q 1.45,r 1.959

Index j of wavelet	Center-frequency Hz	Time-resolution ms	Band-width Hz	gm %
0	6.90	76.50	9.77	5.29
1	19.29	59.00	15.63	4.14
2	37.71	40.50	21.48	3.75
3	62.09	31.50	27.34	3.58
4	92.36	26.00	35.16	3.50
5	128.48	21.50	41.02	3.46
6	170.39	19.50	46.88	3.44
7	218.08	16.50	52.73	3.43
8	271.50	15.00	58.59	3.43
9	330. 63	13.50	66.41	3.43
10	395. 46	12.00	72.27	3.44

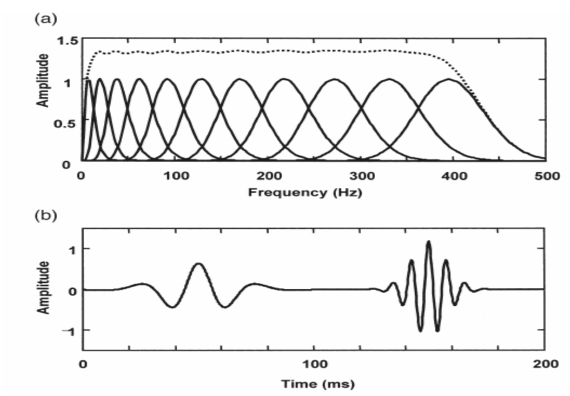


Fig. 1. (a)Filter-bank of 10 wavelets in frequency space (solid line). Sum of all wavelets (dotted line). (b) Wavelet 2 and 5 in time space shifted by 100 and 150 ms respectively.

Usually wavelets are derived from a mother wavelet by linear scaling. The present wavelets are however, not obtained by linear scaling of a mother wavelet. A wavelet consists of a couple of oscillations and as a consequence of nonlinear scaling those at lower j have few oscillations whereas the higher order ones show more oscillations (Fig. 1b).

2.2 The Wavelet Transform and Wavelet Analysis

For a signal of finite duration the convolution with one wavelet can be done according to the convolution theorem as a multiplication in frequency space followed by an inverse Fourier transform. Thus if Fs_k represents the Fourier transformed signal in frequency space and k represents the index of the frequency increments of width f then the convolution in frequency space is:

$$Fw_{j,k} = F\psi_{j,k} \cdot Fs_k \tag{3}$$

This represents the Fourier transform of the wavelet transformed signal for wavelet j. The inverse Fourier transform yields w the signal convoluted by the wavelet. Repeating the wavelet transform for all wavelets is called wavelet analysis. The sum of all w divided by the plateau value yields the signal as it would have been submitted to the band pass filter given by the sum of all wavelet type filters in frequency space. This sum is therefore an approximation of the original signal within the accuracy given by the SD of the plateau value.

2.3 The Analysis of Intensities

To calculate the intensity one starts performing the following wavelet transform:

$$Fv_{j,k} = F\psi c_{j,k} \cdot Fs_k \tag{4}$$

The intensity in time space was defined for each wavelet as:

$$p_{j,n} = \left(v_{j,n}\right)^2 + \left(\frac{1}{2\pi c f_j} \cdot \frac{d}{dt} v_{j,n}\right)^2 \tag{5}$$

The derivative can be calculated numerically or using the property of the Fourier transform on time derivatives[17]:

$$\frac{d}{dt} v_{j,n} = F^{-1}\left[i\left(2 \cdot \pi \cdot f_k \cdot Fv_{j,k}\right)\right] \tag{6}$$

F^{-1} indicates the inverse Fourier transform.

The goal to be achieved was that the amplitude A^2 of a pure sine wave of frequency f_k in frequency space equals the total intensity in wavelet space. This was taken care of by multiplying the wavelets with an intensity factor c_k before performing the calculation of the intensities. The intensity factor was obtained considering the following:

$$(A_K)^2 = (A_k c_k)^2 \left[(F\psi_{K,j-1})^2 + (F\psi_{K,j})^2 + (F\psi_{K,j+1})^2 \right] \tag{7}$$

With

$$c_k = \frac{1}{\sqrt{\left[(F\psi_{K,j-1})^2 + (F\psi_{K,j})^2 + (F\psi_{K,j+1})^2 \right]}} \tag{8}$$

j has to be selected to encompass the three wavelets where the values of $F\psi$ are not 0. Inserting ck in the above equation yields the equality A2=A2. The approximation was further improved multiplying the wavelet by a small damping factor d taking especially into account that the lowest frequencies were overestimated by the intensity calculation. Thus the wavelets used for the intensity calculation were:

$$F\psi c_{j,k} = F\psi_{j,k} \cdot c_k \cdot d_j \tag{9}$$

Where d_0 and d_1 of 0.816 and 0.912 and d 0.96 otherwise yielded optimal results.

3 Methodology

3.1 Subjects

The sEMG data was collected from ten healthy young volunteers (no skeletal and muscular diseases, age between 22 and 25, eight males and two females) during walking at about 0.45 m/s speed. It was recorded by a four-channel acquisition system. There was a pair of electrodes with the distance of 1.5 cm along the direction of the muscle tension mounted by medical rubberized fabric. The electrodes were placed on two muscles: Vastus Rectus and Biceps on each leg. A reference electrode was placed at the surface of the knee on the right leg. In this paper, the electrodes on the Vastus Rectus and Biceps of the left leg were identified as L-1 and L-2 and those of the right leg were marked as R-1 and R-2.

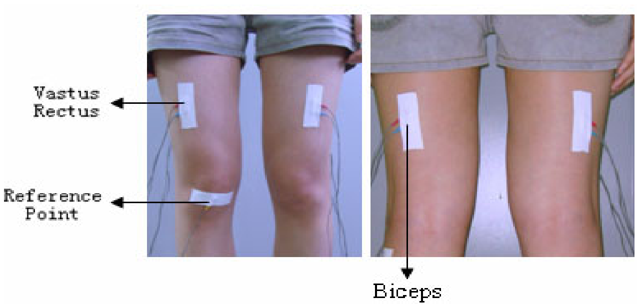


Fig. 2. Placement of electrodes for the experiments

3.2 Experimental Setup

In the experiment, the sampling rate was 1000 Hz per channel. Analog filters were used to filter the sEMG signal to the 20–500 Hz frequency range preventing the interference of 50 Hz. To divide the phases of walking circle, two trigger signal circuit switches were placed under feet to generate synchronous signal. When Mechanical switch was pressed (heel landing) the circuit outputs a high level; when Mechanical switch bounced (heel lift) the circuit outputs a low level. The trigger signal circuit switch is shown in Fig.3. Corresponding relationship of combination state of synchronous signal and key gait phases are shown in Table 2.

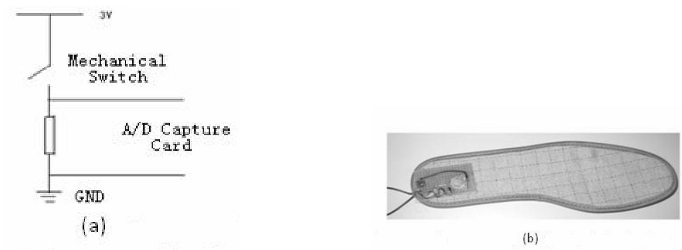


Fig. 3. (a) Circuit of trigger signal circuit switch (b) Trigger signal circuit system

3.3 Experimental Procedure

The subject was asked to walk along one direction for about 30 gait circles. A complete gait cycle began when the heel of one foot hit the ground and ends when it hit the ground again. Each gait cycle was made up of four key gait phases, L-DS(Left Double Stance), L-SS(Left Single Stance),R-DS (Right Double Stance),R-SS(Right Single Stance). A gait cycle and its phases are shown in Fig.4.

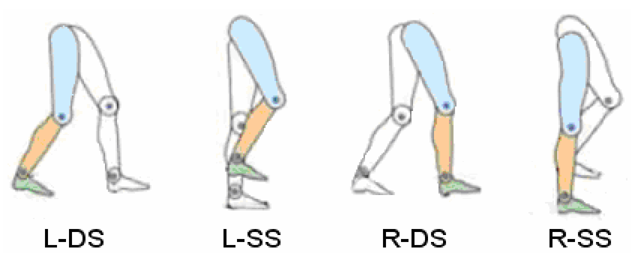


Fig. 4. A walking cycle and its phases

Table 2. Corresponding relationship of combination state of synchronous signal and key gait phases

Key Gait Phases	Right Level	Left Level
L-DS	1	1
L-SS	0	1
R-DS	1	1
R-SS	1	0

4 Results and Discussions

The sEMG signal for each key gait phase was resampled by 500 points. In Fig. 5 and Fig. 6, the data was the average intensity of different key gait phases in the first ten-cycle of one subject. As the sEMG of right leg was symmetrical to the left, the sEMG signal of two channels on the left leg was made into topographical map of intensity.

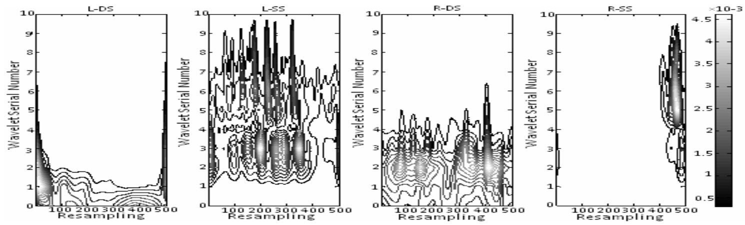


Fig. 5. The intensity analysis by wavelets in the first 10 circle on L-1

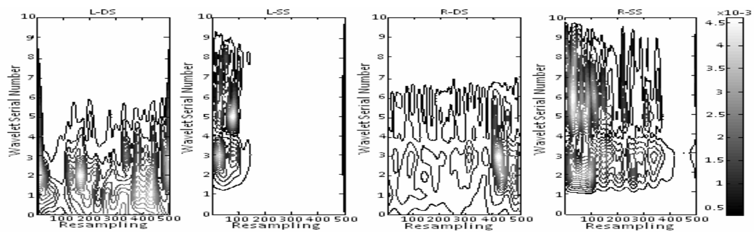


Fig. 6. The intensity analysis by wavelets in the first 10 circle on L-2

Fig. 5 showed that in the L-DS phase, the high intensity of the sEMG signal on L-1 (rectus femoris) mainly distributed on wavelet 1 and wavelet 2 in the 0 to 10% of this phase; in L-SS phase, the high intensity of the sEMG signal mainly distributed on wavelet 3 and wavelet 6 in the 40% to 80%; in the R-DS phase, the high intensity of the signal occurred mainly on wavelet 1 and wavelet 2 in the 20% to 40% and 80% to 100%; in R-SS phase, the high intensity of the sEMG signal mainly distributed on wavelet 3,5,6,7 in the 90% to 100%.

In the L-DS phase of Fig. 6, the high intensity of the sEMG signal on L-2 (biceps femoris) mainly distributed on wavelet 1,2,3 in the 0 to 10% and 80% to 100% of this phase; in L-SS phase, the high intensity of the sEMG signal mainly distributed on wavelet 2,3,5 in the 0% to 20% ; in the R-DS phase, the high intensity of the signal mainly occurred on wavelet 3 in the 80% to 100%; in R-SS phase, the high intensity of the sEMG signal mainly distributed on wavelet 5,6,7 in the 0% to 20%.

To prove the stability of the data, the average intensity distribution of different key gait phases in the second ten-cycle was shown in Fig.7.

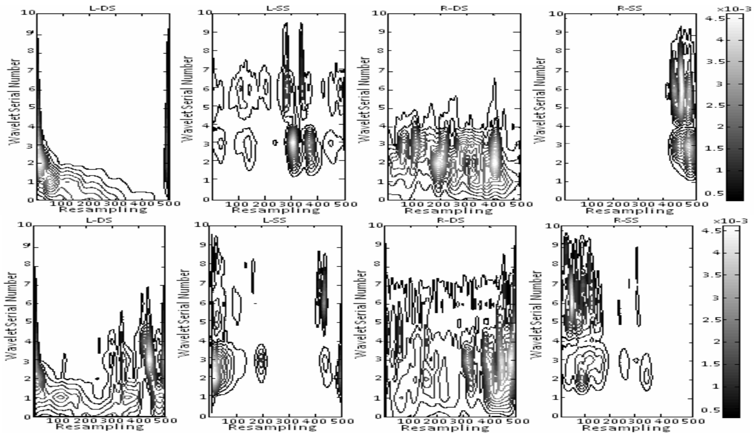


Fig. 7. The intensity analysis by wavelets in the second 10 circle from L-1 and L-2

The intensity distribution in the topographical map showed that there were obvious differences in the distribution of intensity of four key gait phases.

As shown in Table 3, the identification accuracy of 7 subjects using intensity analysis reached 97% or more-up to 99.42%.The others were about 95%.It was obviously better than root mean square and AR Coefficient.

Table 3. Rate of key gait phase Identification of limb walking

Subject	Rate of key gait phase Identification of limb walking (%)		
	RMS (root mean square)	AR Coefficient (N is the order)	Intensity Analysis
1	97.25	99.25(N=4)	98.21
2	93.33	97.79(N=4)	99.42
3	96.26	98.08(N=4)	98.50
4	91.96	92.92(N=4)	94.13
5	99.08	96.54(N=4)	92.92
6	98.32	94.08(N=4)	95.37
7	99.08	99.25(N=3)	98.46
8	93.17	99.21(N=3)	99.17
9	99.29	99.42(N=5)	98.71
10	90.06	98.42(N=3)	97.54

5 Conclusion

This study demonstrated an initial attempt to develop and apply a time-frequency analysis of the intensities for identifying walking phases. The results of this study suggested that the concept of intensity analysis was viable for the feature extraction of sEMG from lower limbs during key gait phases.In the future, the feature of sEMG signal under different key gait phases may be used in the control of Functional Electrical Stimulation (FES) and other intelligent artificial limb.

Acknowledgment. This research was partially supported by National Natural Science Foundation of China (No.90920015, No.30970875, No.60501005), Joint Project of National Natural Science Foundation of China –Royal Society of Edinburgh of UK (No.30910494/C1009), The National High Technology Research & Development Program of China (No.2007AA04Z236), Biomedical Engineering Key Program (No.07ZCKFSF01300) and Intentional Cooperation Key Program (No.08ZCGHHZ00300) of Tianjin Science Technology Support Plan.

References

1. Ganesh, R.N.: Use of sEMG in identification of low level muscle activities features based on ICA and Fractal dimension. In: 31st Annual International Conference of the IEEE EMBS, pp. 364–367. IEEE Press, Minnesota (2009)
2. Duchêne, J., Goubel, F.: Surface electromyogram during voluntary contraction: processing tools and relation to physiological events. J. Cri. Rev. Biomed. Eng. 21(4), 313–397 (1993)
3. Karlsson, S., Yu, J., Akay, M.: Time-frequency analysis of electromyogram during dynamic contractions: a comparative study. IEEE Transactions on Biomedical Engineering 47(2), 228–238 (2000)
4. Chen, X., Zhu, X., Zhang, D.: A discriminant bispectrum feature for surface electromyogram signal classification. J. Medical Engineering & Physics 32, 126–135 (2010)
5. Englehart, K., Hudgins, B., Parker, P., Stevenson, M.: Classification of the myoelectric signal using time-frequency based representations. J. Medical Engineering and Physics 21(6-7), 431–438 (1999)
6. Hudgins, B., Parker, P., Scott, R.: A new strategy for multifunction myoelectric control. IEEE Transactions on Biomedical Engineering 40(1), 82–94 (1993)
7. Englehart, K., Hudgins, B.: A robust, real-time control scheme for multifunction myoelectric control. IEEE Transactions on Biomedical Engineering 50(7), 848–854 (2003)
8. Oskoei, M., Hu, H.: Support vector machine-based classification scheme for myoelectric control applied to upper limb. IEEE Transactions on Biomedical Engineering 55(8), 1956–1965 (2008)
9. Shenoy, P., Miller, K., Crawford, B., Rao, R.: Online electromyographic control of a robotic prosthesis. IEEE Transactions on Biomedical Engineering 55(3), 1128–1135 (2008)
10. Huang, Y., Englehart, K., Hudgins, B., Chan, A.: A Gaussian mixture model based classification scheme for myoelectric control of powered upper limb prostheses. IEEE Transactions on Biomedical Engineering 52(11), 1801–1811 (2005)

11. Hargrove, L., Englehart, K., Hudgins, B.: A comparison of surface and intramuscular myoelectric signal classification. IEEE Transactions on Biomedical Engineering 54(5), 847–853 (2007)
12. Chiou, Y., Luh, J., Chen, S., Lai, J., Kuo, T.: The comparison of electromyographic pattern classifications with active and passive electrodes. J. Medical Engineering and Physics 26(7), 605–610 (2004)
13. Hu, X., Nenov, V.: Multivariate AR modeling of electromyography for the classification of upper arm movements. J. Clinical Neurophysiology 115(6), 1276–1287 (2004)
14. Peleg, D., Braiman, E., Yom-Tov, E., Inbar, G.: Classification of finger activation for use in a robotic prosthesis arm. IEEE Transactions on Neural Systems and Rehabilitation Engineering 10(4), 290–293 (2002)
15. Micera, S., Sabatini, A., Dario, P., Rossi, B.: A hybrid approach to EMG pattern analysis for classification of arm movements using statistical and fuzzy techniques. J. Medical Engineering and Physics 21(5), 303–311 (1999)
16. von Tscharner, V.: Intensity analysis in time-frequency space of surface electromyogram by wavelets of specified resolution. Journal of Electromyography and Kinesiology 10, 433–445 (2000)
17. Mallat, S.G.: A wavelet tour of signal processing. Academic Press, San Diego (1998)

Handle Reaction Vector Analysis with Fuzzy Clustering and Support Vector Machine during FES-Assisted Walking Rehabilitation

Weixi Zhu[1], Dong Ming[1,*], Baikun Wan[1], Xiaoman Cheng[2], Hongzhi Qi[1], Yuanyuan Chen[1], Rui Xu[1], and Weijie Wang[3]

[1] Department of Biomedical Engineering, College of Precision Instruments and Optoelectronics Engineering, Tianjin University, Tianjin, China
richardming@tju.edu.cn
[2] School of Science, Tianjin University of Technology, Tianjin, China
[3] Department of Orthopaedics and Traumatology, Ninewells Hospital, University of Dundee, UK

Abstract. This paper proposed Fuzzy clustering of C means and K means methods to extract the lateral features of lower limbs movement from handle reaction vector (HRV)data. With C-means clustering, the SVM recognition rate of lateral features was usually above 90% while, with K-means clustering, the recognition rate was close to 85%. The best recognition rate was even reaching up to 97% for some individual subject. Then the samples from all subjects were processed together with the cross-validation. Our experimental results showed that the HRV signal could be used with fuzzy clustering and support vector machine to effectively classify the lateral features of lower limbs movement. It may provide a new choice for FES control signal. The optimizing of the algorism parameters can be introduced to get better control in the future.

1 Introduction

Paraplegic is impairment in motor and/or sensory function of the lower extremities. It is usually the result of spinal cord injury (SCI) [1] which affects the neural elements of the spinal canal. This means that paraplegic patients lose the voluntary control of neurological functions below the level of instinct. Inability to stand and walk, loss of sensation, bedsores, joint contractions, worsening of the cardiopulmonary function and loss of bladder and bowel control, etc. are the catastrophic impacts.

Rehabilitation is one of the methods to improve the life qualities of paraplegic patients. The most important aspect in the rehabilitation of paraplegic patients is to help them enhance or regain their lost functions, such as standing and walking [2]. To improve the gait and efficacy of paraplegic walking, functional electric stimulation (FES) is one of the good rehabilitation tools in helping paraplegic patients to regain their abilities to walk [3]. During FES-assisted paraplegic walking, a standard walker is used to help a patient to support his/her body and move forward by means of his/her upper limbs.

* Corresponding Author.

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 489–498, 2011.

In the development of a FES system, one of FES key techniques is to choose its effective control signals. This paper proposed the force applied by the upper body of paraplegic patients to the walker, i.e. handle reaction vector (HRV), as a new FES control signal and tested its feasibility through the recognition of the lateral features of lower limbs movement from HRV signals during FES-assisted walking.

The force component of the vector in the x-,y-and z-axis can be respectively characterized the forward force, balance force and support force obtained by using walker. Among them, the definition of the coordinate system is set by the x-axis positive to the right of patients, the y-axis positive to the forward of patients,the z-axis positive to the upward of patients. Then the definition formula of HRV can be written as

$$[\mathrm{HRV}]=[\mathrm{HRV}_{\mathrm{l}},\mathrm{HRV}_{\mathrm{r}}]^{\mathrm{T}}=[F_{lx},F_{ly},F_{lz},F_{rx},F_{ry},F_{rz}]^{T}$$

To measure the HRV data during FES-assisted walking, a walker dynamometer system was developed with a 12-channel strain gauge bridge network as shown in Fig. 1.

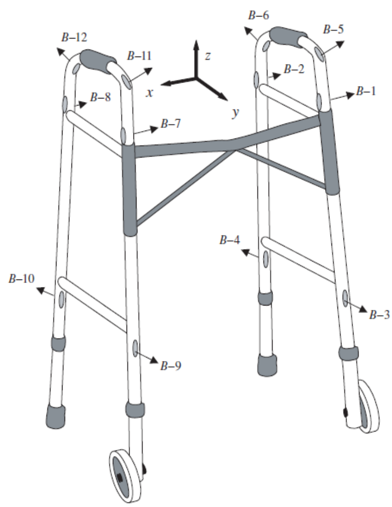


Fig. 1. Walker instrumentation with 12 strain gauge bridges [4]

Twelve strain gauge bridges, B-1 to B-12, were mounted on a standard walker frame to detect resultant forces in 3-dimensions. The forces were in lateral (X-axis), forward (Y-axis) and upward (Z-axis) directions.A full Wheatstone bridge was used with a bending pattern for the measurement of bending moments. The positions of the 12 strain gauge bridges were determined according to the bending moment distributions of the walker frame, and were calculated using the finite element analysis (FEA) method. The sensing data got by the twelve strain gauge bridges were amplified and filtered. Then they were transmitted by a cable to a PC for further analyzing and calculating the HRV.

2 Methods

2.1 Experiment and Data Collection

There were ten subjects, including 5 males and 5 females, involved in our experiment to collect the HRV data during FES-assisted walking. The subjects were required to follow a four-step training for walking until they were able to correctly complete a single continuous movement, then the formal test began. After the experiment, we obtained the upper three-dimensional force information of each subject. They were respectively marked as Flx, Frx, Fly, Fry, Flz, Frz. The figures 2-4 show the three-dimensional information of a subject.

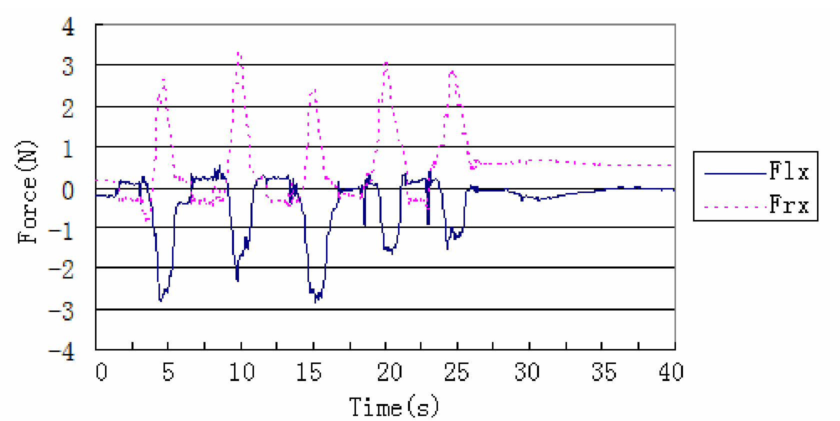


Fig. 2. Upper three-dimensional X direction force signal information

As shown in Fig.2,the X direction force which obtained by using walker characterized the forward force. In the experiment we got 5 standard gait cycles.The red line(Frx) was almost positive and the max amplitude was just more than 3 N. On the contrary,the blue one(Flx) was almost negative and the max amplitude was less than 3 N.

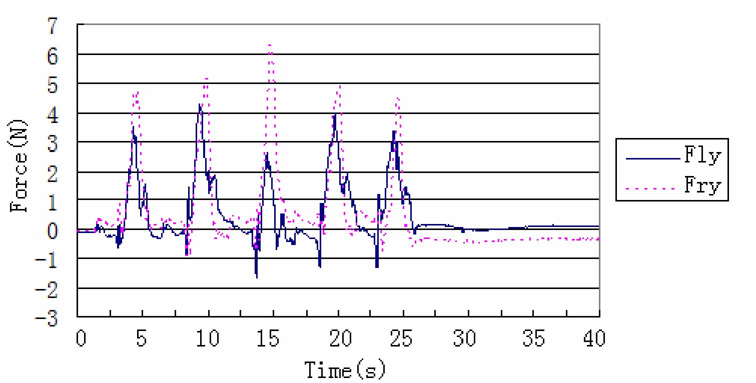


Fig. 3. Upper three-dimensional Ydirection force signal information

In Fig.3,the Y direction force characterized the balance force. In the 5 standard gait cycles,the red line(Fry) was almost positive and the max amplitude was a little more than 6 N. The blue one(Fly) was also positive and the max amplitude was a little more than 4 N.

In Fig. 4, the Z direction force characterized the support force. In the 5 standard gait cycles,the red line(Frz) was almost negative and the max amplitude was a little more than 20 N. The blue one(Flz) was also negative and the max amplitude was less than 20 N.

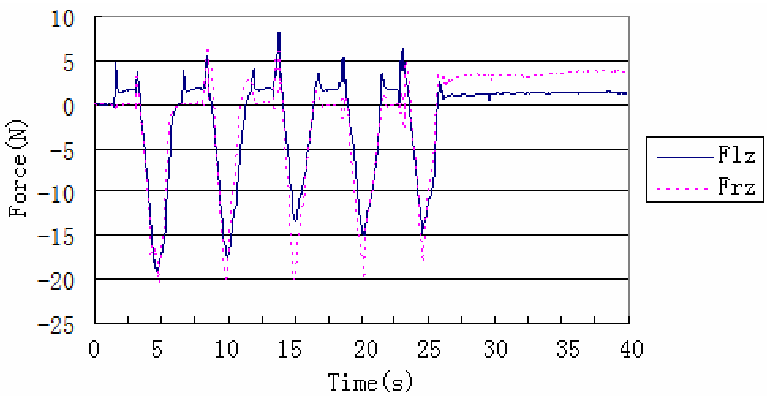


Fig. 4. Upper three-dimensional Z direction force signal information

2.2 Algorithm

Based on six-axis forces to explore walker signals reflecting different users situation of the lower extremities moving, we attempt to use a feature vector non-supervised learning method that is data clustering to evaluate the reliability of HRV measurement.

Fuzzy c-means Algorithm (FCM). Fuzzy c-means (FCM) [5-9]is a data clustering technique where in each data point belongs to a cluster to some degree that is specified by a membership grade. This technique was originally introduced by Jim Bezdek[10,11] as an improvement on earlier clustering methods. It provides a method that shows how to group data points that populate some multidimensional space into a specific number of different clusters.

This iteration is based on minimizing our objective function that represents the distance from any individual event to our cluster center weighted by individual event's membership grade. The algorithm is an iterative optimization that minimizes the objective function defined as follow:

$$J_m = (u,v) = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^m d^2(x_j, v_i) \tag{1}$$

With the following constraints:

$$\sum_{i=1}^{c} u_{ij} = 1, \forall j \tag{2}$$

Where $X = (x_1, x_2, \cdots x_j, \cdots x_n)$ is dimensional vector space for each force and n represents the number of feature vectors. u_{ij} is the membership of the j th data in the i th cluster c_i , m is a constant, the parameter m controls the fuzziness of the resulting partition. Using the Euclidean norm, the distance metric d measure the similarity between a feature vector x_j and a cluster centroid v_i in the feature space, i.e.:

$$d^2(x_j, v_i) = \left\| x_j - v_i \right\|^2 \tag{3}$$

The objective function is minimized when data points close to the centroid of their clusters are assigned high membership values, and low membership values are assigned to data points far from the centroid. The membership functions and cluster centroids are updated by the following expressions:

$$u_{ij} = \{ \sum_{k=1}^{c} [\frac{d(x_j, v_i)}{d(x_j, v_k)}]^{2/m-1} \}^{-1} \tag{4}$$

And

$$v_i = \frac{\sum_{j=1}^{n} u_{ij}^m x_j}{\sum_{j=1}^{n} u_{ij}^m} \tag{5}$$

The FCM algorithm proceeds by iterating the two necessary conditions until a solution is reached. Each data point will be associated with a membership value for each class after FCM clustering. By assigning the data point to the class with the highest membership value, a segmentation of the data could be obtained. Concrete steps are as follows:

Step 1: Set the number of clusters c and the parameter m in (1). Initialize the fuzzy cluster centroid vector $V = [v_1, v_2, \cdots v_c]$ randomly and set ε.

Step 2: computer u_{ij} by (4).

Step 3: update v_i by(5).

Step 4: update u_{ij} by (4).

Repeat Steps 3 and 4 until the following termination criterion is satisfied:

$$\left| V_{new} - V_{old} \right| < \varepsilon$$

With fuzzy c-means feature extraction of the upper force information, there are two characteristic values of each dimension force in a gait cycle .They are respectively function value of each clustering center. Six groups forces are all together 12 eigenvalues.

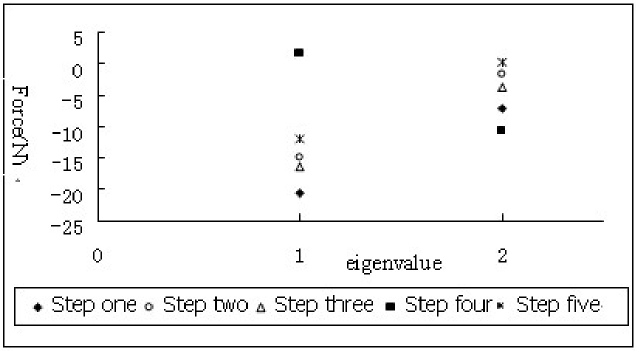


Fig. 5. FLz dimension force C - mean clustering results

As shown in Fig.5, the ordinate is the changes of FLz dimensional force in gait cycles.We can see that the right move is first in step two and the remaining are the left first.So the two eigenvalues obtained by C-means clustering are divisible.

Fuzzy k-Means Algorithm(FKM). The FKM algorithm [12] classifies each vector to all clusters with different values of membership between 0 and 1. This membership value indicates the association of a vector to each of the k clusters. Notice that the FKM algorithm does not classify fuzzy data, but crisp data into fuzzy clusters. We bring forward a modified subordination function u_{ij} [13], u_{ij} refers to the subordination degree of the ith sample to the jth category. We define a creterion function of clustering, that is

$$J = \sum_{i=1}^{k} \sum_{j=1}^{n} [u_{ij}]^{\beta} \left\| x_i - m_j \right\|^2 \tag{6}$$

and the constriction condition is

$$\sum_{i=1}^{k} \sum_{j=1}^{n} u_{ij} = c \tag{7}$$

For every vector V_i , i = 1,…n, its subordination function u_{ij} is compute as the following:

$$u_{ij} = \frac{(1/\left\| v_i - x_j \right\|^2)^{1/c-1}}{\sum_{i=1}^{k} (1/\left\| v_i - x_i \right\|^2)^{1/c-1}} \text{,j=1,2,…,k;c>1} \tag{8}$$

According to fussy theory, the FKM algorithm is described as following.
Input: clustering number K, parameter b and database with N objects.
Output: K clusters and with a minimum sum of the square of deviation.
S1: get clustering number K and clustering center according to tree structure.
S2: in the process of the kth iteration, use u_{ij} to update the clustering center c_i of each category,

$$c = \frac{\sum_{j=1}^{n}(u_{ij})^c x_j}{\sum_{j=1}^{n}(u_{ij})^c}, i = 1,\cdots,k \tag{9}$$

S3: for all the samples, the process will be finish if their subordination function won't change any more, if not, go back to S2.

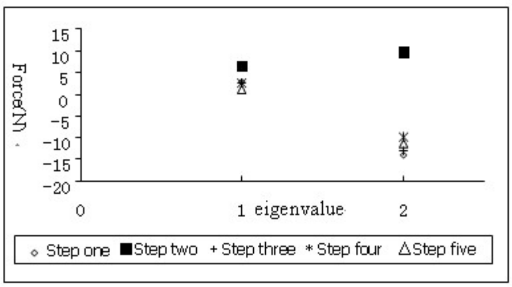


Fig. 6. FLz dimension force K- mean clustering results

This is k-means clustering results shown in Fig.6. We can get the same conclusion that the right move is first in step two and the remaining are the left first. So the two eigenvalues obtained by K-means clustering are also divisible.

Support vector machine (SVM). Support vector machine (SVM) have been developed by Vapnik (1995). SVM was developed to classify data points of linear separable data sets. The target result after finished training is separated into 2 groups which divided set by a separating hyperplane. The distance between the separated hyperplane and closest data points of dataset is called "the margin".SVM is gaining popularity due to many attractive features and promising empirical performance. It is based on Vapnik-Chervonenkis (VC) dimension of statistical learning theory and structural risk minimization principle.

SVM is evolved from the optimal separating line of a linear separable, and its basic idea is showed with a two-dimensional figure (see Fig.7). The square and hollow dot represents two types of samples, H for the classification line, H1 and H2 respectively represents the lines which are both nearest to and parallel to the classification line among the various types of categories in the sample line, the distance between them is called margin. The so-called optimal separating line is the lines which will not only be able to separate two groups (training error rate is 0) but also the largest classification margin.

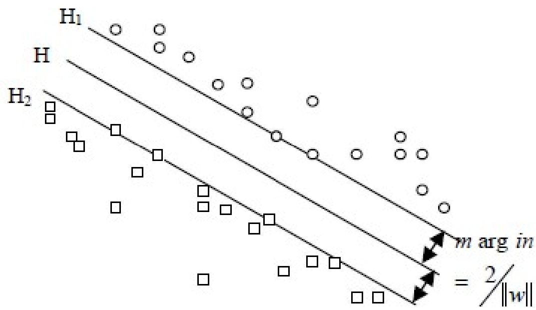


Fig. 7. Optimal separating hyperplane

3 Results and Discussion

Firstly, we tried to classify the lower extremity movement patterns of a single individual. With the C and K-means clustering in feature extraction, the recognition rate of SVM with RBF kernel functions is shown in Table1.

Table 1. Single individual RBF SVM recognition result

Clustering / the number	C	K
1	80.00	80.13
2	82.04	81.67
3	82.04	82.04
4	75.92	82.17
5	82.92	63.96
6	76.17	55.75
7	86.29	74.21
8	96.83	96.83
9	94.88	92.79
10	91.17	86.48
The average classification rate	84.83	79.55

The results of other nuclear functions are similar. So there is no need to list them separately. From the table we can see that for most subjects the recognition rate of lateral features is above 80.00%.With C-means clustering, the average recognition rate is close to 85.00% and the max recognition rate is up to 96.83 %. With K-means clustering, the average recognition rate is nearly 80.00% and the max recognition rate is close to 96.83%.It is highly separable.

Then the 50 samples from 10 subjects were processed together with the cross-validation The linear, RBF, polynomial and sigmoid kernel function were selected to test with SVM and the K-fold of Cross-validation values was chose as 2,3,4,5 separately in order to optimize SVM structure . The C and K-means clustering were used to extract feature firstly. The results are shown in Table 2 and 3. The recognition rate is the average of K-fold of Cross-validation values.

It can be seen from the two tables that when the data were expanded to all 10 subjects, the SVM recognition rate was decreased. With C-means clustering, the SVM recognition rate of lateral features was above 80.00% while, with K-means clustering, the SVM recognition of lateral features was just between 70.00% and 80.00%. There was a highest recognition rate of 85.50% with sigmoid kernel function and 4 fold cross-validation. In fact, the kernel function effect of SVM classifier is not very significant. The recognition rate with sigmoid kernel function was very slightly higher than the other kernel functions.

Table 2. The recognition results of SVM and C-means

kernel function / k	Linear	RBF	Polynomial	Sigmoid
2	78.85	75.50	81.25	84.31
3	82.11	72.96	80.02	78.06
4	83.95	83.81	78.53	85.51
5	79.92	78.55	73.56	83.66

Table 3. The recognition results of SVM and K-means

kernel function / k	Linear	RBF	Polynomial	Sigmoid
2	79.67	76.00	75.00	68.69
3	79.66	73.41	66.30	79.78
4	77.52	72.60	72.60	61.08
5	79.13	76.55	60.55	72.20

4 Conclusions

Our experimental results showed that the HRV signal could be used with fuzzy clustering and support vector machine to effectively classify the lateral features of lower limbs movement from HRV signals during FES-assisted walking. It may provide a new choice for FES control signal. In conclusion, this study provides a solid theoretical and experimental foundation for future work, such as the designing of more advanced functional electrical stimulation system in the future. The optimizing of the algorism parameters can be introduced to get better control in the future.

Acknowledgment. This research was partially supported by the National Natural Science Foundation of China (No.90920015, No.30970875, No.60501005), Joint Project of National Natural Science Foundation of China - Royal Society of Edinburgh of UK (No.30910494/C1009), The National High Technology Research & Development Program of China (No.2007AA04Z236), Biomedical Engineering Key Program (No.07ZCKFSF01300) and Intentional Cooperation Key Program (No.08ZCGHHZ00300) of Tianjin Science Technology Support Plan.

References

1. Sisto, S.A., Forrest, G.F., Faghri, P.D.: Technology for Mobility and Quality of Life in Spinal Cord Injury, Analyzing A Series of Options Available. IEEE Engineering in Medicine and Biology, 56–68 (2008)
2. Popovic, M.R., Keller, T., Pappas, I.P.I., Dietz, V., Morari, M.: Surface-stimulation technology for grasping and walking neuroprostheses. IEEE Eng. Med. Biol. Mag. 20, 82–93 (2001)
3. Agarwal, S., Kobetic, R., Nandurkar, S., Marsolais, E.B.: Functional electrical stimulation for walking in paraplegia: 17-year follow-up of 2 cases. J. Spinal Cord Med. 26, 86–91 (2003)
4. Donaldson, N., Yu, C.H.: A strategy used by paraplegic patients to stand up using FES. IEEE Trans. Rehabil. Eng. 6, 162–167 (1998)
5. Trivedi, M.M., Bezdek, J.C.: Low-Level Segmentation of Aerial Images with Fuzzy Clustering. IEEE Transactions on Systems, Man and Cybernetics 16, 589–598 (1986)
6. Hathaway, R.J., Bezdek, J.C.: Fuzzy c-means clustering of incomplete data. IEEE Transactions on Systems, Man and Cybernetics, Part B 31, 735–744 (2001)
7. Hathaway, R.J., Bezdek, J.C., Yingkang, H.: Generalized fuzzy c-means clustering strategies using L_p norm distances. IEEE Transactions on Fuzzy Systems 8, 576–582 (2000)
8. Pal, N.R., Pal, K., Bezdek, J.C.: A mixed cmeans clustering model. In: 6th IEEE International Conference on Fuzzy Systems, 1997 (1997)
9. Pal, N.R., Pal, K., Keller, J.M., Bezdek, J.C.: A new hybrid c-means clustering model. In: IEEE International Conference on Fuzzy Systems, 2004 (2004)
10. Bezdek, J.C., Hathaway, R.J.: Progressive sampling schemes for approximate clustering in very large data sets. In: IEEE International Conference on Fuzzy Systems, 2004 (2004)
11. Bezdek, J.C., Tsao, E.C.K., Pal, N.R.: Fuzzy Kohonen clustering networks. In: IEEE International Conference on Fuzzy Systems, 1992 (1992)
12. Dunn, J.C.: A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. J. Cybern. 3, 32–57 (1973)
13. Ye, P.: Fuzzy K-means algorithms based on membership function improvement. Changchun Institute of Technology(Natural Sciences Edition) (2007)

Part V

Universal Access to Education and Learning

From "Reading" Math to "Doing" Math: A New Direction in Non-visual Math Accessibility

Nancy Alajarmeh, Enrico Pontelli, and Tran Son

Department of Computer Science, New Mexico State University, New Mexico, USA
{nalajarm,epontell,tson}@cs.nmsu.edu

Abstract. The ability to understand, apply, and manipulate mathematical concepts is a cornerstone of any scientific discipline; as such it is an irreplaceable component of the training and education of any students. While the advent of online and e-learning technologies has enabled the breaking of several barriers and promoted wider access to educational opportunities, it has also furthered the disenfranchisement of visually impaired students. Despite the several efforts in the field of enhancing accessibility for those with visual impairments especially in the educational discipline, it is obvious that there is still a lack of contributions in making mathematics manipulation processes accessible. This paper presents a framework that facilitates doing and manipulating mathematical algebraic content in a way that is convenient, accessible, and usable as well for students with visual impairments.

Keywords: Accessibility, Math manipulation, Visual impairments.

1 Introduction

Mathematics is a key conceptual framework that is essential for everybody, even for those who are not actively pursuing educational objectives. The learning of mathematics commences in the very early grades of school, which is an evidence of the significance of this science. The level and difficulty of mathematics learning goes up sharply when advancing from one grade to another in the educational system; and that is clear in contents like algebra and geometry [10]. Proficiency in mathematics is a requirement to be enrolled into a variety of scientific majors, and consequently, it is a gateway to find a good place in the world of many critical professions, i.e. scientific, academic, technological, industrial, etc.

Mathematical notations rely on a predominantly visual language, which creates serious challenges for individuals with any type of visual impairments. Mathematics in nature is a complex knowledge that challenges learners in all its levels: basic, intermediate, and advanced. Complexity of mathematics spans also to involve the way it is represented and encoded. The complexity of mathematical notation comes from its two-dimensionality, spatial nature, non linearity [11], and being context sensitive. What adds to its complexity and makes it more challenging for both sighted and unsighted individuals, [10] mathematics is a language that highly

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 501–510, 2011.

necessitates exactness, definiteness, totality and comprehensibility of presentation. To summarize:

- Mathematical formulae are inherently two-dimensional entities, and the multidimensionality is critical to interpret the meaning of the formulae e.g.

$$\frac{\sqrt{b^2 - 4ac}}{2a} \tag{1}$$

- The interpretation of mathematical content is highly *contextual* - thus focusing the presentation on individual formulae could lead to ambiguous or incorrect interpretations; for example, a formula like $g(x + y)$ could be alternatively interpreted as the product between g and $x+y$ or as the application of the function g to $x + y$.

All of these issues lead to the conclusion that visually impaired individuals are passively affected by barriers that limit their chances from getting fair advantage of education. To better estimate the size of the problem, according to the Braille Institute [3], about 5 million or one out of 20 preschool-aged children and around 12.1 million children ages 6-17 have visual impairments. 80% of what students learn in school is visually presented information [13].

In mathematical problem solving process, the visually impaired students are challenged in three ways: (1) how to access the information that addresses the mathematical problem in hand; (2) the act of mapping information extracted from the given problem to the proper representation style; (3) how to provide the arrived answer in a readable way for both sighted and unsighted individuals [2,11]. Reading and writing mathematical expressions are among the most challenges visually impaired individuals encounter. The literature has highlighted the difficulties associated with access to mathematical content by individuals with visual impairments [14]. Several researchers have identified the problem of math accessibility for individuals with visual disabilities as a bottleneck in the access to training and careers in science, technology and mathematics (STEM) [12].

An extensive literature has been developed to overcome barriers and problems associated to non-visual accessibility of mathematics. The literature has provided a plethora of approaches aimed at promoting accessibility of mathematical formulae (see [14] for a recent survey). However, there have been relatively limited efforts to enhance *mathematics manipulation technologies* for people who are visually impaired. Focusing exclusively on the reading of mathematical formulae - which has represented the main focus of the current literature - provides only limited help to visually impaired students. In particular:

- The focus on formulae detracts from the opportunity of enhancing understanding by placing the formulae in their context and using the context to develop more effective presentation modalities;
- The greatest challenge for most students is not on reading mathematics, but on understanding and using/applying its concepts.

The process of learning mathematical content is inseparable from the ability to practice it - by solving exercises or by manually developing the sequence of logical or algebraic steps that are often implicit in the arguments present in the studied text. The literature has, to date, provided very limited efforts aimed at enabling visually impaired students to writing and practicing mathematics (e.g., [19]).

Let us consider a simple example: we wish to simplify (a + b) (2a - b):

1. The first step is to develop the product, by multiplying each term in the parenthetical expression by each term in the second parenthetical expression; specifically,

   (a) Extract the first term from the first expression, a, and multiply this by each term in the second expression: $2a^2 - ab$

   (b) Extract the second term from the first expression, b, and repeat the process: $2ab - b^2$

   (c) Combine the two expressions: $2a^2 - ab + 2ab - b^2$

2. The second step is to identify occurrences of terms that can be combined and replace them with their combination: $2a^2 + ab - b^2$.

This example illustrates a process that is challenging without any visual aid - selecting sub-expressions, scanning expressions, create copies of expressions. A purely textual presentation of expressions would require repeated scans and jumps.

In this project, we propose a framework to facilitate manipulation of mathematical concepts by individuals with visual disabilities, starting at the level of basic algebra. The framework is to enable visually impaired individuals to have a complete control over navigation functions through a full history of manipulation steps they carry out. In particular, the framework allows the visually impaired students studying middle school level algebra to manipulate algebraic expressions in recorded accessible consecutive steps that at the end lead them to a final answer to the initial problem. The framework also enables visually impaired students to access each component in any step, review their manipulated expressions, correct mistakes done at any point of time in case of having any, work on particular components separately, save interesting results for later usage, and use a scratch paper like environment.

The central concept of the proposed framework is the idea of workspace:

- The workspace is composed of blank pages, where the student can conduct his/her exercises; pages can be named and directly accessed;
- The workspace provides the ability to place arbitrary bookmarks, within distinct formulae, and position navigation on any requested bookmark (i.e., the student can keep their “fingers" pointing to different locations of the formulae); bookmarks are hierarchically organized to facilitate retrieval of different positions within exercises and formulae;
- The workspace allows the student to select formulae (represented by roots of MathML sub-trees), copy them, or delete them; each formula can be manipulated through selection of sub-formulae, insertion of new terms or removal of terms;
- The workspace provides distinct domain-specific layers that combine the basic term manipulations (selection, copying, removal, replacement) with domain-specific transformations (e.g., multiplication of terms, reversal of fractions);
- The workspace makes it possible to temporarily hide parts of formulae.

The design of workspaces relies on aural presentation of formulae and bookmarks, and the use of hot access keys to control scanning and manipulation of terms. The implementation relies on the representation of expressions in MathML (or its realization in the MSOffice OOXML format); workspaces are implemented as a web tool that provides a virtual notebook with the described features.

2 Prior Research

Among the commonly used approaches in accessibility, screen readers, e.g., JAWS [9] and Microsoft narrator, are considerable software tools that help visually impaired individuals to have an access to many on screen presented content, i.e., text and image or video descriptions (alternative text). However, this is not truly applicable for all kinds of content, especially the mathematical one, for the earlier mentioned characteristics of mathematics, in addition to the fact that some mathematical content is represented as images. Software magnifying tools on the other hand, are very useful aids in case of low vision, while unfortunately, they are not of any advantage in a situation of complete blindness. The limitations of these approaches strongly encouraged the development of other tools.

The field of promoting mathematics accessibility for the visually impaired individuals has attracted researchers and developers that dramatically changed the way of handling mathematics. Research projects influenced the gradual development of other approaches. As a result of introducing MathML, a W3C standard [12], mathematics became available online in a convenient format that can be neatly and easily rendered. However, availability is not everything towards usability, in a sense that many user categories like those of visual impairments were not able to make use of mathematics and take advantage of the availability of mathematical content - even after introducing MathML. Static approaches that use Braille to render mathematics, and Dynamic approaches in which audio is used in rendering are the main categories of accessibility approaches [14].

Math2Braille for converting MathML to Braille standard [4], LAMBDA which uses linear mathematical encoding for mathematics representation [6], and MAVIS to overcome the problem of backtranslation between Braille and LaTeX [12] are examples of the static approaches in which several new Braille notations were created for mathematics. Examples of dynamic presentation approaches include AsTeR for reading mathematics through parsing LaTeX documents to produce a tree structure which in turn facilitates navigation and tagging [15], MathGenie for reading and navigating equations, also for conversions between different formats i.e., Nemeth to LaTeX [18], and the Microsoft Internet Explorer plug-in MathPlayer for reading mathematical content that is represented in MathML [5, 17]. InftyEditor [20], Infty converters for conversion between different standards [20], and WinTriangle for reading and editing documents [8] are examples of tools that enable the non-visual creation and editing of mathematical content. A different direction is represented by tools to interactively translate MathML into voiceXML, render pages by audio, and facilitate voice interaction from end users [16]. [1] proposed a framework to allow both the linear and hierarchical navigation of mathematical expressions.

3 The New Framework Development

Recent technological advances changed the way students receive education, moving from complete dependency on classroom and materialistic tools to more of virtual communication mediums and software based curriculums. For educational purposes, visually impaired students are increasingly required to deal with online available mathematical content. Visually impaired students, while navigating online materials, might be exposed to mathematical content, which has to be properly detected. Once detected, mathematical expressions should be rendered separately from the original navigated page. Next comes the need to provide the ability to reason about mathematical content and perform algebraic manipulations. The visually impaired student should be able to manipulate and reach to an answer to the original algebraic expression. Through all these stages, accessibility is an issue that must be present for the visually impaired. Fig. 2 shows the overall structure of our system.

The ultimate goal of this new framework is to enable visually impaired students to access online mathematical content and to perform algebraic manipulations for solving algebraic problems. Unlike LiveMath and other online creating and editing tools, this new direction of educational tools is not meant to directly or indirectly give answers or carry out manipulations for end-users; instead, it allows visually impaired individuals to practice algebra and solving techniques without any guidance and results given from the system. Moreover, it keeps a history of steps of all manipulations carried out so far by the visually impaired students, so that they can easily go back and forth to review what has been so far made, or simply saying, navigation between different components in the hierarchy of steps is made possible without tedious efforts from the end-user side, which facilitates building a full and clear image of what they have been doing.

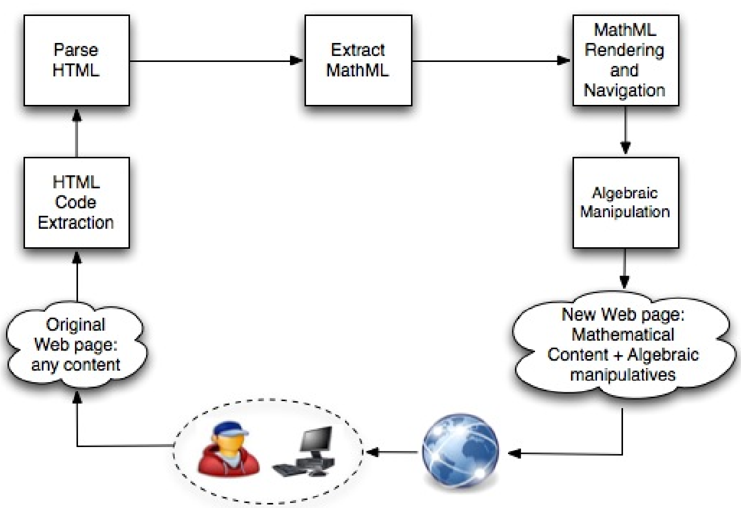


Fig. 1. The Framework structure

The easy to access history of manipulations gives the visually impaired students the capability to make sure what they are making, how so far they handled the expression under consecutive manipulation steps they made, and what expression they got at the end of each step. This approach reduces the amount of visually impaired individual's dependency on sighted people help, decreases the amount of time the visually impaired students would invest in manipulation when using other traditional tools, lowers the need to memorize temporary results, and facilitates practicing of mathematics skills.

4 Functional Requirements and Design

To come up with an effective solution, the following functional requirements were taken into consideration:

- Accessibility support
- It entails understandable representation, clear rendering, user friendly interface, and easily manageable functionalities. The use of audio rendering and hot access keys are among the features the framework has.
- Self controlled environment
- The visually impaired student controls what functions to use, either right or wrong like pencil and paper.
- Hierarchical recording of manipulation steps
- Every step carried out should be recorded for later revision and making sure that things are going in the right direction.
- Error correction and backtracking
- The ability to backtrack, correct errors, and cancel mistaken steps. This helps avoid redoing things from the very beginning.
- Easy navigation
- Going back and forth between steps should be an easy mission that visually impaired students should not bother.
- Bookmarking and working on certain sub-expressions
- Bookmarking prevents the recreation of interesting sub-expressions, also helps reducing the amount of information to be memorized. Working on certain sub-expressions helps flexible dealing with smaller parts that collectively form the original expression.
- The ability to save work
- The work can be saved for later revisions, study, and training. Teachers also can benefit from this option so as they see how their students practice algebra.

The basic operations one needs in solving an equation are taken into consideration. No solution hint or guidance will be given either implicitly or explicitly, since the approach acts the same as giving a piece of paper and pencil for a student and asking them to solve an equation. Non-visual accessibility is supported through the deployment of audio feedback, hot access keys, and interaction with the MathPlayer plug-in [5,17].

The system includes a plug-in that works on parsing MathML source code to capture any mathematical content on the navigated page. After the mathematical content detection step, and with a control from the visually impaired student, a new space opens to enable the visually impaired student to manipulate mathematical expressions freely. Moreover, TTS and audio feedback approaches are used to guide the visually impaired student to the content of the page. In addition, hot access keys will be deployed for the ease of access to help visually impaired in carrying out manipulations. Mouse and keyboard navigations are supported, and for later, voice commands and haptic device navigation and control might be supported.

Initially, the following operations (distinguished in bold) which are mainly derived from the mathematical algebraic perspective, are considered:

- Eliminate - To delete a term (get rid of it when no longer needed) from the whole expression being manipulated. It is necessary to mention that when talking about a term; it can be a factor, variable, number, operator, etc.
- Combine like terms - To collect terms that fall in same category.
- Cancel - To cancel terms together when the effect of one cancels the effect of the other.
- Enclose - To make a set of consecutive terms look like one unit.
- Add - To add an existent term to the expression. This is clear and straightforward in algebraic manipulations.
- Subtract - Like the way the previous function works, this function is to subtract an existent term from the expression.
- Add new - To add a nonexistent term to the expression.
- Subtract new - To subtract a nonexistent term from the expression.
- Insert - To insert a new term, this operation is commonly used in the context of manipulation although the terminology is not formal in mathematical definition of operations for manipulating mathematics.
- Replace - To replace an existent term by a new one, like what said about the previous function, common but not formal.
- Multiply - To multiply the expression by an existent term. This is clear and straightforward in algebraic manipulations.
- Divide by - To divide the expression by an existent term. This is clear and straightforward in algebraic manipulations.
- Factor out - To handle the process of extracting factors.
- Expand - To carry out the distribution process in case of multiplications to what is inside brackets.
- Apply roots - To apply any degree root to the selected term.
- Apply powers - To apply any power to the selected term.
- Work on selected term - To separately manipulate the selected term away of the whole original expression, facilitating more manageable manipulations.
- Open a scratch paper - To work on an empty space to carry out intermediate operations and manipulations.
- Open Calculator - To open an accessible basic calculator that helps doing calculations.

- Save for later - To keep interesting components or results for later use, in other words to bookmark desired components.
- Undo - To go back one step from the currently under focus step, to correct mistakes made in any step prior to the active one.

5 Future Work

It has been found that, while there exists extensive work contributed in accessibility, availability, and usability of mathematics for people who are visually impaired, there is still an obvious absence of attention paid for promoting and driving mathematics manipulation accessibility for students with visual impairments. The contribution presented in this paper is considered a new direction and a starting point for an extendable research work that aims at handling the problem of lacking accessible mathematics manipulation frameworks to help students who are visually impaired overcome their print disabilities. Going beyond what is addressed in this framework, the next natural step is to provide coverage for a larger body of mathematical content, beyond elementary algebra, e.g., by including support for all content covered in college algebra and extend to basic geometry.

The current framework has been designed and developed to provide the basic functionalities required to support extraction of mathematical content and execution of algebraic transformations. The next step involves creation of layers that implement mathematical transformations that are relevant to different mathematical content areas. The design of these layers should be informed by the modalities of mathematical manipulation used by visually impaired students. We have initiated a series of interviews and focus groups with visually impaired students from the New Mexico School for the Blind and Visually Impaired. These user studies will be used to develop the set of transformations and to guide the way they are presented to the users.

6 Conclusions

Visually impaired individuals are productive and are of equal importance for the societies they are involved in like sighted people. Especially in the context of mathematics, the students who are visually impaired are taught exactly the same content and curricula sighted students receive at school. This indicates the need to provide visually impaired students with tools that enable them to reason with mathematical content in a manner comparable to that of sighted students. While the existing literature has offered tools to enable the ``reading'' of mathematical content, there are no tools offering the equivalent of ``pen-and-paper'' to solve mathematical problems.

In this paper, we present the preliminary design of a framework, currently under development, for enhancing math manipulation accessibility for middle school students who are visually impaired. Unlike previous contributions, which are exclusively aimed at enhancing rendering and reading of mathematical content, this framework provides tools to support the processes of mathematical manipulation available. The framework is accessible and offers a convenient solution to allow students who are visually impaired to practice with basic algebra and perform algebraic manipulations of simple formulae. The framework currently support transformations for elementary algebra and it is being informed and validated by focus groups, in cooperation with the local school for the Blind and Visually Impaired.

Acknowledgments. The research has been partially supported by NSF grants CBET-0754525 and IIS-0812267.

References

1. Abu Doush, I., Pontelli, E.: Building a programmable architecture for non-visual navigation of mathematics: Using rules for guiding presentation and switching between modalities. In: Stephanidis, C. (ed.) UAHCI 2009. LNCS, vol. 5616, pp. 3–13. Springer, Heidelberg (2009)
2. Beal, C.R., Shaw, E.: Working memory and math problem solving by blind middle and high school students: implications for universal access. In: Proceedings of the 19th International Conference for Information Technology and Teacher Education, Las Vegas (2008)
3. Braille institute, `http://brailleinstitute.org`
4. Crombie, D., Lenoir, R., McKenzie, N.R., Barker, A.: math2braille: Opening access to Mathematics. In: Miesenberger, K., Klaus, J., Zagler, W.L., Burger, D. (eds.) ICCHP 2004. LNCS, vol. 3118, pp. 670–677. Springer, Heidelberg (2004)
5. Design science, `http://www.dessci.com`
6. Edwards, A., McCartney, H., Fogarolo, F.: Lambda: a multimodal approach to making mathematics accessible to blind students. In: Proceedings of the 8th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS 2006), Portland,Oregon, USA, October 23-25, pp. 48–54. ACM, New York (2006)
7. FireVox, `http://firevox.clcworld.net`
8. Gardner, J., Stewart, R., Francioni, J., Smith, A.: Tiger, Agc, And Win-Triangle, Removing The Barrier To Sem Education. In: CSUN International Conference on Technology and Persons with Disabilities, Los Angeles, CA (2002)
9. JAWS for Windows, Freedom Scientific, `http://www.freedomscientific.com`
10. Kahanova, I.: The ways of teaching mathematics to visually impaired students (unpublished)
11. Karshmer, A., Pontelli, E., Gupta, G.: Helping visually impaired students in the study of mathematics. In: 29th Annual Frontiers in Education Conference, FIE 1999, vol. 2, pp.12C4/5-12C410(1999)
12. Karshmer, A., Gupta, G., Geiiger, S., Weaver, C.: Reading and writing mathematics: the MAVIS project. In: Proceedings of the Third International ACM Conference on Assistive Technologies, pp. 136–143. ACM Press, Marina del Rey (1998)
13. Murphy, R.: Learning-related vision problems (March 2010), `http://www.allaboutvision.com/parents/learning.htm`
14. Pontelli, E., Gupta, G., Karshmer, A.: Mathematics Accessibility. In: Universal Access Handbook. CRC Press, Boca Raton (2009)
15. Raman, T.V.: Audio Systems for Technical Reading. Ph.D. thesis, Department of Computer Science, Cornell University, NY, USA (1994)

16. Reddy, H., Gupta, G.: Dynamic aural browsing of mathml documents with voicexml. In: Human-Computer Interaction. Lawrence Erlbaum and Associates, Mahwah (2005)
17. Soiffer, N.: Mathplayer: web-based math accessibility. In: ASSETS 2005. ACM Press, Baltimore (2005)
18. Stanley, P.B., Karshmer, A.: Translating MathML into Nemeth Braille Code. In: Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I. (eds.) ICCHP 2006. LNCS, vol. 4061, pp. 1175–1182. Springer, Heidelberg (2006)
19. Stöger, B., Miesenberger, K., Batušić, M.: Mathematical Working Environment for Blind. In: Miesenberger, K., Klaus, J., Zagler, W.L., Burger, D. (eds.) ICCHP 2004. LNCS, vol. 3118, pp. 656–663. Springer, Heidelberg (2004)
20. Suzuki, M., Tamari, F., Fukuda, R., Uchida, S., Kanahori, T.: INFTY – An integrated OCR system for mathematical documents. In: Proceedings of ACM Symposium on Document Engineering (DocEng), Grenoble, France, pp. 95–104

Accessible Education for Autistic Children: ABA-Based Didactic Software

Silvia Artoni[1], Maria Claudia Buzzi[1], Marina Buzzi[1], Claudia Fenili[2], and Simona Mencarini[1]

[1] CNR-IIT, via Moruzzi 1, 56124 Pisa, Italy
Silvia.Artoni@iit.cnr.it, Claudia.Buzzi@iit.cnr.it, Marina.Buzzi@iit.cnr.it, Simona.Mencarini@iit.cnr.it
[2] ASA (Associazione Sindromi Autistiche), Sesto Fiorentino (FI), Italy
Claudiafenili14@libero.it

Abstract. Behavior Analysis is the science that studies human behavior. Applied Behavior Analysis (ABA) is a process of systematic application of interventions whose principles are based on Learning Theory. The aim is to improve socially significant behaviors; the interventions are responsible for changing the behavior. ABA is successfully applied for educating autistic children. This paper discusses the design of an eLearning environment, based on ABA principles, for delivering accessible education to autistic children. ABA is a complex methodology and it is crucial to maintain therapy coherence between all the caregivers while allowing them maximum flexibility on the ABA application to better adapt to individual child responses.

Keywords: autism, ABA, didactic software, usability, accessibility, design.

1 Introduction

Autism Spectrum Disorder (ASD) is a pervasive development disturbance characterized by the difficulty or inability to communicate and consequent impaired social integration. The autism spectrum is very large and its symptoms vary significantly from child to child. Typical behaviors include avoiding eye contact and stereotypical behaviors such as rocking, head rotations or flapping hands. Common issues are: the loss or reduction of functional language (utters sounds that are cries or screams or echolalia), different perception (extreme sensibility to noisy environment). Some cases also present additional cognitive or perception deficits (Comorbidity).

The Diagnostic and Statistical Manual of Mental Disorders, *DSM-IV* TR [1] and the international statistic Classification of Mental and Behavioral Disorders [2] define autism in term of delays or anomalies that appear within 3 years of age, in at least one of following areas: 1) social interaction, 2) communicative and social use of language, 3) symbolic or imaginative play.

This amplitude of problems and needs requires a personalized educational approach to best exploit the child's abilities. Conventional educational techniques are ineffective in autistic children for several reasons, including their different sensorial perception (for instance, difficulty remaining in a noisy environment such as a

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 511–520, 2011.

classroom), difficulty concentrating, inability to communicate needs or to accomplish tasks (provoking frustration and "problem behaviors").

According to several studies [3], [4], early intervention in children affected by autism disorder is more effective for learning and developing social abilities. It is very important to act quickly in the first two years of life to take advantage of the plasticity of children's brains [5] [6], [7]. Often autism symptoms appear at 18-36 months, and some alarming signals may be observed very early, from birth. Early individuation of autism is certainly an advantage: it is crucial to have an early diagnosis to quickly start behavior therapy [8], [9], [10], [11], [12].

Several studies have shown that early and intensive intervention, beginning in preschool with a load of at least 6-8 h a day, is the only way to produce significant improvement in knowledge and skills, in some cases (high-functional children) with development corresponding to the normotype [13], [12], [14], [15]. Lovaas published the first study (1973) on effectiveness of intensive behavioral intervention in children with autism. The results of this study led the author to develop the ABA model, validated by subsequent research in 1987 [11]. Eikeseth and others showed that ABA treatment, compared to an eclectic type of intervention, produced the most progress in all areas of evolutionary [16]. Sallow and Graupner have recently replicated the studies of the father of ABA, showing that approximately half the children subjected to early and intensive behavioral intervention have reached the age of 7 years with a level of adaptive functioning and intellectual equal to the standard [17]. In the state of Maine Administrators of Services for Children with Disabilities regarded ABA as the only scientifically validated intervention, and concluded that "early diagnosis of autism should be accompanied by the proven effectiveness of an intensive ABA intervention to produce optimal results and benefits of long-term costs" [18]. In Italy the SINPIA guidelines (May 2005) approved ABA as an intervention for teaching autistic children. Thus, early diagnosis of autism is essential to implementing behavioral intervention of this kind. Results of this first study led the author to develop a model of ABA intervention, validated by subsequent investigation in 1987 [11]. Several other approaches are in large part based on methods used in ABA, but most authoritative results in the scientific environment are related to ABA itself [16].

More than 30 years of research have created a highly effective model of intervention through the systematic and intensive application of basic behavioral principles and structured teaching techniques, based on Discrete Trial Teaching (DTT) [10], [19]. ABA therapy requires caregivers to implement precise educational protocols (error-less learning, reinforcement, etc.) and rigorous verification of the child's learning progress (by filling out structured forms), and has proved to be particularly effective [14], [20], [19].

In short, by conducting a series of sequential trials of increasing difficulty, the child, initially driven by the therapist's prompts to avoid errors, will gradually be able to operate independently and assimilate concepts/skills. The success of the child's therapy depends on the coherence of the program, which should be respected not only at school and during the therapy sessions, but also at home with parents and relatives. The family's participation is fundamental, and parents' active participation in the ABA program is strongly recommended. Lovaas was the first to suggest that parents play an important therapeutic role in the treatment of children with ASD [6], [11].

In this paper we describe the participatory design (also involving therapists and children's parents) of a didactic software program for children with autism spectrum disorder. The ABCD SW project (Autistic Behavior & Computer-based Didactic software) [21] aims at designing and developing Open Source didactic software conforming to the ABA learning method to ensure an usable product that responds to these children's special needs.

2 Related Work

ABA therapy is based on Augmentative and Alternative Communication (AAC) and Discrete Trial Training (DTT). AAC is a technique that by increasing the user's perceptions provides an alternative method for communicating, and thus is used in treating learning disabilities and neurological pathologies. Creating new technologies for autistic subjects specifically concerns implementing communication through AAC and the behavior analysis ("problem behaviors" and social interactions [22], [23]).

Recent studies have confirmed the efficacy of electronic therapy based on AAC. Hirano et al. [24] designed and implemented a visual scheduling system (vSked) for planning and organizing children's activities in the classroom, and observed not only increased efficiency for caregivers but also benefits and improvements in student-student and student-teacher communication.

Recently, Pino and Kouroupetroglou have created an Open Source framework for building adaptable, modular, multilingual, cheap and sustainable component-based AAC products [25]. However at the moment the framework is distributed only by the authors.

Mobile communication tools are a promising field in AAC research. Moving AAC from specialized devices to a standard mobile platform offers many advantages: first of all drastic cost reduction, greater flexibility, simpler and faster customization, small size and ubiquity, and familiar environment (cell phone) for the children. Monibi and Hayes implemented a library of virtual cards for autistic children's activities on a Nokia N800 (Mocoto prototype). The pre-installed card library may be easily extended with pictures or other digital images. A desktop software program allows setup and customization of activities (e.g. size and number of cards, audio cues) [26].

Sampath et al. propose a system for autism using AAC that allows bidirectional communication between child and caregivers. Specifically, a gateway on a handheld device was built, allowing conversion between pictures and spoken language, enabling the completion of the communication loop (receptive and expressive) [27].

Although many digital products are available for augmentative communication (e.g., GoTalk, Tango, Dynavox, Activity Pad), teachers and therapists have experienced low usability and flexibility; training is required for set-up and customization, making it difficult for parents to use it at home. Furthermore, they are expensive [22].

Putnam and Chong investigate the real use at home of software specifically designed for autism. Results of an on-line survey (spread out throughout associations) filled out by 114 responders shed light on the limited diffusion of specific software (only 8%) while software for cognitive disability was used by 25% [28]. This result indicates the need to move from research prototypes to engineered software specifically designed for autistic children, possibly according to their preferences.

Participatory design actively involves all stakeholders in the design process to help ensure that a created product meets their needs and that results are highly usable [29]. Participatory design is essential in designing for autistic persons. Hirano et al. successful applied participatory design in developing the vSked system [24]. De Leo and Leroy involved special education teachers in designing software for facilitating communication with severely autistic children via smart phones [30].

A large branch of research is devoted to providing usable tools to assist therapists of children with autism. Kientz, Monibi and Hayes based their studies and their software development on participatory design [22], [26]. Kientz et al. [31] designed and developed two systems for facilitating efficient child monitoring (both progress and behavior): 1) Abaris, the supporting team executing Discrete Trial Training therapy, building indices into videos of therapy sessions and allowing easy data search; 2) CareLog, for collecting and analyzing behavioral data ("problem behaviors"). Furthermore sensors were used to monitor stimming behavior (self-stimulatory movements) to understand the cause of an uncomfortable situation.

Hailpern et al. [32] investigated the use of computers for assessing the behavior of nonverbal children. By defining a set of dependent variables for use in video annotation, called A3, it is possible to systematically analyze the interactions of nonverbal children, with the computer capturing feedback related to attention, engagement and vocal behavior.

Recently, pervasive technologies are investigated for monitoring user behavior. To enhance social skills and abilities of persons with ASD, Kaliouby and Goodwin built a suite of wearable technologies (cameras, microphones, sensors) for capturing, analyzing, and sharing (via wireless network) their social-emotional interactions, in a fun and engaging way [23]. Current studies offer usable technologies that favor communication and analysis of data of behaviors. There is a lack of specific studies for the systematic application of instructions regarding ABA, which prioritizes learning through the DTT system by increasing degrees of difficulty.

3 ABA Fundamentals

The first step for participatory design was to acquire knowledge of the ABA method. We interviewed the therapists, watched videos of therapy sessions, and we were also present at ABA sessions with different children and therapists. This was possible only with a few selected children who were able to work in spite of an external presence. The observer stands behind the child at a certain distance and does not interfere in any way with the teaching session. ABA therapy is conducted in a simple environment that offers no distractions for the child. A simple empty table is used for arranging the article on acquisition (to be learnt) and the additional materials (such as distracters) needed for the trial. The teacher sits facing the child, the table between them, and prepares the child's work environment and then announces the command. The child executes the directive either independently or in response to the therapist's prompt. A reinforcement may be given after a sequence of trials such as a game, or something else particularly pleasing.

The teacher decides on which program to work:

1. **matching:** image to image, image to word, word to image, word to word
2. **receptive** program (e.g. using the command: "Touch apple")
3. **expressive** program (e.g. using the command: "What is this?").
   and consequently, (s)he presents the child with the material (usually hand-made colored square cards), organized in categories. In addition to cards, real elements such as plastic objects, dishes, fruit or other elements may be used.
   In ABA each program is implemented by using a sequence of trials of increasing degrees of difficulty:
4. **Mass Trial:** basic trials ensuring children's success with the item on acquisition in a personalized program
5. **Distracter phase:** first a neutral distracter is added to the item on acquisition, and next two neutral ones. Then one non-neutral distracter is added, and next, two non-neutral ones.
6. **Extended Trials:** a choice involving 3 items, 2 must be mastered and the third is on acquisition
7. **Random Rotations** of learned items. If the program is just at the beginning and 3 items have not been mastered, it is possible to work on a rotation of the 2 items mastered in that program (Block Rotation).
8. **Generalization:** consists in proposing new ways of acquiring each mastered item. After acquisition, articles are subjected to an abstraction process.

Within this sequence of learning, different types of prompts may be provided, according to the learning needs of the subject, to help him complete the trial successfully.

4 The System

We believe that technology can enhance lives of children with autism, by creating more sophisticated eLearning tools. Our project mainly aims to define an educational ad hoc methodology for teaching autistic children and create didactic computer-based courses in order to render therapy more effective and efficient. The idea is to map ABA principles in creating a specific software program suitable for therapists and children. To simplify child-computer interaction and allow a modality "similar" to physical ABA therapy, we chose to use touch-screen devices and vocal synthesis to announce the commands of the learning modules (exercises).

The software must enable great flexibility and be adaptable to the learning abilities of any autistic child. For this reason we decided to offer an adaptation of the presentation modality of the activities to be carried out, first of all fitting into the topography of the prompt. For example: a child who needs more visual support can be facilitated by the introduction of the written discriminative stimulus (SD) and not only the vocal, pronounced by the speech synthesizer. Caregivers may also speak in order to integrate commands in the best way to stimulate the child. The language required is generally simple and minimal (short sentences without articles, e.g. touch apple, match yellow). It may take place in more structured sentences depending on the level of difficulty reached by the child and its ability to generalize the concept expressed.

Figure 1 shows a logical scheme of the software architecture of the eLearning environment. The three main software components are:

- Didactic software, i.e. basic ABA programs, for learning categories of articles (such as colors, numbers, etc.). To this aim an intuitive User Interface (UI) was studied, in order to be as clear as possible, easy to use for the less ABA-experienced individuals.
- Monitoring software, as required by the ABA therapy, to control the child's progress. Computer-based therapy is only one medium for promoting learning but classic face-to-face one-to-one ABA therapy will be alternated. The monitoring software will allow recording data from both face-to-face and computer-child therapy.
- Data analysis software allows conversion of raw data into easy-to-use graphics, showing the child's learning progress. Web interfaces allow caregivers to easily and rapidly access child learning data, facilitating decisions on how to best direct the educational program and render the learning process more effective.

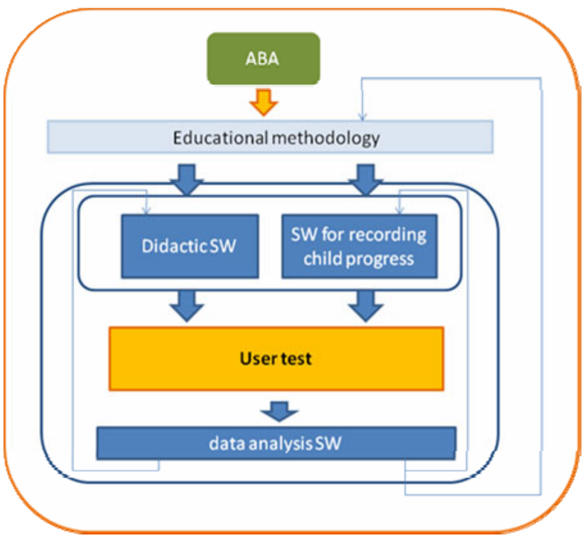


Fig. 1. Scheme of the ABA eLearning Environment for child with autism

Our project covers all these aspects as a whole and not as separate components. After developing an initial prototype, we will perform a pilot test with a few low-functioning autistic children to perform software tuning and customization.

A crucial aspect is to customize the UI. Each child has personal abilities that vary over a wide range, so the software must be easily configured to meet both specific (receptive communication, expressive communication, etc.) and temporary needs (if the child fails a trial on a mastered item, the therapist would immediately jump to a lower level [see Section 3] to reiterate and consolidate the acquisition).

Data from each trial are stored in a database in order to create a path of acquisitions (mastered items), difficulties and prompt adaptation to reach the objective. The historical archive is useful for ensuring greater consistency and coherence between the caregivers in their daily work, and make the subject's "learning history" available for the scientific analysis of gathered data.

5 The User Interfaces

Participatory design with therapists and caregivers is essential for reaching reliable results. In order to avoid influencing the imagination of therapists we asked them to picture how they image the user interface of the software, without showing them our personal ideas. Figure 2 shows three examples of organization.

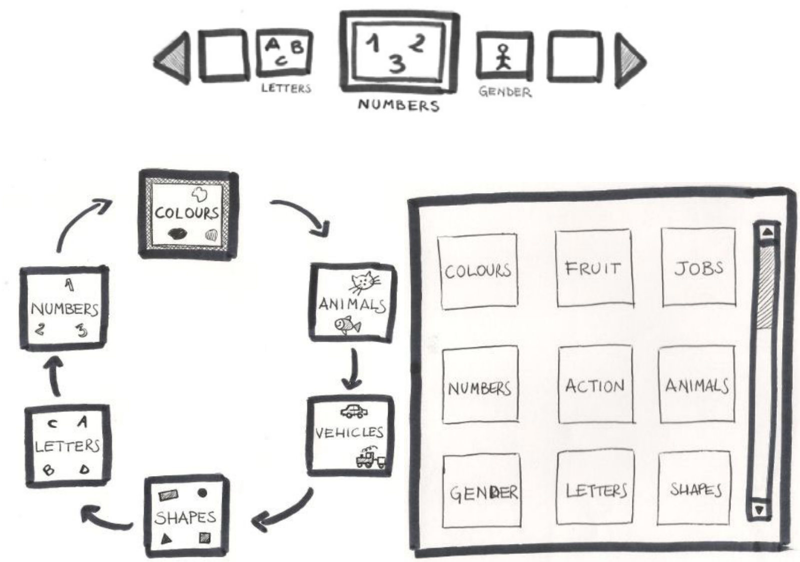


Fig. 2. The UIs proposed by therapists: carousel (up), circular icons (bottom left), square icons (bottom right)

Starting from the top one, there is a carousel (for selecting the category on which work), a circle of icons, and icons (arranged in a square) with a search function.

As previously mentioned, the complexity of Discrete Trial Training and teaching by levels increases the complexity of the UI (Fig.3). Another crucial point regards the organization of data related to the teaching session. This information concerns the child, but it is also linked to therapists and programs.

Each performance of each elementary unit is recorded in a database. Data then needs to be aggregated and transformed in graphic form to highlight the progress or weakness of the child and allow directing the therapy as needed.

The generalization applicable to the software is obviously limited compared to what happens during a typical ABA session (without computer). Within the software modules the generalization is carried out changing:

- the Discriminative Stimulus (i.e. the therapist's command)
- the position of items on the screen
- the stimulus, the visual features of items (photo, drawing, outline, sketch).

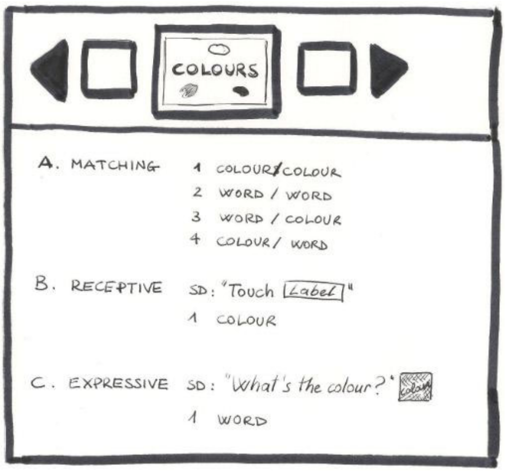


Fig. 3. Programs (matching, receptive or expressive) for the selected category (color)

6 Conclusion

In this paper we have described the design of didactic software dedicated to autistic children, involving the Applied Behavior Analysis. This method was selected since it has scientific validation, is based on structured teaching and is customized.

Numerous studies demonstrate the effectiveness of ABA. The teaching is structured by DTT, follows the "errorless learning" procedure, based on the use of prompts, and adapts the educational program to the child's abilities. The software design of the proposed eLearning environment closely follows the procedure for teaching ABA, suggesting Discrete Trials through levels of increasing difficulty.

The software ensures a high degree of flexibility thanks to its ability to customize the type of prompt to the learning needs of each student. It relies on a consolidated scientific approach since it is based on an effective and validated method for teaching children with disorders of the autistic spectrum. It also provides accurate data collection for monitoring the individual learning path, essential for focusing the educational effort where necessary for the child.

Currently, the system design has been completed, and the user interfaces were drawn with the valuable support of children therapists and parents. The next step will be implementing the didactic software component, which is the core of the eLearning system.

Acknowledgments. We thank the Regione Toscana that funded this project within the framework of the "FAS 2007 2013 Delibera CIPE 166/2007 PAR FAS Regione Toscana Action Line 1.1.a.3" (Gen 2011-Gen 2013). Furthermore, we thank all therapists who participated in this user interface design, in particular Simone Di Cesare and Francesca Lenci.

References

1. American Psychiatric Association. DSM-IV-TR. Diagnostic and Statistical Manual of Mental Disorders. 4 edn. Text Revision (2000)
2. World Health Organization, The ICD-10. Classification of Mental and Behavioural Disorders. World Health Organization, Geneva (1992)
3. Anderson, S.R., Romanczyk, R.G.: Early Intervention for Young Children with Autism: Continuum-Based Behavioral Models. The Journal of The Association for Persons with Severe Handicaps 24(3), 162–173 (1999)
4. Corsello, C.M.: Early intervention in autism. Infants & Young Children 18(2), 74–85 (2005)
5. Myers, S.M., Johnson, C.P., and the Council on Children With Disabilities: Management of Children With Autism Spectrum Disorders. Pediatrics 120, 1162–1182 (2007), doi:10.1542/peds.2007-2362
6. Maurice, C., Green, G., Luce, S.C.: Behavioral intervention for young children with autism: A manual for parents and professionals (1996)
7. Dawson, G., Zanolli, K.: Early intervention and brain plasticity in autism. In: Novartis Found Symp., vol. 251, pp. 266–74 (2003); discussion 274–80, 281–97
8. Baron-Cohen, S., Scott, F.J., Allison, C., Williams, J., Bolton, P., Matthews, F.E., Brayne, C.: Prevalence of autism-spectrum conditions: UK school-based population study. The British Journal of Psychiatry 194, 500–509 (2009)
9. Leaf, R., McEachin, J.: A work in progress: Behaviour Management Strategies and a Curriculum for Intensive Behavioral Treatment of Autism, pp. 1417–1418 (1999)
10. Lovaas, O.I.: Teaching developmentally disable children: The me book (1981)
11. Lovaas, O.I.: Behavioral treatment and normal education and intellectual functionating in young autistic children. Journal of Consulting and Clinical Psychology 55, 3–9 (1987)
12. Rogers, S.J.: Brief Report: Early Intervention in Autism. Journal of Autism and Developmental Disorders 26(2) (1996)
13. Fenske, E.C., Zalenski, S., Krantz, P.J., McClannahan, L.E.: Age at intervention and treatment outcome for autistic children in a comprehensive intervention program. Analysis and Intervention in Developmental Disabilities 5, 49–58 (1985)
14. Rosenwasser, B., Axelrod, S.: The contribution of applied behavior analysis to the education of people with autism. Behavior Modification 25(5), 671–677 (2001)

15. Savner, J.L., Myles, B.S.: Making Visual Supports Work in the Home and Community: Strategies for Individuals with Autism and Asperger Syndrome. Autism and Asperger Publishing Co., Shawnee Mission (2000)
16. Eikeseth, S.: Outcome of comprehensive psycho-educational interventions for young children with autism (2008)
17. Sallow, G., Graupner, T.: Intensive behavioural Treatment for Children with Autism: Four year outcome and Predictors. American Journal on Mental Retardation 110(6), 417–438 (2005)
18. Maine Administrators of Services for Children with Disabilities, Autism Task Force Report (2000), http://www.madsec.org/
19. Weiss, M.J.: Expanding ABA intervention in intensive programs for children with autism: The inclusion of natural environment training and fluency based instruction. The Behavioural Analyst 2(3), 182–186 (2001)
20. Rosenwasser, B., Axelrod, S.: More contributions of applied behavior analysis to the education of people with autism. Behavior Modification 26(1), 3–8 (2002)
21. Artoni, S., Buzzi, M.C., Buzzi, M., Fenili, C.: Didactic software for autistic children. In: Cipolla Ficarra, F.V., de Castro Lozano, C., Pérez Jiménez, M., Nicol, E., Kratky, A., Cipolla-Ficarra, M. (eds.) ADNTIIC 2010. LNCS, vol. 6616, pp. 73–80. Springer, Heidelberg (2011)
22. Hayes, G.R., Hirano, S., Marcu, G., Monibi, M., Nguyen, D.H., Yeganyan, M.: Interactive visual supports for children with autism. Springer Personal and Ubiquitous Computing, 18 (2010), doi: 10.1007/s00779-010-0294-8
23. Kaliouby, R., Goodwin, M.S.: iSET: Interactive Social-Emotional Toolkit for Autism Spectrum Disorder. In: Proceedings of IDC 2008, pp. 77–80 (2008)
24. Hirano, S., Yeganyan, M.T., Marcu, G., Nguyen, D.H., Boyd, L.A., Hayes, G.R.: vSked: Evaluation of a System to Support Classroom Activities for Children with Autism. In: Proceedings of CHI, pp. 1633–1642 (2010)
25. Pino, A., Kouroupetroglou, G.: ITHACA: An Open Source Framework for Building Component-Based Augmentative and Alternative Communication Applications. ACM Transactions on Accessible Computing (TACCESS) 2(4), art. no.: 14, 30 pp. (2010)
26. Monibi, M., Hayes, G.R.: Mocotos: Mobile Communications Tools for Children with Special Needs. In: Proceedings of IDC 2008, pp. 121–124 (2008)
27. Sampath, H., Sivaswamy, J., Indurkhya, B.: Assistive systems for children with dyslexia and autism. ACM SIGACCESS Accessibility and Computing (96) (2010)
28. Putnam, C., Chong, L.: Software and technologies designed for people with autism: what do users want? In: Proceedings of the 10th ACM SIGACCESS Conference, pp. 3–10 (2008)
29. Wikipedia, Participatory Design, http://en.wikipedia.org/wiki/Participatory_design
30. De Leo, G., Leroy, G.: Smartphones to facilitate communication and improve social skills of children with severe autism spectrum disorder: special education teachers as proxies. In: Proceedings of the 7th Int. Conf. on Interaction Design and Children (IDC 2008), pp. 45–48 (2008)
31. Kientz, J.A., Hayes, G.R., Westeyn, T.L., Starner, T., Abowd, G.D.: Pervasive Computing and Autism: Assisting Caregivers of Children with Special Needs. IEEE Pervasive Computing 6(1), 28–35 (2007)
32. Hailpern, J., Karahalios, K., Halle, J., Dethorne, L., Coletto, M.: A3: HCI Coding Guideline for Research Using Video Annotation to Assess Behavior of Nonverbal Subjects with Computer-Based Intervention. ACM Trans. Access. Comput. 2(2), Article 8, 29 pages (2009)

Educational Impact of Structured Podcasts on Blind Users

Maria Claudia Buzzi[1], Marina Buzzi[1], Barbara Leporini[2], and Giulio Mori[1]

[1] CNR-IIT, via Moruzzi 1, 56124 Pisa, Italy
{Claudia.Buzzi,Marina.Buzzi,Giulio.Mori}@iit.cnr.it
[2] CNR-ISTI, via Moruzzi 1, 56124 Pisa, Italy
Barbara.Leporini@isti.cnr.it

Abstract. Podcasts are increasingly present in education. However, sequential and continuous listening to didactic material might not be the most appropriate form for learning. For instance, if not well-organized, it can be frustrating to retrieve specific information from educational podcasts. We believe that structured podcasts (i.e., audio files that exploit the document structure) could facilitate navigation and search of content. In this paper we investigate the impact of structured podcasts on blind users, through a user test and satisfaction questionnaire administered to eighteen visually-impaired persons. Results confirm that structured podcasts are simple to use, rapid to explore and are greatly appreciated by blind users.

Keywords: audio podcast, mp3, learning, blind, screen reader.

1 Introduction

Audio podcasting is increasingly used in education. Especially appreciated as an ubiquitous tool, podcasting is a flexible technology that adapts to users' needs and habits and facilitates learning anywhere, anytime, at any pace [1]. Numerous studies have shown the effectiveness of podcasting for many students who learn better when listening to educational material instead of accessing the written version [2], [3], [4], [5]. When podcasting is used for educational purposes, easy-to-use content facilitates the learning process. Sequential reading of all content may not be the best way to promote learning, while exploration and the opportunity to skip quickly to a specific part might be more appropriate for memorizing the main concepts and building and consolidating ideas. In our approach we consider features that are useful for providing audio content in educational settings. We propose to offer structured audio materials that permit students to easily obtain an overview of the content as well as find specific parts quickly.

Previous studies have shown that short podcasts (e.g., max 10-15 min) are more effective for learning than a single long unit. Long podcasts may decrease attention, reducing comprehension [6], [4]. Establishing rules and methods for achieving an appropriate content structure is a challenge.

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 521–529, 2011.

A text-to-speech (TTS) system is software that converts text into speech [7]. Available text-to-audio converters (such as Text2mp3, DSpeech) usually produce a single audio podcast or a set of time-based mp3 files. Adding “break strings” to the document also makes it possible to split the content into several audio files, but this requires considerable time and effort by the user.

In [8] we proposed a Web-based Text-To-Speech system for automatic generation of a set of structured audio podcasts from a single source document. The system takes a document as input, and generates as output a set of podcasts that reflects the content structure (one mp3 file for each document section), ready to be downloaded and uploaded on an mp3 player.

Listening to structured audio content offers many advantages over sequential listening: the user can easily locate audio parts of interest, and better orient him/herself when searching for specific information within the podcast, aiding the learning process.

In this paper we investigate the educational impact on blind users of listening to structured podcasts, versus a time-based division method. We have defined a set of tasks to perform on a collection of podcasts and a questionnaire for visually impaired users to fill out. We discuss and analyze results collected through this designed and implemented Web user test. After discussing related work, we briefly present the Web-based Text-To-Speech (TTS) system [8] for automatic generation of a set of structured audio podcasts, a description of the user test, analysis of collected data, and some concluding remarks along with indications for future work.

2 Related Works

Many studies confirm that podcasting is increasingly useful for exchanging information and making learning approaches more efficient in terms of resource consumption. For many people, listening may be more attractive and less tedious than reading [9, 10]. Podcasts are often used for additional support in teaching and learning. Some common uses in higher education are: taped lectures, guest speakers, group presentations, tutorials, exam reviews, reinforcement of key concepts, and drill or repetition [10].

Recently, to facilitate the preparation of audio materials, several tools have been proposed and developed that transform a text document into a spoken version by using text-to-speech (TTS) technology and a voice synthesizer.

A TTS system converts a text into speech. The output generated can be heard immediately while the audio is being produced or can automatically be recorded in audio files. Typically, these audio files are in mp3 format and can be listened to on an mp3 portable player or on smartphones. Tools like Robobraille and vozMe are examples of this converting process. RoboBraille is an email-based service which automates translation of text documents into Braille and speech. Users submit documents (e.g., text files, Word documents, HTML pages) as email attachments. The translated results are returned to the user via email [12]. VozMe (http://vozme.com/) is an easy-to-use online service for creating mp3s from text. It only requires typing in or pasting the text and pressing a button. However, audio content created is not structured (is a single file) so it is not very suitable for studying.

Some software such as Natural Reader [13] allows one to jump to the previous or next section. In [9] authors combine sophisticated speech-processing techniques (including audio-based skimming) to create a multimedia player that allows audio-centered browsing and navigation of recorded presentations. However, these systems require interaction with a PC and thus are not suitable for Mobile Learning.

3 Structured Audio Podcast System

Structured audio podcasts are useful for all kinds of people and are especially appreciated by blind users, who can easily skip from one section to the previous or the next (on an mp3 player by using backward and forward command keys).

Our hypothesis is that for blind users, structured audio content facilitates navigation within the content, reducing frustration when searching for the desired content when learning and memorizing specific parts. Conversely, searching for the desired content in a single long sequential audio file can be quite difficult.

The system we developed is web-based [8]: blind users can interact (via screen reader) with the Web interfaces thanks to the accessible features obtained by applying appropriate design criteria of the Web Content Accessible Guidelines (WCAG 2.0) [14] as well as WAI-ARIA (Accessible Rich Internet Applications) [15] attributes. Our Web system receives a document in input (.doc, .rtf, .txt), and in output provides a set of audio files that reflects the document's internal structure (one file for each document section).

Figure 1 shows the architecture of the system: a document is uploaded through a Web page and saved on the server. The system identifies the section titles so that the document can be split and processed by a TTS module to be transformed in an audio streaming, coded and saved as an mp3 file. Each generated audio podcast represents a section of the document, and it contains the title of the section at the beginning in the first few seconds of listening.

In addition, the system transforms each title of sections to audio streaming i.e. talk files (http://www.rockbox.org/). With .talk files, an mp3 player equipped with the Rockbox firmware makes files and folders directly accessible through audio.

Analysis of font properties allows detection of important information highlighted by means of bold or italic font properties by the author, emphasizing them in the generated audio podcasts (for example words in bold are delivered between two short tones in the audio podcast).

The system is also able to detect tables that are converted into single podcasts, since listening to a large table without seeing it can be extremely difficult and frustrating.

After generating structured podcasts, the system shows their list on a web user interface. A list of downloadable podcasts is a sort of index of document sections. Users can choose whether to download the entire package on their mp3 players, or select only a part: this enables better orientation during a quick search for specific information within a podcast, improving the learning process.

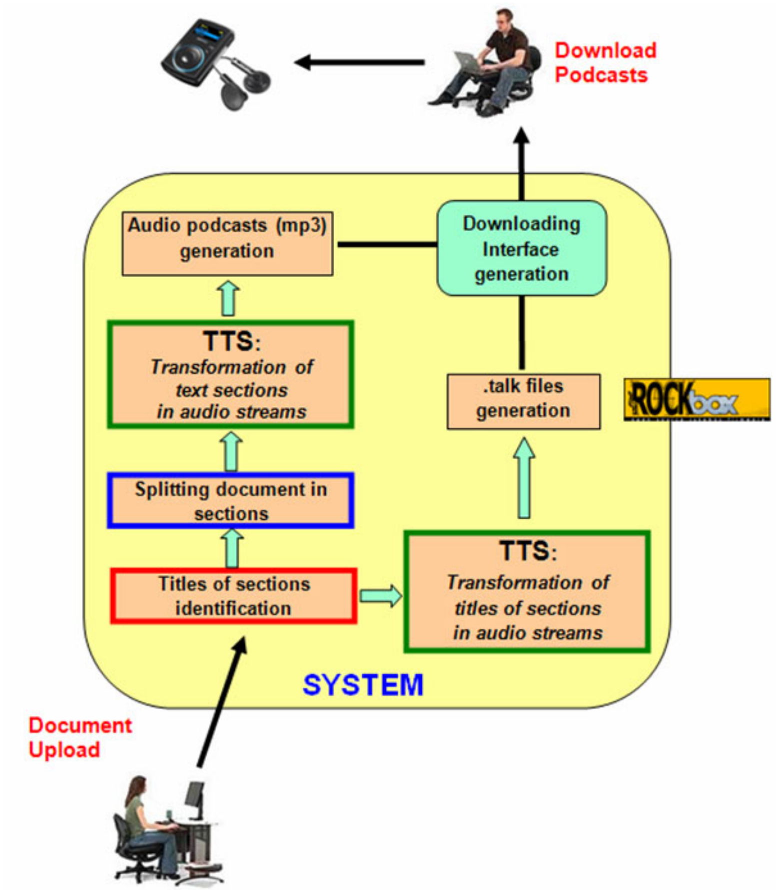


Fig. 1. Architecture of the web-based system for generating structured audio podcasts

4 User Test Design

This study aims to evaluate the real impact of structured podcasts on learning and on the "going over" and memorizing phase for blind users. The memorizing phase is critical for blind users, especially for quickly retrieving specific information. To evaluate our proposal, we set up a user test involving a group of blind users.

4.1 Target

To involve potential users in our test, a message was sent to a collection of email addresses obtained thanks to collaboration with the Italian Association for the Blind.

Eighteen people answered; 14 are blind and 4 are visually impaired. The sample was composed of 8 women and 10 men, with age ranging from 18 to 74 years. All participants use assistive technology in the Windows environment: the blind use

JAWS for Windows[1] screen reader, and the visually impaired use both the screen reader (to rest their eyes) and the magnifier (to enlarge screen content).

Regarding technical skill, 12 users have an intermediate degree of knowledge of the Internet, while 3 users were beginners and 3 users were experts. All of them used an mp3 player device for listening to audio books, music, educational audio files, or downloaded audio podcasts from Internet. In particular, 14 participants used an mp3 player supporting Rockbox firmware for vocal navigation on the device.

4.2 Method

A 12-page document on an e-Commerce topic was created, containing useful information for sightless users. The document entitled “Tips for on-line shopping” was organized into three parts: Psychology of e-commerce web sites, Online Security and Payment methods.

Three audio versions of the document were created to be used in three different phases:

1. A single podcast of duration 28 min and 24 s (first phase).
2. A set of 5-min podcasts (6 mp3 files) where the document broke off after 5 min (when the paragraph ended)
3. A set of structured podcasts (21 mp3 audio files), each one representing a logical section of the document, created using our system [8].

The set of 5-min podcasts was generated using Dspeech (http://dimio.altervista.org/eng/), and the set of structured podcasts was created using our system (described in the previous section). Both sets of podcasts were generated using the same SAPI 5 voice, in order to avoid some differences in user perception during listening. Each set of podcasts was grouped in a directory: time-organization (containing 6 files) and content-division (with 21 file audio).

The two sets of podcasts also included a playlist .m3u file, useful for listening to the podcasts using a computer, because the user can skip from one audio track to the next or to the previous one as on an mp3 player, in a sequence order as listed in the .m3u file.

Both folders are compressed in a self-extracting file directly downloadable from the user test pages. The folder with the files can be uploaded on an mp3 player device (or listened to on a computer). The audio files are necessary for performing a set of tasks, which mainly consist of answering questions (carrying out seeking tasks).

The user test is organized in three phases. The first phase, for all users, requires choosing any username (e.g. a nickname), to login, and to download the podcast corresponding to the entire document to be listened to.

The second phase requires downloading the **same** document split in different tracks according to the time-based division method (for group A) or the structure-based division method (for group B), which the participant had to use to carry out the first three tasks. The tasks consist in answering three questions, seeking the answer in the folder containing the set of downloaded podcasts (different versions for the two groups).

[1] http://www.freedomscientific.com/products/fs/jaws-product-page.asp

Analogously, the third phase requires downloading the **same** document split into different tracks according to the structure-based (Group A) or the time-based (Group B) algorithm, in order to make each group listen to both versions, and to execute three other simple seeking tasks in order to answer three questions.

The sequence of time and structure-based tasks (and vice versa) are alternately balanced between participants in order to minimize the learning effect of the structured audio contents. Each group has six questions – i.e. tasks – to complete: three for a set of time-based podcasts, three for the structured ones. The user is alternatively assigned to the group A or B at the first login.

Each of the six questions has the objective of finding the answer within one specific podcast of the set, and the questions do not follow the same sequential order of sections of the original document. This is to force the user to spend time navigating and retrieving the required information.

4.3 Remote User Test

The user test, accessible remotely via Web, has been developed using PHP. Web pages of the user test conform the Web Content Accessibility Guidelines (WCAG 2.0) [14] and they have been developed using WAI-ARIA suite [15] criteria to facilitate navigation and interaction for blind and visually impaired users using a screen reader or other assistive technology.

The Web pages of the user test were fully accessible with the most recent version of Firefox while not completely with Internet Explorer. The remote system leads each user to perform the three phases. For each task the starting and ending time was captured and recorded in order to obtain information on how long each task took to find the information. Results evaluate the correctness of the answers to the six questions (three for a set of time-based podcasts, and three for the structured ones), the time for answering, and the answers of the post-test questionnaire.

The time it took to answer any of the six questions was calculated from the time the user uploads the page containing the question until the “Go to the next question” button has been pressed; if the user decided to complete the task at a different time (s)he pressed the “Pause the test” push button and restarted the test later at the point where (s)he left off. Analysis of the test results is described in Section 5.

4.4 Questionnaire

At the end of the three phases, each user was asked to fill out a questionnaire on personal information (sex, age), computer skills, conditions of test executions (browser, screen reader, etc.), user preferences and a subjective evaluation of both the division methods.

5 Results

As previously mentioned (Section 4), the sequence of time- and structure-based tasks (and vice versa) was alternately balanced between participants. In fact, a pool of users had to look for the answer to a question within the time-based set of audio podcasts, while another pool of users had to look for the same question inside the structure-based set of audio podcasts. We considered task accomplishment and time required for the completed tasks (i.e., for the correct answers).

By analyzing the log files, we observed that not all users completed all the tasks successfully. Probably due to the length of the test, some gave an answer after trying to recall information from memory, resulting in a wrong or approximate answer. If the answer was incorrect we discarded this time.

The total number of correct answers (successful seeking tasks) in the two versions was higher in the organization structured-based (93%) than the time-based (89%). Furthermore, it took 24% less time to complete all six tasks (total average time) using structured rather time-based organization.

Considering the total average time it took to successfully complete each task, we observed that using the set of structured podcasts led to shorter times for four of the tasks (T1, T2, T4 and T6), compared to using the time-based set, and vice versa for the other two tasks.

Possible factors favoring time-based organization are:

1. the position of the information sought is at the end of the track in the structured podcast, while it is at the beginning in the time-based file;
2. the title did not indicate clearly which track to explore. In these cases the exploration tasks were nearly equivalent.

Instead, where the title clearly suggested where the answer could be found, structure-based organization proved more useful than the time-based one.

However, a possible factor contributing to the time differences (and some users confirm this hypothesis) might be that some users answered a question by memory, after the first phase (listening to the full mp3 version) -- thus executing the task without actually searching for the answer.

To summarize, after analyzing all times spent carrying out the assigned tasks, we can conclude that there is a time savings of about one quarter (24%) by using the structured-based version. This result is important, considering that when the user reads again (i.e. listens over and over to) a podcast he/she is able to identify easily and immediately where to find information, thanks to structured organization using the logical structure of the document, and it would be easy to remember.

Concerning user satisfaction and preferences (expressed in the post-test questionnaire), the feedback of participants is clear, concordant and unambiguous: all participants prefer listening to small audio podcasts instead of a single long audio file, and they prefer structured podcasts to the set composed of 5-min podcasts, for many reasons:

a) it takes less time to retrieve specific information
b) titles of sections simplify seeking specific information
c) shorter average duration of each podcast
d) more intuitive and practical search
e) focus and perception of more details
f) simplifies memories and information access
g) a structure-based division method maintains the entire audio flow of a section logically, while the time-based division method can interrupt it abruptly.

All this positive feedback encourages the adoption of the structure-based method to produce more usable podcasts that might better satisfy user needs.

6 Conclusions

In this paper we present a study to verify the impact of structured and time-based podcasts on learning for blind users. The study was conducted with 18 users (14 blind and 4 visually impaired) via a user test to compare the two division methods: structured (according to document structure) and time-based (fixed duration) versions. Analyzing the collected data, all users greatly appreciated having a set of podcasts instead of a single long file, and they expressed a preference for structure-based podcasts over the time-based (5-min) division method.

Considering the time it took to complete the proposed tasks, structured podcasts seem to support the search for information, decreasing the time needed to search (at least in 4 questions out of 6). Results highlighted a time savings of about 24% obtained using the structured-based sequence.

Results regarding time were not totally in accord with our expectations for greater time-saving, possibly due to several factors: external factors may have influenced the test, increasing the overall time (due to remote testing that did not allow precise monitoring of user activities); difficulty and position of the answer inside the content; and the test was long so some users may have preferred to answer by trying to remember what they heard before, without actually searching for the answers in the podcasts.

All participants expressed their preference for structured podcasts since they required less time to retrieve specific information, titles of document sections provided a clue for seeking specific information, there was shorter average duration of each podcast, and overall they offered a more intuitive and practical search. In addition, reflecting the logical structure of the document, a structured podcast allows the focus and perception of more details, simplifies memories and information access, and maintains the entire audio flow of a section logically, while the time-based division method can interrupt it abruptly.

In conclusion, the positive feedback of all participants encourages the adoption of a structure-based method to produce more usable podcasts that better satisfy user needs by decreasing the time required to explore and seek information.

In future work we aim to refine the user test and repeat it with more blind people, but also with sighted users in order to compare results of the two samples. Moreover, we are planning a new user test with blind users, aiming to evaluate the web system we have developed to create structured podcasts, to verify its usability, the impact of structured podcasts on learning, and the delivery of special information (such as bold, italic, tables, etc.) contained in the original document.

References

1. Mermelstein, B., Tal, E.: Using Cellular Phones in Higher Education. In: Wireless and Mobile Technologies in Education, WMTE 2005 (2005)
2. Aldrich, D., Bell, B., Batzel, T.: Automated Podcasting Solution Expands the Boundaries of the Classroom. In: Proceedings of the 34th Annual ACM SIGUCCS Conference on User Services, pp. 1–4 (2006)
3. Lauer, T., Wolfgang, H.: Audio-based Methods for Navigating and Browsing Educational Multimedia Documents. In: Proceedings of the Inter. Workshop on Ed. Mult. and Mult. ed. Poster and demo session, pp. 123–124 (2007), doi, `http://portal.acm.org/citation.cfm?id=1290166`

4. Pitt, I., Edwards, A.: Design of Speech Based Devices. A Practical Guide. Springer, London (2003)
5. Wolff, T.B.: Podcasting Made Simple. In: Proceedings of the 34th Annual ACM SIGUCCS Conference on User Services, pp. 413–418 (2006)
6. Deibel, K.: Course experiences of computing students with disabilities: four case studies. In: Proceedings of the 39th SIGCSE Technical Symposium on Computer Science Education, pp. 454–458 (2008)
7. Allen, J., Sharon Hunnicutt, M., Klatt, D.: From Text to Speech: The MITalk system. Cambridge University Press, Cambridge (1987) ISBN 0521306418
8. Leporini, B., Buzzi, M.C., Buzzi, M., Mori, G.: Automatically Structuring Text for Audio Learning. In: Stephanidis, C. (ed.) UAHCI 2009. LNCS, vol. 5616, pp. 73–82. Springer, Heidelberg (2009)
9. Cebeci, Z., Tekdal, M.: Using Podcast as Audio Learning Objects. Interd. Journal of Knowledge and Learning Objects 2 (2006)
10. Ormond, P.R.: Podcasting enhances learning. Journal of Computing Science in Learning 24(1), 232–238 (2008)
11. Campbell, G.: There's Something in the Air: Podcasting in Education. Educause Review 40(6) (November/December 2005), `http://connect.educause.edu/Library/EDUCAUSE+Review/TheresSomethingintheAiPo/40587`
12. The RoboBraille Consortium. The RoboBraille email service, `http://www.robobraille.org`
13. NaturalSoft Text-to-Speech, `http://www.naturalreaders.com/index.htm`
14. W3C. Web Content Accessibility Guidelines 2.0. (December 11, 2008), `http://www.w3.org/TR/WCAG20/`
15. W3C. WAI-ARIA Technical Specification, `http://www.w3.org/TR/wai-aria/`

A New Structure of Online Learning Environment to Support the Professional Learning

Wenzhi Chen and Yu-How Lin

Departmrnt of Industrial Design, Chang Gung University,
33302 Tao-Yuan, Taiwan, R.O.C.
wenzhi@mail.cgu.edu.tw, uricul@pchome.com.tw

Abstract. The online learning environment (OLE) is an essential element in using information and communication technology (ICT) for learning and instruction. This paper proposed a new structure of an OLE for design professional domain that tried to integrate the learning community (forum) and courseware for increasing the using motivation and learning efficiency. First, a framework consisted with four components - Instructor, Learner, Course, and Internet, was formulated. Then, the OLE was developed based on the framework that constructed by 3 levels - Community, Classroom, and Studio. Community provides an open space for users to interact and exchange the information. Classroom provides various design courses, while Studio provides several communication tools for design studio activities. The experiment courses and surveys were conducted to evaluate the need-ness and satisfaction of the OLE. The results show the positive responses from the learners that regarded the Community can help their learning.

Keywords: e-learning, online community, online learning environment, social network.

1 Introduction

The rapid development of the Information and Communication Technology (ICT) make the e-learning become a main trend of the education. There are many online learning environments (OLE) and learning management systems (LMS) were developed to support the students' learning and the teaching of the instructors. The OLE/LMS is a platform that provides functions to help the learning and teaching [1]. Some researchers argued that the most of OLE/LMS consider the management of the learning content, activities, and process, but do not focus on the learning [2].

The learner is the central of the learning system, especially in e-learning related theories [3]. In the real world, the learners may need to integrate some knowledge, or to share some opinions that related with several courses, especially in professional domain. But courses are isolated that has its own space and function in the most of OLE/LMS.

This paper proposed a new structure of an online learning environment served on design professional domain that set a public forum on the top level of the OLE,

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 530–537, 2011.

provide a space to formulate a community for users. It provides the chances for learners to discuss the issues across courses or connect with people or experts outside the campus.

2 Framework and Prototype

2.1 The Framework for Development of Online Design Learning Environment

The OLE/LMS is an essential element in using ICT for learning and instruction [4]. Many frameworks and models were proposed for developing or evaluating the OLE/LMS in general e-learning domain [5-12]. Based on the features of knowledge and educational needs of the field of design. The OLE for design education would be different from that of other domain. A framework for developing the online design learning environment (ODLE) was formulated according the related literature and the experiences of using Internet to mediate the design education in the previous study [4]. Figure 1 illustrates the framework that consisted with four components: *Instructor*, *Learner*, *Course*, and *Internet*. Each component has three key factors that are considered along with the relationship between component and ODLE.

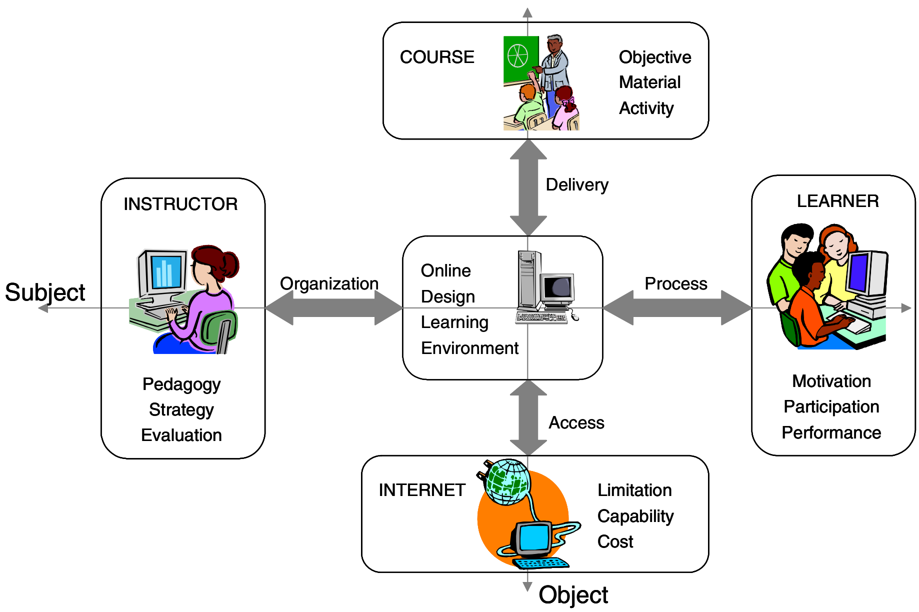


Fig. 1. The framework of online design learning environment [4]

From this framework, the ODLE is the center of the whole structure through which all the components are connected to. At the horizontal ends of the ODLE structure are the *Instructor* and the *Learner*, who are also the main users of OLE. *Instructor* transfer knowledge and skills to *learner* through the ODLE while *learner* obtain the relevant knowledge and skills through this ODLE medium. The *Course* and *Internet* are then at the vertical ends of the ODLE structure where the *course* contains the main content for the transfer between *instructor* and *student*, and the *Internet* provides the main channel for linking the components to the ODLE. Similarly, 4 main functional components of the ODLE provide instructors with tools for courses and learner managements, establish a learning objective for learners based on their own or the plan and procedures set by the instructors, provide the courses content and activities with appropriate communication by using Internet as a link, while also providing more online resources to support teaching and learning. The relationship between these four components and ODLE itself are *organization*, *process*, *delivery*, and *access*. The *instructors* need good interface and *organization* functions to manage the learners and course materials. The *learners* need flexible design learning *process* make learning more easier. A suitable *delivery* methods is important to convey the *courses* materials and activities. *Internet* is the key channel to *access* all components in ODLE. The components, factors and variables of the ODLE framework shows as Table 1.

Table 1. Components, factors and variables of ODLE framework

Components	Factors	Variables
Course	Objectives	Single, compound …
	Materials	Text, graphic, audio, video, interactive multimedia …
	Activities	Lecture, discussion, presentation, practice…
Instructor	Strategy	Project, workshop, cooperative project …
	Pedagogy	Cooperative, constructivism …
	Evaluation	Personal, group …
Learner	Motivation	Degree of the intention …
	Participant ion	Present, log times, duration …
	Performance	Quality, creative, innovation, finish …
Internet	Cost	Money, time, infrastructure …
	Limitation	Bandwidth, stability …
	Capability	Bandwidth, information …

2.2 The Prototype of Online Design Learning Environment

The ODLE prototype “CoCreaThink Design (CCTD)” (http://thinkdesign.cgu.edu.tw) was built based on the framework mention above. The structure of CCTD consisted with three levels: *Community*, *Classroom* and *Studio*. Figure 2 demonstrates the main aims, functions, the interactive type and the media used in each level. Figure 3 shows the screen shot of each level. The details of each level are described below.

The first level, *Community*, provides the CCTD users with an open space to exchange opinions about design. This is an open forum for everyone and provides a space for both learners and instructors to interact and communicate with individuals from inside and/or outside the formal course. Based on a collaborative learning theory, the aim is to cultivate the ability of presentation and communication, and to increase the interchange of design thinking and motivations for learning.

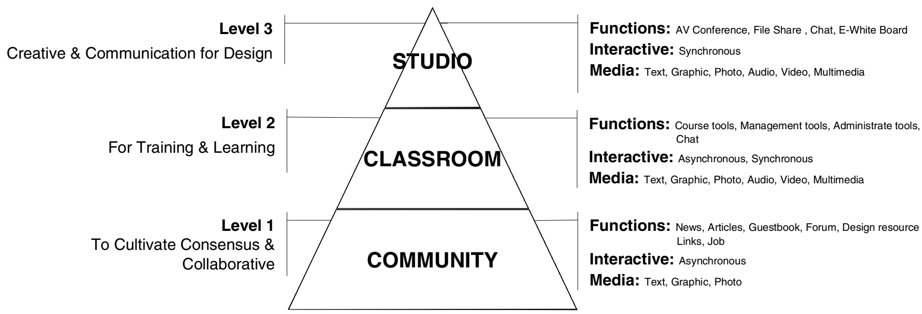


Fig. 2. The levels, functions and interactive media types of the "CoCreaThink Design"

Fig. 3. The screen shot of the "CoCreaThink Design" (Community, Classroom, Studio)

The second level, *Classroom*, provides various asynchronous design courses. This is the main space for learning, and has several tools provided for course learning and management. This level provides an asynchronous learning approach for knowledge and technique-based courses. File format is not limited and both learners and instructors can upload/download course materials or assignments, allowing them to share, to submit and to exchange course resources.

The third level, *Studio*, provides synchronous tools for communication and discussion in design studio activities. This is designed for the design studio course. chat room, video conferencing, file exchange and electric white board have been integrated into this environment, allowing the learner to do design collaboratively.

3 Evaluation

For evaluating the CCTD and verify the effective of new structure of OLE in design professional learning. Two experiment courses and projects were performed, and the surveys were conducted to collect the responses of students. There were 94 students from two universities participated the study. The participants of the experiment courses are showed as Table 2.

Table 2. The number of the participants in experiment course

University\Surveys	Survey 1 (Course)	Survey 2 (Co-design)
U1	25	17
U2	36	16
Total	61	33

3.1 Experiment Course and Project

Two experiments were conducted to collect the data. The first experiment used CCTD to mediated the learning and teaching in studio based courses. The second experiment applied CCTD for co-design projects. Two collaborating universities were participated these experiment course and projects.

Third-year undergraduate students participated in these studies. The students participated enrolled in the course at their own respective schools. Nevertheless, the courses had the same title, "Product Design". Product design was the core course of the undergraduate industrial design education program in both cases. It was a one-year course divided into two semesters.

The course aims to cultivate the following abilities of students: 1) professional attitude, 2) design practice techniques and skills, and 3) application of design methods and theories. Studio (project-based) pedagogy was adopted, in which students learn through the processes of design research, ideation, simulation, and presentation to develop their own way of design, as well as learning how to define a problem, create ideas, and select the solution for the problem. Finally, through multimedia presentations, prototype making and display, and reports, students learn to present their learning outcomes. These courses using the CCTD as a mediator to assist learning and teaching. The courses was based on the blended learning approach [13, 14] which combines personal interaction in live class sessions with online education for greater flexibility. In addition to the traditional face-to-face interaction in a studio, the teachers also used the CCTD to announce course information and to hand out class notes; the students could use it to submit assignment files and to share and exchange design concepts and ideas. The CCTD also provided a variety of tools for communication, including an asynchronous discussion forum and a synchronous chat room and video conference facility; students and teachers could use these for various design activities. The detail and the relevant results of these experiments could referred to previous studies [15-17].

3.2 Data Collection and Analysis

A questionnaire designed to collect the data of the need-ness and satisfaction of the CCTD. The questions of need-ness and satisfaction were focus on the functions that design for different level of the CCTD. All the questions use the Likert 7 scales for responded that the range was from "strongly not need (-3)" to "strongly need (+3)" in need-ness category, and from "vary unsatisfied (-3)" to "very satisfied (+3)" in satisfaction category. The analysis used the SPSS statistical software package.

Table 3. Internal Consistency Analysis of the questionnaire

Categories	Sub-Categories	Cronbach's α	
		Survey 1	Survey 2
Need-ness	Community	0.82	0.89
	Classroom	0.92	0.89
	Studio	0.84	0.91
Satisfaction	Community	0.85	0.91
	Classroom	0.92	0.92
	Studio	0.91	0.82

The reliability as a measure of internal consistency analysis was calculated, the result is showed in Table 3. All the reliability of the different questions were higher than 0.80.

4 Results

Table 4 presents the results of descriptive statistics and analysis of variance. The need-ness of *Community* (mean = 1.28, SD = 0.76) is higher than *Classroom* (mean = 1.07, SD = 0.89) and *Studio* (mean = 0.87, SD = 1.10). The satisfaction of *Community* (mean = 0.86, SD = 0.77) is also higher than *Classroom* (mean = 0.61, SD = 0.78) and *Studio* (mean = 0.37, SD = 1.07). The ANOVA indicates a significant difference of need-ness (F = 4.528, p = 0.012 < 0.05), and a significant difference of satisfaction (F = 6.225, p = 0.002 < 0.01). The Student-Newman-Keuls (SNK) test shows that the significant difference was between *Community* and *Studio* both in need-ness and satisfaction.

Table 4. The analysis of the need-ness and satisfaction

Category	Level	N	Mean	SD	ANOVA		
					F	P	SNK
Need-ness	Community	89	1.28	0.76	4.528	0.012*	A
	Classroom	93	1.07	0.89			A B
	Studio	93	0.87	1.10			B
Satisfaction	Community	84	0.86	0.77	6.225	0.002**	A
	Classroom	80	0.61	0.78			A B
	Studio	82	0.37	1.07			B

Note: Need-ness scale: -3 (strongly not need) ~+3 (strongly need)
Satisfaction scale: -3 (vary unsatisfied) ~ +3 (very satisfied)
* p < 0.05, ** p < 0.01

The results demonstrated that learners regarded the public forum can help their learning. The collaborative learning is one of the important theories of e-learning. But most of the collaboration is inside the course with classmate in general courses. There are many people can help learner to solve the problems. Dalsgaard [2] argued that learning management systems do not support a social constructivist approach which emphasizes self-governed learning activities of students. They only cover administrative issues. It is argued that students' self-governed learning processes are supported by providing students with personal tools and engaging them in different kinds of social networks. The results of this study are consistent with Dalsgaard's arguments.

5 Conclusions

The online learning environment is an essential element of using information and communication technology for learning and instruction. But there are many factors affect the learning and instruction in e-learning. Online learning environment should not just emphases the management of the learning objects and activities. It also needs to consider the specific needs of learning to help learners. This paper proposed a new structure of the online learning environment for design domain that try to integrate the learning community and courseware for increasing the using motivation and learning efficiency of OLE. The results show the positive responses from the learners. It also encourages instructors to participate the discussion in learning community, not just doing the administration on OLE.

Social networks provide various advantages in educational context as collaboration, knowledge sharing, common interests, active participation and reflective thinking [18]. This kind of structure that integrated social software into OLE seems beneficial to students. But, more extensive research would be necessary to be know about the way of using ICT in the design profession learning.

Acknowledgments. This research was partly funded by the National Science Council of Republic of China under grant numbers NSC 96-2411-H-182-003, NSC 92-2520-S-182-002, and NSC 90-2218-E-182-004.

References

1. Dringus, L.P., Terrell, S.: The framework for DIRECTED online learning environments. The Internet and Higher Education 2(1), 55–67 (1999)
2. Dalsgaard, C.: Social software: E-learning beyond learning management systems. European Journal of Open, Distance and E-Learning (2) (2006)
3. Gillani, B.B.: Learning theories and the design of e-learning environment. University Press of America, Inc., Oxford (2003)
4. Chen, W., You, M.: A framework for the development of online design learning environment. The bulletin of the JSSD 56(1), 83–92 (2009)
5. Ozkan, S., Koseler, R.: Multi-dimensional students' evaluation of e-learning systems in the higher education context: An empirical investigation. Computers & Education 53(4), 1285–1296 (2009)
6. Bradley, C., Oliver, M.: The evolution of pedagogic models for work-based learning within a virtual university. Computers & Education 38(1-3), 37–52 (2002)
7. Hsu, S., Marques, O., Hamza, M.K., Alhalabi, B.: How to design a virtual classroom 10 easy steps to follow. T.H.E. Journal 27(9), 96–98 (1999)
8. Masterman, L., Sharples, M.: A theory-informed framework for designing software to support reasoning about causation in history. Computers & Education 38(1-3), 165–185 (2002)
9. Passerini, K., Granger, M.J.: A developmental model for distance learning using the Internet. Computers & Education 34(1), 1–15 (2000)
10. Zahner, J.: Teachers explore knowledge management and E-learning as models for professional development. TechTrends 46(3), 11–16 (2002)
11. Cloete, E.: Electronic education system model. Computers & Education 36(2), 171–182 (2001)
12. Jones, A., Barnard, J., Calder, J., Scanlon, E., Thompson, J.: Evaluating learning and teaching technologies in further education. Association for Learning Technology Journal 8(3), 56–66 (2000)
13. Abrams, G., Haefner, J.: Blending online and traditional instruction in the mathematics classroom. The Technology Source Archives (2002)

14. Bender, D.M., Vredevoogd, J.D.: Using online education technologies to support studio instruction. Journal of Educational Technology & Society 9(4), 114–122 (2006)
15. Chen, W., You, M.-L.: The differences between the influences of synchronous and asynchronous modes on collaborative learning project of industrial design. In: Schuler, D. (ed.) HCII 2007 and OCSC 2007. LNCS, vol. 4564, pp. 275–283. Springer, Heidelberg (2007)
16. Chen, W., You, M.: The problems and influences of online collaborative design learning projects. Research in Arts Education 15, 105–137 (2008)
17. Chen, W., You, M.: Student response to an Internet-mediated industrial design studio course. International Journal of Technology and Design Education 20(2), 151–174 (2010)
18. Mazman, S., Usluel, Y.: The usage of social networks in educational context. Proceedings of World Academy of Science, Engineering and Technology 37, 404–407 (2009)

Behaviour Computer Animation, Communicability and Education for All

Francisco V. Cipolla Ficarra[1,2], Jacqueline Alma[3], and Miguel Cipolla-Ficarra[2]

HCI Lab. – F&F Multimedia Communic@tions Corp.
[1] ALAIPO: Asociación Latina de Interacción Persona-Ordenador
[2] AINCI: Asociación Internacional de la Comunicación Interactiva
[3] Universidad Nacional de La Pampa
Via Pascoli, S. 15 – CP 7, 24121 Bergamo, Italy
ficarra@alaipo.com, jacqueline_alma@yahoo.com, ficarra@ainci.com

Abstract. We are making a heuristic study to determine the benefits of computer animation to improve communicability in the educational processes. A diachronic analysis allows to research the influence of the design of the first animations of the 20th century on the current 2D and/or 3D characters. We are also examining the existing link between the artistic and technical context at the moment of carrying out computer animations. Finally, the results of a heuristic assessment of 2D and/or 3D characters with and without artificial intelligence components in the last two decades in Spain and Italy are presented.

Keywords: Behaviour Computer Animation, Education, Communicability, Computer Graphics, Multimedia, Evaluation.

1 Introduction

In the last few years we have been watching an almost perfect realisation of virtual characters or 3D animations on video consoles, computer screens, mass media communication, etc. [1] [2] [3] that make it hard to differentiate whether they are real human figures or not. This degree of the quality of the computer-made images applied to virtual tutors in e-learning courses prompts a rhetorical question: Are human beings ready to accept that the concepts are 100% taught by virtual characters, considering the realism of these “professors”, or will the characters continue to fulfil merely an assistance role in the learning process? Currently, because of economic cost in many national or private firms in the area of telecommunications, these virtual characters are steadily usurping real people’s tasks. However, automatic on-line information systems, for instance, do not solve veraciously the problems that users have, such as the non-functioning of an Internet service [4].

Additionally, the quality control systems of those services are not published on the websites of these telecommunications companies. Consequently, it is not possible to know for sure whether the result of inserting the virtual characters as assistants in front of users with real problems of non-functioning services is a positive result or not [4]. These virtual assistants are also starting to appear in the college environments for

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 538–547, 2011.

information and orientation tasks within the portal structure. Others, in contrast, have been used with didactical purposes for a long time, especially in the industrial context. In both cases we see how, little by little, there is a move from characters taken from comic books to more and more real characters, who have gone on to assist adult users, as well as young and adolescent users. In this process of continuous evolution in the area of communicability of interactive systems, we are trying to detect in the current work those elements that may contribute through an excellent quality of visual realism and synchronized movements to a greater integration of those characters in the whole educational process, from childhood right through till the continuous training of adults [5].

Addtionally, there will be an attempt to research the details of the elements that currently bolster the communicability of those characters, bearing in mind not only the on-line and off-line multimedia interactive systems, but also social mass-media. At the same time, a state-of-the-art of behavioural computer animation from its origins up to the present day will be drawn in order to detect which components of graphic computing bolster communicability. For example: the parts of the body of those characters which are frequently used, compared to those deriving from cinematic computer animation; the instruments and means of traditional teaching available for help with difficulties in the learning process; the behavioural style and the different personalities that can be assumed in relation to the context or environment in which users interact, and the personalization of the tutor by the potential users of the interactive systems. The latter, in relation to age, education and the knowledge/experience in the use of computers and motivation for attending the courses or contents taught by these characters. The different e-learning experiences, campus and virtual classrooms, etc. that we are trying to compile in our work will enable us to answer the rhetorical question we have stated. This answer will allow us to set up a series of research themes both for the immediate future and the long term, via Web 2.0 and more especially via Web 3.0

2 A Diachronic Study about Pioneers and First Animations

In the traditional animation literature, Emil Cohl is the pioneer who is always mentioned since its dawn [6]. He was a newspaper cartoonist who entered film production in 1905. Cohl applied the principle of the comic strip to the screen. “Fantasmagorie” [6] represents without any doubt the greatest expression in the metamorphosis of the characters. Besides, it is the intersection of the real world with the virtual world because you can see the hands of the makers of the drawings.

Therefore, there is in that work as in the other works of the time the emulation and simulation of reality [7]. These are principles that would be transferred to computers. Another of the historic novelties of Cohl and his work team was the technique of stop-motion photography which would set up the basis of the current “stop action” in computers. Moving or physically manipulating objects and setting up each one of the pictures or frames in relation to the television 25 frames per second or 24 frames per second in the cinema, the viewer perceives the objects in motion.

In this context it is worth mentioning the work of Cohl in “The Dentures” in 1909 [6]. Perhaps the most accomplished in the history of animation in that decade

"Mobelier Fidele" (The Automatic Moving Company) in 1910 where Cohl breathes life into an automatic moving (chairs, tables, lamps, baskets, etc.) move of their own accord. Another of the great subjects of those times was transoceanic communications. In this set of works we have "Professor Bonehead is Shipwrecked" by Emil Cohl. It was in these animations where many animators of those times focused to give movement to water. Certainly, works of few minutes runtime would entail many weeks of work for the production team.

Nowadays, commercial programs such as Maya incorporate the wave effect of water or movement of liquids, as one among many options. However, in the beginning animators required several tricks or effects to achieve those results, and have an artistic unity to coordinate the draughtsmen in order to generate a style of homogenous design in the drawings that later on would be animated. Something that even nowadays exists in the comics, especially with US, British and Japanese classical characters, such as: Astro Boy, Snoopy, The Flintstones, The Simpsons, Dragon Ball, etc.

Gertie, the Dinosaur (1914) a 12 minutes animation by Winsor McCay laid down [6], perhaps for the first time, the foundations of behaviour animation of a living animal, from which there were no direct references of movement. That is, that animation set a landmark in the history of animation and consequently of the current computer animation. That moment of giving a personality through movement to beings extinguished millions of years ago opened a long path of research and development where the animators needed work teams made up by a great number of professionals hailing from the most diverse disciplines. It is not for nothing that the vast majority of computer animation works that have won the movie Oscars have professionals from the physical sciences, computer programmers, mathematicians devoted to develop algorithms for graphic computing, fine arts artists, set designers, dress designers, historians to recreate scenes, botanists, meteorologists, astronomers, etc. who work on the characters and on the environment where these characters appear. These professionals may join on a steady basis the work staff or be external collaborators on a constant basis.

In these works we find the current foundations to carry out animations such as can be photography, montage techniques, the first special effects, etc. Although these works sprang from vignettes in the newspapers, in the case of Felix the Cat [8] or true events, such as "The Sinking of the Lusitania", by Winsor McCay [6]. In many works there was a textual reference, that is, traditional tales that were animated by introducing the Felix character. In this case, creativity was more focused on the technical aspects of the novel means like the effect of metamorphosis than on the storyboard.

A metamorphosis has always been very popular in the animal computer animations, especially in the advertisement or cinema environment. Metamorphosis still remains a valid resource to draw the attention of the potential users of interactive systems, regardless of the goal of the use of the computer. Obviously these early works had incorporated the text in the shape of comic balloons, since they stemmed from the comic. Besides, the movies were silent, and in the theatres in which these early animations were shown, they were generally accompanied by music instruments, such as the piano. In the case of "The Sinking of the Lusitania", the story is based on a real event. In this case each and every one of the 25,000 drawings that make up the animation were photographed. Besides, the context factor of the story took an essential role in the set design such as the animation of the water waves, while the sky was static. The theatrical factor was present in the animation since there was a curtain that opened up to give start to the animation. In order to give a greater realism to the narrative, some pictures of the passengers of the ship were included. The value of the static photography inside a dynamic means boosts the idea of diachronism and synchronism.

In the cartoon "Puss in Boots" by Walt Disney real image were combined with 2D animations [6] [9]. That is, animated characters that got out of the drawing sheet, and came into contact with the draughtsman's environment. It was even combined with plasticine sculptures. That is, with the animations of the sculptures and photographing the movements of the characters and in the milieu, photogram by photogram, just like the current Wallace & Gromit works, "Chicken Run", etc.

2.1 Mass Media and Animations Evolution

The dawn of the past century hit a landmark in the evolution of the animations linked to the constant breakthroughs of the social communications of that time. A short synthesis of those first 40 years can be listed in the following way (in brackets are the author/s and/or the dates of the presented events [8] [9] [10]:

- The first humour comic animation cartoons appear in 2D (Blackton, 1907).
- The first experiments of TV start in Russia (Boris Rosing, 1907).
- Animation of a comic strip published in a New York newspaper, whose main character was Little Nemo (Winsor McCay, 1911).
- The first experiments of synchronizing the sound to the movie images begin (Gaumont, 1912).
- The expression "animated cartoons" is coined in the film 'Newly weds' (Emil Cohl, 1912).
- Experiments and/or discoveries with the first effects of cinematographic camera: accelerated motion, slowed down, reverted or backwards, images fading, etc. (1913).
- Appearance of the first colour animated cartoons. It was the debut of Thomas Cat (John Bray, 1920).
- Incorporation of the key frame techniques in the cinema animations (Disney, 1926).
- Presentation of montage theory with the classical film 'Battleship Potemkin' (Sergei Eisenstein, 1926).
- Invention of the cinemascope, whose use will spread in 1953 (Chretien, 1928).
- Presentation of the 16 mm. Kodakcolor film (Eastman Kodak, 1928).
- The first TV channel is inaugurated in Washington DC 'W3XK' (Jenkins, 1928).
- Felix the Cat is the first cartoon that appears in the New York television station NBC (1928).
- Dinner Time was the first cartoon with sound, voice and music (Paul Terry, 1928).
- Premiere of the first Disney film in the Colony Theatre of New York (1928).
- Use of a system of photomechanical coloration in the film 'King of Jazz' (Walter Lantz, 1930).
- Use of the Technicolor in the animations 'Flowers and Trees' gets the first Oscar (Disney, 1932).

- Eastman Kodak presents the first 8 mm film (1932).
- The first TV educational network appears in the University of Iowa: W9XL (1932).
- The Three Little Pigs is the first cartoon designed with a storyboard (Disney, 1933).
- ‘Fantasia’ was the pioneering film in the use of stereophonic sound (Disney, 1941).

In this listing are the keys of the audiovisual which in the late 80s and early 90s would move into the computer screens, from the workstations in the computer animation industries to the desktops of the users keen on the images in motion (2D and/or 3D). The main sectors since the origins of animation can be depicted through the following ellipsis and in the middle we would place animation.

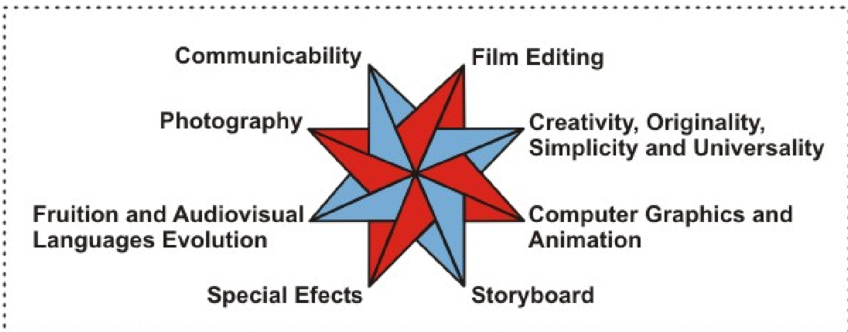


Fig. 1. Each one of the points of this star depicts an area of traditional animation

In principle that star has sectors of traditional animation which are redundant or duplicate, but with the passing of time some of them have converged among themselves, linking each other indefinitely, such as can be the coloration of images through the computers. That is, resorting to commercial programs of self-edition such as can be Photoshop or Photo-Paint. However, what has remained unchanged since the origins is the differentiation between the creative or artistic sectors and the technological [11]. Clearly these are two split sectors.

There may be intersections between them but the technological sector has been taking up a bigger area inside the set of elements that make up the artistic objects. For instance, the possibility of making the drawings directly on the computer without the need of drawing traditionally, that is, by hand and then moving those drawings to the computer through the use of the scanner.

This division of sectors has found several solutions of its own depending on the geographical environment where the computer animations are made.

In the previous examples of the early 20th century were already the foundations of the computer animation cinematographic industry, whose success or failure derives from many factors such as:

- Achieving the communicability of the contents, for instance, resorting to the style guidelines to avoid the confrontation between the localization and the globalization of the explicit and implicit message of the animations.

- Resorting to the utmost simplicity of style to reach the greatest universality of the message and a wider national and international circulation in the least possible time.
- Differentiating between emulation and simulation of reality so that the user can identify with the characters and keep the attention and interest in the content that is presented in an interactive way; whether it is sequential or not.
- Generating work teams where the several disciplines and professionals agree on the actions in a democratic and balanced way.
- Keeping constantly the spirit of curiosity in the face of the audiovisual novelties, software, hardware, etc. by the members of the production team and establishing mechanisms of constructive discussion among all the members so that these breakthroughs can be evaluated for its later incorporation in the productive processes.
- Fostering the interest towards those pastimes that can improve the common dialogue between technicians and artists. For instance, a computer science expert who participates in photography courses and a virtual characters designer who attends the presentations of the latest versions of the software for computer animation and/or self-edition.
- Counting on communicability experts in each one of the production process stages, that is, from design until post-production.
- Defining quality attributes and aim them inside the cardinal points of computer animation, especially with the characters to whom they instill life and personality: universality, simplicity, originality and humour.

3 Human Characters and Behaviour: Animations for Users

The 2D animation deriving from comic like its evolution towards 3D and the later combination of 2D and 3D has made its importance in the educational process apparent.

Millions of students before starting to write or read have seen cartoons in the traditional communication media, that is, in print media, television and cinema [12] [13]. We can see an example in the following Snoopy abecedary (www.peanuts.com) [14]:

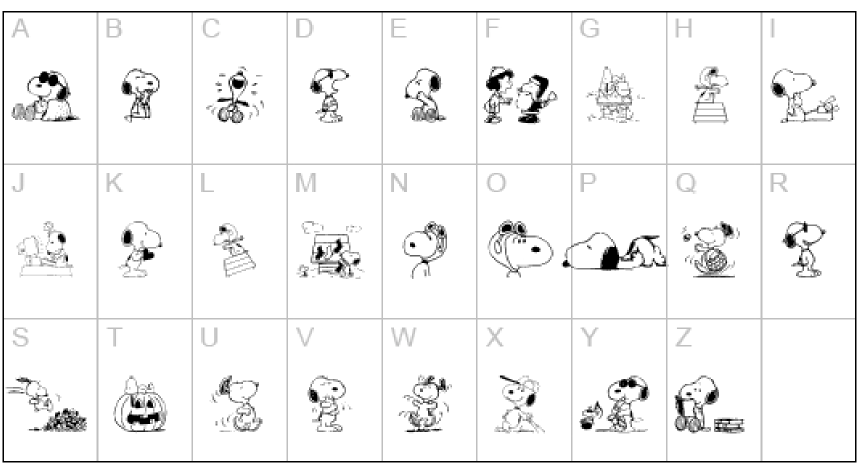


Fig. 2. Snoopy accompanies each one of the letters of the abecedary in different situations

This possibility of associating something unknown such as the number or a letter with a sound through a character already known or seen on television or in magazines facilitates the reading learning process, especially among the youngest kids, as long as they know Snoopy. This character from the comic strips becomes automatically a guide or study mate – so to speak a virtual agent – for the children. Obviously, for reasons of social psychology, pedagogy, sociology, although we have a virtual agent of Snoopy in 3D, this can't replace the teacher in the classroom. A virtual agent can't take the place of the teacher in the first moments of the learning how to read, not only with the students in the nursery school or early courses even with those adults who can't read or write [14]. The virtual agents have a helping function in the educational process. In the industrial case they may have a role of constant support, in case of doubts about the usage of certain equipment, machinery, etc., or in situations which occur suddenly, with the need of making a quick decision.

As it almost always happens with the quality attributes in the communicative process, such as empathy, inference, self-evidence, etc., those virtual characters which are known beforehand by the users or viewers of the audiovisual contents know what is the behaviour of said characters. As it happens with self-evidence in an interactive system, the user has the feeling of gaining time in navigation, since he anticipates the design and the purpose of the components which are shown to him for the first time. The anticipation of meaning has a positive influence on another attribute of quality such as motivation. Self-evidence is essential to increase the trust in those users who lack previous experience in the use of the multimedia interactive systems, for instance, what prompts a greater motivation to keep on interacting with the system.

Traditionally, Kahn [15] divided self-evidence into several categories inside a hypermedia system: First, the navigation styles of the application of the different structures without getting acquainted with the whole system, that is, those navigation styles may be related to the permanent possibility of advancing or going backwards from a bidirectional guided link that is made up by a set of screens; second, the anticipation of the destination of a link from the original node (in the case of the nodes with definitions which are activated from the words that are underlined in a text); third, the understanding of the degree of control that is available in the active means to operate with them. The control is exerted through the forward, pause, backward keys, etc. However, Kahn excluded other components of self-evidence which are related to the category of the design known as presentation or layout such as certain constituents of the metaphor, to widen the dimension of the screen at the moment of watching a YouTube video or modifying the volume. That is, if in the context there are design isotopic lines, all those factors linked to the previous will help self-evidence and very specially the empathy and inference in the communicative process. Some of these quality attributes have been the target of study and set in motion of projects related to the expert systems generated through artificial intelligence.

3.1 Behaviour Computer Interaction, Artificial Intelligence and Education

In the previous examples of the animations the user has always had a passive or active role in the face of the animations, whether as a receptor of the images on the cinema screen or the television or in the interactive process with computers. In this latter case, the main reasons why the user interacts may be to search for information, training, pastime (videogames) and social relationships. Now the Internet has incorporated in the last two decades the notions of artificial intelligence. Consequently, we have virtual agents or autonomous animated characters which can remain for some length of time without the dependence of the human being. That is, characters who act following the parameters or models that human beings use to make their decisions. These expert virtual agents are capable of solving problems in fractions of a minute, as long as they are not easy to solve by the human being (generating an expert system is not worth the trouble) or whether the problem is very big, since it requires strong investments in hardware.

The resolution of the problems must be based on experience and knowledge, leaving aside the random factor. Besides, the resolution of the problem must not be based on intuitive or common sense methods. All of that with the purpose of turning the expert agents into an ideal assistant to get complete information in the face of decision making.

In the case of on-line information such as the virtual tutors of the universities, phone information services, Internet, etc., endless communicability errors can be seen [4]. Errors which can be put down to their design stage, since they are based on intuitive aspects. From the point of view of the quality of the animation and graphic information, they have a low quality from the point of view of movement, textures, modelling, etc. This reality has lead us to carry out a study of the behaviour computer animation among the cinematographic, television, virtual tutors on-line and in the commercial and free-access on-line videogames, in the Spanish and Italian portals (300 randomly selected works). In the lower part are the quality attributes that refer mainly to the aspects of graphic computing (static and dynamic) and communicability. Besides, characters and virtual agents are synonyms in the figure 3. The interested readers may find a wide definition of those notions in the following bibliographic references [16--20].

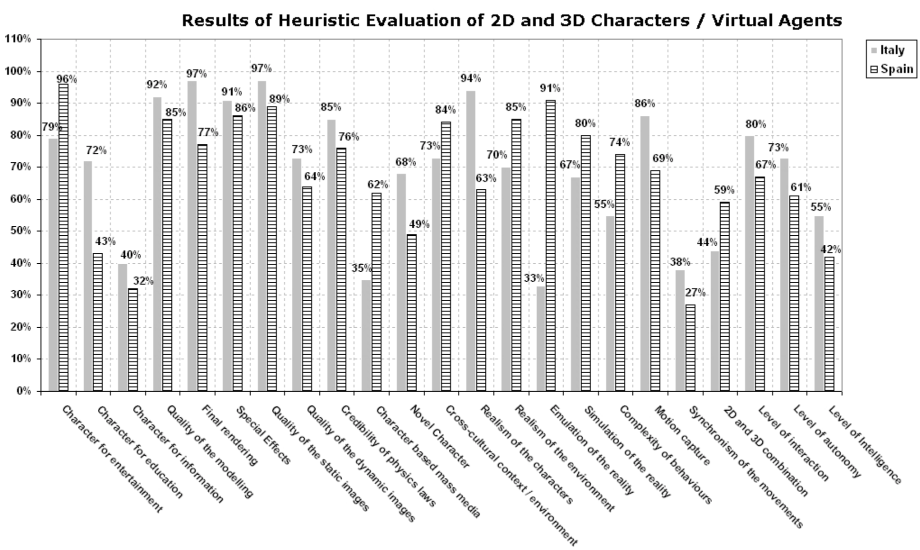


Fig. 3. The results make apparent the variability of the quality attributes between Italy and Spain

4 Conclusion

The first animations in the social mass media underwent a parallel evolution to the breakthroughs in the technologies of those times. However, the foundations of the design for the animations have remained intact along the decades, even in the computer-made animations and in special effects. The key elements to draw the attention and prompt the motivation of the user or international viewer of the animations in cinema, television, computer, mobile phone, etc., are the creativity and the originality of the characters linked to the behaviour and character that they have at the moment of interacting with their virtual peers or with the user. Humour also needs to be included, the same as the simplicity in the communicability and the context where these characters move, 2D and/or 3D. In the multimedia interactive systems, the category of layout or presentation of the information generated through computer graphics play an essential role. In that category it is necessary to differentiate between the emulation or simulation of reality at the moment of designing and animating virtual characters. In the expert agents, it is necessary to add a middle level in the resolution of problems. The results reached in the heuristic evaluation make apparent the greater communicability in the characters for the entertainment rather than for education in Spain and Italy. Although the quality of the computer graphics, including the realism of movement, is greater in the Spanish systems. In the Spanish systems prevails the quality of the modelling. That is, there is a greater quality in the static images in Spain than in Italy. The final rendering (textures, lightening techniques, combination of colours, environmental effects, etc) and the refining of each one of the frames that make up a scene are better in Spain than in Italy. The quality of the behaviour computer animation in our universe of study along the last two decades has turned out to be low in both countries. The reason is the lack of interaction between the formal and the factual sciences in the university academic training and the industrial sector of the audiovisual.

References

1. Badler, N., et al.: Virtual Humans for Validating Maintenance Procedures. Communications of ACM 45(7), 56–63 (2002)
2. Scharver, C., et al.: Designing Cranial Implants in a Haptic Augmented Reality Environment. Communications of ACM 47(8), 32–39 (2004)
3. Lu, R., Zhang, S.: Automatic Generation of Computer Animation. Springer, Berlin (1998)
4. Barnes, M., Meyers, N.: Mobile Phones: Technology, Networks and User Issues. Nova Science Publishers, New York (2011)
5. Cipolla-Ficarra, F., Cipolla-Ficarra, M.: Computer Animation and Communicability in Multimedia System and Services: A Trichotomy Evaluation. In: Damiani, E., Jeong, J., Howlett, R.J., Jain, L.C. (eds.) New Directions in Intelligent Interactive Multimedia Systems and Services - 2. Studies in Computational Intelligence, vol. 226, pp. 103–115. Springer, Heidelberg (2009)
6. Video. Los comienzos de la animación. Divisa Ediciones, Valladolid (1995)
7. Cipolla-Ficarra, F., et al.: Human-Computer Interaction. Tourism and Cultural Heritage. Springer, Berlin (2011)

8. Barrier, M.: Hollywood Cartoons: American Animation in Its Golden Age. Oxford University Press, Oxford (2003)
9. Bendazzi, G.: Cartoons: One Hundred Years of Cinema Animation. Indiana University Press, Indiana (1999)
10. Rosebush, J.: Historical Computer Animation. ACM Siggraph, New York (1992)
11. Cipolla-Ficarra, F., Nicol, E., Cipolla-Ficarra, M.: Computer Graphics and Mass Media: Communicability Analysis. In: Cipolla Ficarra, F.V., de Castro Lozano, C., Pérez Jiménez, M., Nicol, E., Kratky, A., Cipolla-Ficarra, M. (eds.) ADNTIIC 2010. LNCS, vol. 6616, pp. 182–192. Springer, Heidelberg (2011)
12. Wells, P.: The Animated Bestiary: Animals, Cartoons, and Culture. Rutgers University Press, New Jersey (2008)
13. Kerlow, I.: The art of 3D Computer Animation and Imaging. John Wiley, New York (2000)
14. Cipolla-Ficarra, F.: Heuristic Evaluation of Animated Help in Hypermedia. In: CD Proc. HCI International, Las Vegas (2005)
15. Kahn, P.: Global and Local Hypermedia Design in the Encyclopedia Africana. In: IWHD 1995, pp. 115–129. Springer, Berlin (1995)
16. Cipolla-Ficarra, F.: Eyes: A Virtual Assistant for Analysis of the Transparency and Accessibility in University Portal. In: CD Proc. Applied Human Factors and Ergonomics, Las Vegas (2008)
17. Cipolla-Ficarra, F.: Virtual Learning Environment: Quality Design for Foreign Languages in Multimedia Systems. In: Damiani, E., Jeong, J., Howlett, R.J., Jain, L.C. (eds.) New Directions in Intelligent Interactive Multimedia Systems and Services - 2. Studies in Computational Intelligence, vol. 226, pp. 117–127. Springer, Heidelberg (2009)
18. Cipolla-Ficarra, F., Cipolla-Ficarra, M., Vera, P.: Communicability for Virtual Learning: Evaluation. In: Jacko, J.A. (ed.) HCII 2009. LNCS, vol. 5613, pp. 68–77. Springer, Berlin (2009)
19. Cipolla-Ficarra, F.: Quality and Communicability for Interactive Hypermedia Systems: Concepts and Practices for Design. IGI Global, Hershey (2010)
20. Styliaras, G., Koukopoulos, D., Lazarinis, F.: Handbook of Research on Technologies and Cultural Heritage: Applications and Environments. IGI Global, Hershey (2011)

An Intelligent Task Assignment and Personalization System for Students' Online Collaboration

Asterios Leonidis[1], George Margetis[1], Margherita Antona[1], and Constantine Stephanidis[1,2]

[1] Foundation for Research and Technology – Hellas (FORTH), Institute of Computer Science, GR-70013 Heraklion, Crete, Greece
cs@ics.forth.gr
[2] University of Crete, Department of Computer Science

Abstract. This paper discusses a framework that supports collaborative learning activities in smart environments. When designing or developing collaborative environments numerous fundamental requirements should be taken into consideration to maximize their potentials. These best-practices outline strategies regarding (i) group formation, (ii) role assignment, (iii) personalized support, and (iv) activity monitoring. A multi-tier architecture scheme is employed, on top of the "ClassMATE" system, where every module addressees some of these aspects and their combination results in a complete framework that enables both online and offline collaboration in the smart classroom.

Keywords: ambient intelligence, education, smart classroom, collaborative learning.

1 Introduction

The evolution of Information Technology (IT) for more than three decades has drastically affected the way users interact with personal computers and increased their expectations from technology. Towards this objective, researchers developed novel concepts to provide content-rich invisible computing applications, eventually leading to the emergence of the Ambient Intelligence paradigm. Ambient Intelligence is a vision of the future which offers great opportunities to enrich everyday activities (e.g., on the road, at home, at work, etc.). Considering that ICT (Information and Communication Technologies) have been proven to play an important role in education, this work investigates the promising potentials of AmI in the education domain and in particular in the domain of collaborative learning.

The notion of Smart Classroom has been around already for a few years. In a Smart Classroom, conventional classroom activities are enhanced with the use of pervasive and mobile computing. However, the majority of the current approaches towards the realization of the Smart Classroom address various issues unilaterally, either from the technological or the educational perspective, neglecting the main objectives of disseminating knowledge and supporting the student during the learning process.

Education is defined as the process of acquiring general knowledge, developing judgmental thinking and generally preparing oneself intellectually for mature life. The

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 548–557, 2011.

Smart Classroom can be considered as a vehicle towards achieving these goals, as the presence of computer technology not only motivates learners' engagement in the various learning activities but also establishes the groundwork for computer-assisted collaborative learning applications that benefit from context-awareness. The work reported in this paper aims to introduce an innovative solution for efficient learning through collaboration in smart environments.

2 Background

A learning approach can be classified under one of the following classes: individualistic, competitive and collaborative. In individualistic learning, knowledge is acquired as the learner studies at his own pace towards mastering a topic. In competitive learning, a learner acquires knowledge by maximizing performance and overwhelming every opponent. Finally, in collaborative learning, a learner consciously acquires knowledge and masters a topic and unconsciously cultivates his social skills through participating in group activities. Two independent studies, [8] which dealt with traditional learning approaches without any technological aid, and [1] which examined computer supported learning, reached the conclusion that collaborative learning offers both a better learning experience and knowledge gain.

Collaborative learning is often confused with the notion of co-operative learning and used interchangeably in the literature. According to [1] co-operative work is accomplished by the division of labor among participants, where each person is responsible for a portion of the problem solving, whereas collaboration involves the mutual engagement of participants in a coordinated effort to solve the problem together. The mutual engagement in group activities unveils the social aspects of collaboration, as several social skills are cultivated in addition to cognitive skills. In conclusion, through collaborative learning, learners improve their performance [7], achieve results of high quality, shape positive attitude towards instructional experience [8], increase their metacognitive understanding and rate of knowledge capturing [12], and eventually achieve mastery.

Concurrently with the explosion of the World Wide Web, the need for remote (anywhere) and asynchronous (anytime) education revolutionized collaborative learning, while the emerging concept of computer-supported collaborative learning (CSCL) was proven to have a positive impact on students' learning [1, 5, 10, 8, 12, 14, 18, 6]. In this respect any computer – supported collaboration system should follow the features discussed next in order to maximize its potential impact:

- A collaborative task should incorporate both group and individual goals to motivate active participation and maximize the quality of the results [16].
- A group should be formed in a way that not only maximizes cohesion and density but also includes members that can be benefited the most from their participation (e.g. average learners) [5, 12, 7].
- Social skills can be cultivated through collaboration. However, extended engagement with organizational activities may result in focus loss or uneven roles distribution [17]. To address that, predefined collaboration scripts [10] can be introduced at the beginning of the task to control roles distribution and eliminate

potential collisions that may occur due to emerging, unforeseen roles. Scripts usage though should remain limited and gradually fade out, as organization skills, cooperation abilities and intergroup relations are only learned through active involvement in the planning process [2, 16, 14, 5].

- Learners' participation and knowledge inference can be maximized by [14, 16, 5, 12]: (i) enforcing continuous instructor's involvement which motivates contribution and (ii) advertising the "head" participants as collaboration centers where the less-active members can seek for unassigned or incomplete tasks.
- Finally, monitoring of the overall collaboration process can be exceptionally useful as [6, 18]: (i) collaboration patterns and "best-practices" can be identified for future reuse, (ii) personalized assistance can be provided when necessary and (iii) employment of acquainted skills can be evaluated in practice under real conditions.

3 Related Work

During the past decades the emerging field of educational applications and in particular collaborative learning applications, received much attention by the ICT community resulting in a diversity of approaches, ranging from small-scale systems aiming to motivate active participation to large-scale systems supporting every aspect of a collaborative activity.

In more detail, [13] proposed a collaborative platform where a combination of hand-writing recognition hardware and software enabled the exchange of text-based notes between the participants of a group and their instructors. In [20] situation-aware portable devices (i.e., PDAs) facilitated students' enrollment in a course (e.g., deadline reminders, personal area, etc.). In addition to course-related facilities, a limited set of desirable collaborative features was also supported like data synchronization and exchange. As regards the group formation, that task was performed offline by the instructor prior to the activity's initiation.

A similar approach is presented in [4], where interconnected handheld devices were provided to the learners to enable their interaction, whereas a monitoring strategy was applied to collect real-time data. Those data were instantly analyzed by the instructors to determine their corrective actions, which eventually assisted learners with acquiring new knowledge on their own. In [19] a ubiquitous learning environment was proposed which aimed to optimize collaboration and maximize success potentials by selecting: (i) the right collaborator, (ii) the appropriate content and (iii) the suitable service. The various collaboration actions were performed automatically (i.e., group formation, content discovery, content exchange) while every interaction was monitored by the system. Learner's profile is mostly exploited during group formation while during collaboration, its use remains limited resulting in inadequate personalized assistance.

In [3] a new system that could be easily tailored by educators in order to support the realization of scripted collaborative learning situations was presented. On the one hand it enabled instructors to provide a script that specify the sequence of activities, and on the other hand it employed grid-oriented technologies to provide the necessary activity tools to the learners. Personalized guidance was also supported to help learners while performing task activities. Finally, in [15] an agent-based system was proposed, whose core consisted of four major components that offered the necessary collaboration-related functionality (i.e., activity planning, monitoring, evaluation, assistance and administration). In terms of user interface, every "screen" was presented in the form of a web page dynamically generated by servlets and applets.

However, all the aforementioned approaches lack the dimension of environmental context – awareness that could foster collaboration among students. In this paper an innovative framework is proposed to bridge that gap and outline the benefits of computer-assisted collaborative learning in smart environments.

4 Architecture Overview

The proposed framework is built on top of the ClassMATE system [11] and adds support for collaborative educational activities. ClassMATE aims to bridge the gap between the ambient environment and the educational applications in a transparent manner, by enabling applications to exploit contextual information provided by the environment, in order to enhance the educational process. In particular, ClassMATE facilitates student's learning activities by simplifying everyday tasks and personalizing content to every individual learner. On the other hand, ClassMATE assists the teaching process by automating common teachers' activities (e.g., material distribution, homework collection, progress monitoring), thus permitting the teacher to better focus on the educational process.

ClassMATE follows a modular architecture and its core consists of four major components layered in parallel: the Device Manager, the Data Space, the User Profile and the Security Manager, glued together via the fifth major component, the Context Manager. These five major components also provide the building blocks for additional functional modules that can be developed and integrated into the system.

The Context Manager is the orchestration component of the ClassMATE's architecture. It monitors the ambient environment and makes context-aware decisions that control the operation and collaboration of ClassMATE's services and applications. The Device Manager offers a generic mechanism for heterogeneous devices manipulation by any ClassMATE-enabled application in the ClassMATE cloud.

The role of the Data Space is threefold: (i) it implements a centralized content repository, providing transparent content access and management, (ii) it encapsulates a content classification mechanism, based on IEEE's LOM specification [9], providing the necessary content-related rationale to data mining procedures, and (iii) it encapsulates a sophisticated filtering mechanism for personalized content delivery. Finally, The User Profile implements the classroom's users (students and teachers) behavior monitoring and evaluation, in order to provide user related metadata to the ClassMATE's services and applications.

4.1 The Collaboration Manager and Collaboration Strategy Manager

As aforementioned, ClassMATE offers fundamental functionality of a smart classroom enabling thus extensions to be added such as the Collaboration Manager and Collaboration Strategy Manager, which provide intelligent manipulation of collaborative activities in the classroom. The Collaboration Manager acts as a special-purposed Context Manager which handles the various collaborative activities of the classroom's users, in strong cooperation with the Collaboration Strategy Manager, upon the built-in ClassMATE's monitoring mechanism which facilitates the overall notification process. Figure 1 presents the architecture of the Collaboration Extension, described in detail in the next section.

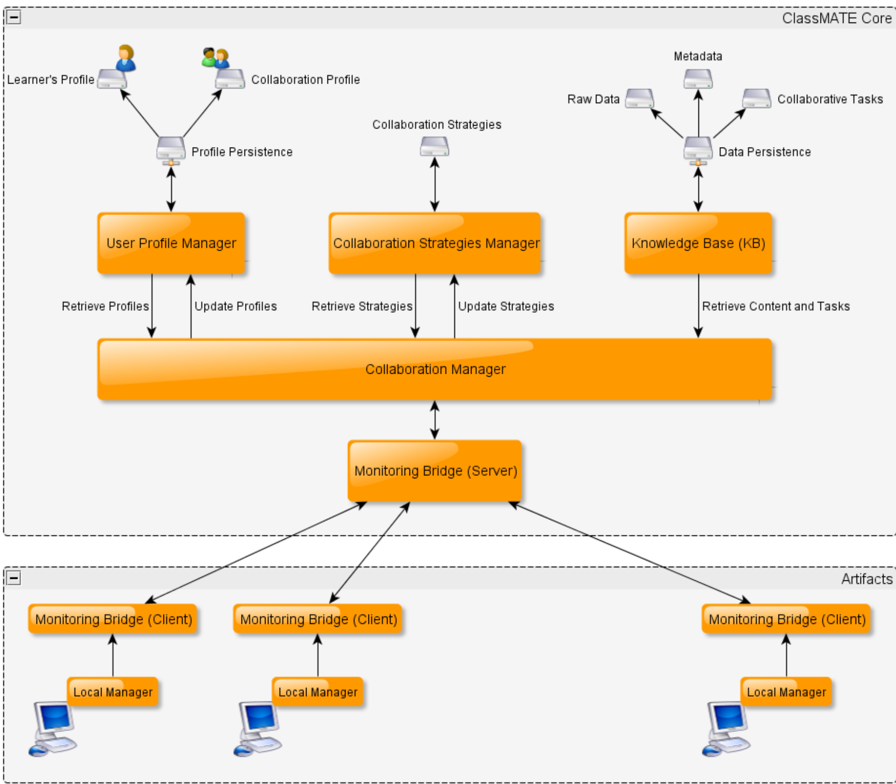


Fig. 1. Collaboration extension of ClassMATE

5 Implementation Details

Every collaborative tasks lifetime includes the following phases: (i) planning, throughout which the initial preparation of collaboration activities takes place, (ii) enactment during which the collaboration plan is deployed and monitored in real-life situations, and (iii) evaluation which aims to qualify and quantify the acquired knowledge against a short- and long- term education plan.

5.1 Collaborative Activity Planning

In the planning phase the Collaboration Manager sequentially executes the following sub-tasks: (i) identification of the task's requirements, (ii) designation of the candidate participants, and (iii) role assignment to each participant. Taking advantage of the ClassMATE platform, the entire process can be executed automatically, while the instructor can modify, at real-time, the role mappings if necessary.

In order for ClassMATE to effectively orchestrate the content delivery and provide context – aware services to the students, it handles every educational content used in the classroom as a Learning Object which is accompanied by LOM-based metadata that describe its educational attributes and facilitate its automatic retrieval. Nevertheless, LOM does not include any collaboration-related attributes, whereas the various extensions [10] proposed to address collaboration have not been standardized yet. In the presented approach a hybrid model is adopted. The LOM schema remains unchanged, the necessary collaboration concepts compile the "Collaboration Taxonomy" and every learning object can be associated with a concept through the content classification mechanism. The combination of native educational-oriented LOM attributes (e.g., part-of, difficulty level, domain, etc.) with collaboration aspects (e.g., two-person assignment, highly collaboration skills needed, etc.) can fully describe a learning object and subsequently facilitate its automated discovery in the context of a collaborative assignment.

In addition, the ClassMATE's User Profile module is able to determine the selection of the best candidate for an assignment, however it is not adequate for selecting coherent collaborative groups. Therefore, the collaborative profile was introduced to aggregate the necessary attributes that characterize the learner's behavior inside a working group and publish them to the Collaboration Manager. The included attributes are: (i) co-operability, that defines learner's ability to operate in a team, (ii) productivity, that measures learner's active involvement and contribution in group activities, (iii) rate of conflict, that reflects the learner's the spirit of contradiction, (iv) punctual behavior, which shows whether the learner delivers the assigned work on-time, (v) organizational skills, that measures learner's participation in the decision-making process towards task assignment and (vi) contribution type, that determines whether the learner is a "thinker" or a "typist".

Given that the collaborative task definition and the participants' collaborative profiles are available, the remaining task which is needed for the collaboration plan to be accomplished is the roles assignment task, which is the most challenging. Given a collaborative task *ct*, with the educational requirements *er* and the collaboration requirements *cr*, find potential participants that satisfy both *er* and *cr*.

$$\boxed{\text{ct: } \{\text{er, cr}\} \rightarrow \{p_1, p_2, \ldots, p_n\}, \forall\, p_i \;\exists\; p_i \text{ attribute} \in \{\text{er} \cup \text{cr}\}}$$

In order for the system to do so, initially it applies a set of semantic rules to *er* and *cr* so as to infer the quantitative attributes that a learner should have at minimum to be considered as a potential candidate, whereas both the educational and the collaboration profile are used to determine whether a learner satisfies these requirements. For example, consider the following scenario where the instructor delivers a collaborative task in which the learners should write a brief essay about Netwon's 1st law. Through the inference process, the system identifies that the following criteria should be met regarding educational aspects: (i) domains: physics and foreign language, (ii) difficulty: medium, (iii) deadline: soon, and collaboration aspects: (a) number participants: 2, (b) collaboration skills: high. As a result, the system should find two participants, one with good writing skills and one good at Physics.

Additionally to the learner's profile, the system utilizes contextual information like learners' personalized schedule to exclude incompatible candidates. Finally, taking into consideration timing constraints, the system can automatically exclude candidates that are known as "trouble makers" and focus on candidates that are either punctual or have excelled in collaborative tasks in the past.

As soon as a preliminary set of candidates is defined, the instructor can either accept the suggested participants or discard them and manually define the desired groups. If manual definition is recognized to be the rule instead of the exceptional case, then the system collects the various profiles and attempts to extrapolate any patterns and associate them with the particular instructor for future use. Finally, if particular strategies are known to offer poor performance in particular cases, then the instructor can be informed beforehand.

5.2 Enactment, Monitoring and Personalized Assistance

During enactment the system initiates personalized assistance and monitoring facilities. The role of assistance is apparent, whereas the role of monitoring is twofold. From an educational perspective, monitoring facilitates progress tracking concerning either an individual, or a group or even a whole class, while from a collaboration perspective it facilitates the dynamic reconfiguration of the process's plan at run-time to handle emerging needs (e.g., participant's unavailability).

Dynamic activity reconfiguration is natively supported by the ClassMATE; however, that mechanism was extended to support dynamic reconfiguration of collaborative activities as well. The initial plan is automatically generated by the system based on the (i) assignment's predefined plan (e.g., part A should be executed before part B) and (ii) the personalized schedule of the participants. Despite that the generated plan eliminates conflicts that may occur due to activities overlap, it is still subject to modification at run-time. Certain priority rules are applied to prevent any potential deadlocks, while this feature can be disabled for particular individuals if abuse is detected.

Dynamic activity reconfiguration directly affects a learner's collaboration profile negatively or positively. If a learner constantly postpones his assigned activities then his rank will decrease. On the other hand, if a learner supplants his own schedule to facilitate the team, then his rank will increase and the system will further encourage him by planning future activities to fit his schedule.

Every collaborative activity is essentially an educational activity, thus the educational mechanisms supported by the ClassMATE system are deployed as well, and in particular the content personalization facility. The support given to a learner is adapted according to (i) his educational profile, (ii) the task's educational requirements encoded in its metadata (e.g., difficulty level, time to learn, etc.), and (iii) teammate(s) educational profile(s). The latter aims to balance contribution between "weak" and "excellent" learners, as the "excellent" will most likely surpass the “weak”. For example, if a group consists of two learners, one “weak” and one “excellent”, and they have to collaborate on a multiple choice assignment, then the “weak” learner’s participation can be encouraged by automatically enabling the extended hint option whenever a hint is asked. Similarly, if all participants need further help for the same question, then the provided hint will be personalized to their individual needs as well. For instance, in the previous example, the “excellent” learner will receive the simple and the “weak” learner the extended version of the same hint.

To ensure that learners do not take advantage of the personalized assistance facility a failsafe mechanism was introduced to eliminate any potential “ghost” users. A “ghost” user is a user who constantly asks for hints or leaves his teammate(s) to answer a question, or uses instant messaging to receive the correct answers. The system, in an attempt to discourage and eliminate such behaviors, monitors the ratio of hints requests versus individual’s contribution and the exchanged messages and upon improper use detection the strategy dynamically updates to resolve them. For instance, if a learner constantly asks for hints then system will not display the hint immediately but instead try to motivate him to answer the question by himself. As a result, the content personalization component not only records every asked hint in the learner’s educational profile to facilitate progress report generation, but also notifies the collaboration manager to update the collaboration profile if needed.

Monitoring facilitates instructors’ participation as well. First of all, computer-aided reporting simplifies class management through a multi-group view that presents the overall progress, while the instructor can focus on a particular group or even an individual participant to evaluate the progress made (e.g., actual progress, number of hints taken, correct / wrong answers ratio, etc.). Moreover, the instructor can view the created content (e.g., an essay) and motivate the participants by providing comments if necessary. Finally, a feedback mechanism is offered, through which the instructor can notify to the system about the activity’s progress and potentially modify the collaboration strategy.

5.3 Process Evaluation and Strategies’ Adaptation

The system accumulates various data at runtime and in combination with instructor’s assessment evaluates the applied strategy. During the evaluation process, both the collaboration efficacy and every individual’s progress are assessed. Based on evaluation findings the strategy will either remain the same, or the system will attempt to improve it for future use.

Collaboration strategies have been mentioned in the planning phase, where the system selects the most appropriate one for use. A strategy associates an educational task with different characteristics (e.g., assignment type, domain, difficulty level, etc.) with collaborative tasks and learners’ characteristics. The analysis of the maintained log can identify and generalize strategies that perform well in many different cases or discard those that do not. Moreover, if an ambiguity emerges where similar strategies perform equally in similar situations, then the instructor can compare them and select the most suitable. Such a decision may result in standardization of “best-practices” that maximize knowledge acquisition, and cultivation of collaborative skills (e.g., in groups of two always combine a “good” and an “average” learner to maximize the performance of both).

6 Conclusion and Future Work

This paper discussed a framework that enables collaborative learning in smart educational environments. It offers the necessary features (i.e. coherent group formation, learners' motivation, activity monitoring, etc.) to maximize its potential impact and offers a collaborative environment where the leaners can efficiently collaborate and educate themselves while the system offer personalized support to each participant.

The next steps for the presented work involve an exhaustive user-based evaluation in order to acquire additional useful feedback from end users regarding the robustness of the system, and the extension of the supported collaborative scenarios and the administration facilities to better orchestrate them.

References

1. Feyzioglu, B., Akcay, H., Sahin-Pekmez, E.: Comparison of computer assisted cooperative, competitive and individualistic learning: An example of Turkey. In: Proceedings of Congrès International d'Actualité de la Recherche en Education et en Formation (2007)
2. Ames, C.: Classrooms: Goals, structures, and student motivation. Journal of Educational Psychology 84(3), 261–271 (1992)
3. Bote-Lorenzo, M.L., et al.: A tailorable grid service based system that supports scripted collaborative learning. Computers & Education 51(1), 155–172 (2008)
4. Cortez, C., et al.: Teaching Science with Mobile Computer Supported Collaborative Learning (MCSCL). IEEE Computer Society, Los Alamitos (2004)
5. Dewiyanti, S., Brandgruwel, S., Jochems, W., Broers, N.: Students experiences with collaborative learning in asynchronous Computer-Supported Collaborative Learning environments. Computers in Human Behavior 23(1), 496–514 (2007)
6. Duran, E., Amandi, A.: Collaborative Student Profile to Support Assistance in CSCL Environment (2008)
7. Hooper, S.: The effects of persistence and small group interaction during computer-based instruction. Computers in Human Behavior 19(2), 211–220 (2003)
8. Humphreys, B., Johnson, R.T., Johnson, D.W.: Effects of cooperative, competitive, and individualistic learning on students' achievement in science class. Journal of Research in Science Teaching 19(5), 351–356 (1982)
9. IEEE LOM. Draft Standard for Learning Object Metadata. IEEE Learning Technology Standards Committee, IEEE 1484.12.1 (2002)
10. Leo, D.H., et al.: COLLAGE: A collaborative Learning Design editor based on patterns. Educational Technology and Society 9(1), 58–71 (2006)
11. Leonidis, A.: ClassMATE: Classroom Multiplatform Augmented Technology Environment (Master's thesis) (2010), e-Locus at `http://elocus.lib.uoc.gr/dlib/d/3/6/metadata-dlib-d19ded9b6c0938f4723672b56b78ebe2_1288598057.tkl`
12. Lipponen, L., Rahikainen, M., Lallimo, J., Hakkarainen, K.: Patterns of participation and discourse in elementary students" computer-supported collaborative learning. Learning and Instruction 13(5), 487–509 (2003)

13. Milrad, M., Perez, J., Hoppe, U.: C-notes: designing a mobile and wireless application to support collaborative knowledge building. In: Proceedings of IEEE International Workshop on Wireless and Mobile Technologies in Education 2002, pp. 117–120. IEEE Computer Society, Los Alamitos (2002)
14. Pawan, F., Paulus, T.M., Yalcin, S., Chang, C.F.: Online learning: Patterns of engagement and interaction among in-service teachers. Language Learning and Technology 7(3), 119–140 (2003)
15. Sheremetov, L.: EVA: an interactive Web-based collaborative learning environment. Computers & Education 39(2), 161–182 (2002)
16. Slavin, R.: When does cooperative learning increase student achievement? Phychological Bulletin 94(3), 429–445 (1983)
17. Strijbos, J.-W., Weinberger, A.: Emerging and scripted roles in computer-supported collaborative learning. Computers in Human Behavior 4, 491–494 (2010)
18. Trentin, G.: Using a wiki to evaluate individual contribution to a collaborative learning project. Journal of Computer Assisted Learning 25(1), 43–55 (2009)
19. Yang, S.J.H.: Context Aware Ubiquitous Learning Environments for Peer-to-Peer Collaborative Learning. Educational Technology & Society 9(1), 188–201 (2006)
20. Yau, S.S., et al.: Smart Classroom: Enhancing Collaborative Learning Using Pervasive Computing Technology. Engineering, 1–9 (2003)

Using Interface Design with Low-Cost Interactive Whiteboard Technology to Enhance Learning for Children

Chien-Yu Lin[1], Fong-Gong Wu[2], Te-Hsiung Chen[3], Yan-Jin Wu[4], Kenendy Huang[4], Chia-Pei Liu[5], and Shu-Ying Chou[6]

[1] Department of Special Education, National University of Tainan, 33, Sec. 2, Shu-Lin St.,Tainan, Taiwan 700, R.O.C
linchienyu@mail.nutn.edu.tw
[2] Department of Industrial Design, National Cheng Kung University, 1, Ta-Hsueh Road, Tainan, Taiwan 701, R.O.C.
fonggong@mail.ncku.edu.tw
[3] National Tainan School for the Hearing Impaired, 52, Xinyi Rd., Xinhua Township, Tainan County 712, Taiwan,R.O.C.
wolla98@gmail.com
[4] Graduate Institute of Assistive Technology, NUTN, 33, Sec. 2, Shu-Lin St., Tainan 700, Taiwan, R.O.C.
genius1106@hotmail.comm
[5] Municipal Jinsyue Elementary School, 47, Nanning St, Tainan 700, Taiwan, R.O.C.
{Kennedy0845,elixir416,swim-s}@yahoo.com.tw
[6] Tainan Municipal Heshun Elementary School, 5, Lane 178, Sec. 5, Anhe Rd.Tainan, Taiwan, R.O.C.

Abstract. This study attempts to make use of interactive whiteboard as an interface for children with learning disabilities. Using flash software to design teaching materials with the assistance of interactive whiteboard could also be developed in the children's learning. When detected by a wiimote and infrared light device, corresponding information appears on a screen to increase the interaction aimed at children with learning disabilities by adopting an enhanced intuitive learning method. This study is divided into training and testing steps. The training step allows researchers involved in special education to acquire low-cost interactive whiteboard skills and to develop a unit course for children with learning disabilities. The participants are children with disabilities in the testing step. In this study, the application of technological innovations relies upon user-interface design, which facilitates users' control ability and interaction with an innovation, to convert the technical capabilities into a usable and friendly teaching material.

Keywords: interactive, assistive technology, teaching materials, children, wiimote.

1 Introduction

The study utilizes wiimote equipments of assistive technology and applications of flash software so that teachers teaching for normal or resource classes in kindergarten

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 558–566, 2011.

and elementary school enable to improve students' learning interest, especially for students who have difficulty following the scheduled learning progress [1].Computer-based instruction is widely used in special education. Special education teachers sometimes may have trouble teaching their students to use different computer tools[2],and advances in computer technology replacing translation of traditional paper questionnaires with novel display versions[3].Through the assistance of assistive technology and the demonstration of multimedia design, teachers have enough ability to produce the learning materials of custom-made design in order to support learning disabled students to absorb knowledge[4][5]; user-friendly design is defined as a structural design of an interface[6][7],and teaching interaction procedure is a systematic form of teaching where teachers use to describe their behavior[8].

This study is suitable for students with learning disabilities in resource class and children in kindergarten. The research takes pupils as the main body and divides into two part—the design of teaching materials and the demonstration of students' works. Teachers teaching for resource classes in elementary schools can not only participate in this study directly but also perform the outcome of the study to students right away. Because this device is easy to use, what participants have to do is to operate infrared light, a prototype which relies on the desire of patients to design. In addition, the apparatus can adjust the size and scope of projection. This kind of customized design is low-cost and less than US 40; therefore, children are also able to complete the whole study at home. The Interactive whiteboard technology consists of a large board connected to both a ceiling-mounted projector and a laptop computer that can connect to the internet via the school's wireless network. The projector displays the image from the computer screen on the board. The computer can then be controlled by touching the board with a wireless IR device that functions much like a PC mouse [9]. Interactive whiteboard systems comprise a computer linked to a data projector and a large touch-sensitive electronic board displaying the projected image. Children or teachers can manipulate objects on the screen directly by IR device[10].

Through the continuous progress on information media, functions of interface become more complicated day by day. The purpose of this research is to help children to learn with new teaching materials. Because intuition is an operating feature of assistive technology, it is unnecessary for operators to use complicated input methods, such as using a mouse or keyboard. Therefore, it is easier for children to use interface, to have realer feelings, and to experience a new interactive way. The study proceeds from case study and develops into operating interface and assistive equipments. Under practical exercises and operating processes of teachers and students from kindergartens and elementary schools, the research can observe the obvious differences between the traditional and the custom-made design of teaching material when displaying students' works. Students also very enjoy custom-made teaching materials. Therefore, according to the outcome of the study, the study is successful. Information and communication technology is a powerful tool for learning, which has prominent influences on helping teachers explain difficult concepts, giving access to a huge range of examples and resources, and inducing pupils to engage in learning easily[11].

HCI exists at the junction of computing sciences, design arts, and social sciences. Human-computer interaction is a discipline concerned with the design, evaluation and implementation of interactive computing systems for human use[12]. The goal of interaction design subfield within HCI is to improve the experience for students with learning disabilities of direct interaction. Computer-mediated communication facilitates the understanding of communication patterns, forms, functions and subtexts, which can in turn engender an understanding of how to derive meanings within such context[13]; patterns or figures are increasingly being used not just in education but also in many other areas such as software engineering, engineering and business management, and are also frequently advocated for teaching human–computer interaction principles[14].

The interfacial design of assistive teaching materials makes use of interactive whiteboard. With one mouse click corresponding sound then appears, which enables to help students learn according to their own preference.In studies related to the interactive whiteboard, it is indicated that the interactive whiteboard technology has the potential of supporting teaching and learning[15].This assistive technology of interactive design describes a new design for teaching materials developed in the frame of a research project supported by information interface tools[16]. The introduction of basic education or special education-based computer aided tools in the routine development process of education is truly important.The display of multimedia teaching materials relies on the operation of a mouse; however, for students with learning disabilities, learning how to use a mouse becomes a learning burden. The study makes use of wiimote and interactive whiteboard. The application of an infrared emitter is similar to that of a stylus of a tablet computer. The design of learning interfaces adopts an intuitive design while the design of teaching materials adopts flash software to invent interesting animations in order to raise students learning motives. In addition, students not only can truly experience the vivid teaching materials but also are impressed by them.

2 Method

Thanks Johnny Chung Lee[17] who creates a program about low-cost multi-point interactive whiteboards using the wiimote and shares freely on his website tp://johnnylee.net/projects/wii/. Wiimote is the controller of wii. The fore part of wiimote is equipped with infrared camera which enables to receive infrared light while its interior provides Bluetooth for communication purpose; that is to say, it is available to connect with other computers possessing by Bluetooth apparatus. Therefore, the study is able to integrate the outcome of flash and to develop teaching materials for children.

2.1 Equipment

The theory that the study applies is making use of Bluetooth to connect a computer and a wiimote. By means of infrared emitter, wiimote can track the accurate launching location of infrared light and then generates a simple whiteboard. According to the course design of participating teachers, image projection in the study is divided into two categories.

First category: A projector shows images on the wall or the projector screen. And let the wiimote point to projection position. Based on the requirement of students, the range of exhibition can be adjusted. For students with deep short sight, for instance, we can make use of amplification displaying a larger size in order to improve inconvenience for short-sighted students because the presentation common computer monitors displaying is too small, as Fig. 1.left illustrated. Second category: Let wiimote point to a computer monitor which originally merely has the display function. With the assistance of wiimote, the computer monitor thus becomes a touch screen, as Fig. 1.right illustrated.

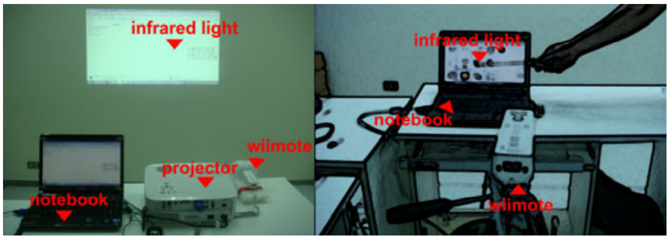


Fig. 1. Display how to use this assistive technology

The display mode of these two methods is an intuitive learning tool. The operator not needs to use a mouse as a tool, especially for children with learning disabilities who therefore do not have to learn the use-pattern of the mouse.Students will raise their interest toward learning and lessen the generation of frustration during their learning processes.

An interactive whiteboard is an interactive display interface that connects to a computer and a projector. A projector projects onto a board's surface that children can control the display by using a pen or other devices, but it cost so expensive; teachers could not use them in many place. Since the wiimote can track sources of infrared light, it can track infrared light led in the tip. In this research, teaching materials are designed by flash software and use interactive whiteboard to develop children's abilities. According to the requirement and preference of students, participating junior and elementary teachers in the study proceed to devise and adjust the design of infrared emitter.

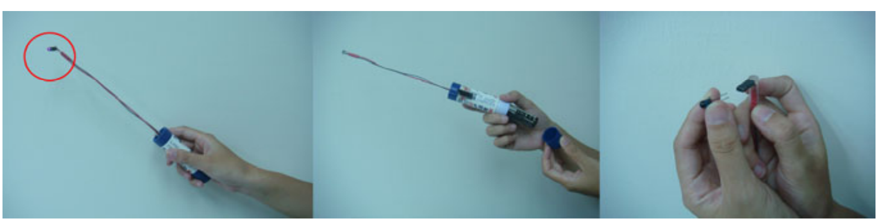


Fig. 2. Infrared emitter design 1

Fig. 2. is designed by teacher Te-Hsiung Chen. Because children with disabilities are generally incompetent at manipulating operational tools and their action

sensibilities are also not nimble enough, teacher Chen therefore makes use of micro-switch to launch infrared emitter. In addition, in the aspect of infrared emitter, in order to prevent wiimote receiver from the interruption of the location of the body resulting from the left or right hand operation of the operator, teacher Chen determines to extend launching section. This improving design creates distance between hand and launching section, and induced effect is better than before.

Fig. 3. is the design of teacher Kenendy Huang. Teacher Huang considers children are not good at holding a pen, so creating a ring style of infrared emitter; therefore, children can just wear the ring on their finger. Furthermore, teacher Huang also devises infrared emitter in a glove form. The tool is in a startup state when children's palms press on the wall, which seems like the machine responds once their hands only touch the wall. It is a certainly humanized design for children to absorb knowledge because it is unnecessary to teach them how to operate learning tools.

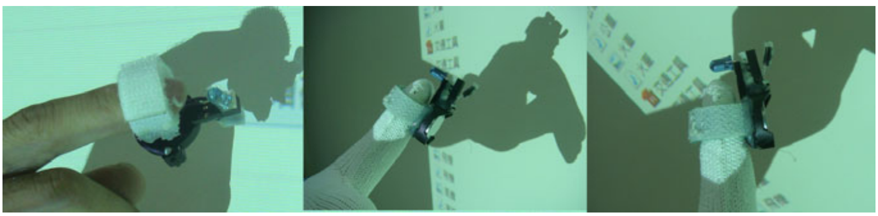


Fig. 3. Infrared emitter design 2

Figure 4 is designed by teacher Chia-Pei Liu. The shape of toy gun is the sample when teacher Liu attempts to design the shape of infrared emitter so as to evoke children's curiosity because they will desire to hold the toy gun to play. As soon as children press the power switch of the toy gun, the screen will make a difference, which not only incite their interest but also stimulate their motivation of learning.

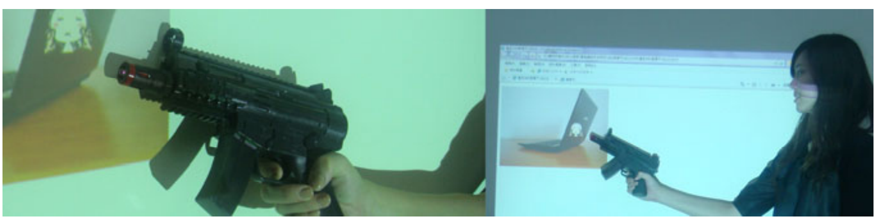


Fig. 4. Infrared emitter design 3

2.2 Participants

This research starts with the design of digitized teaching materials, and with the application of low-cost interactive whiteboard, therefore, this research is accomplished by carrying out two levels. The first is to select teachers from elementary and junior high schools, and teach them how to use flash software. The teachers will design interactive teaching materials, according to the need and age of the students they teach in their schools. Firstly, the researcher demonstrates how to use flash to design teaching materials. Then participatory teachers propose directions of their courses, and begin to design their teaching materials. The second level is that teachers set the teaching materials they designed in the schools, and let students operate digital teaching materials practically. At the same time teachers have to observe the children during their operating process, and record their feedbacks.

3 Case Study

Children always need repeat training and students with learning disabilities have difficulty concentrating on teaching materials; this study therefore attempts to introduce the concept of custom-made learning for teaching materials. The learning materials are customized for students themselves so as to improve their learning interest. The range of custom-made application is considerably extensive, including attire, architecture, medical care, rehabilitative instrument, etc. This study begins from assistive teaching, which provides diverse learning methods in the design of teaching materials. Using infrared emitter and interactive whiteboard can link to corresponding information, which is able to increase the attraction and intimacy of teaching materials. Here are two cases and explanations of this study.

3.1 Case 1

Case 1 is the teaching of Chinese characters. The teaching material of Chinese characters is divided into categories, including stationery, fruit, transportation, sport, furniture, animal and cooker, and some other Chinese phrases being used often in our daily life. Take the category of fruit for instances, teacher Wu puts vocabularies like “watermelon”, “strawberry”, “pineapple” forward as daily words.

Fig. 5.left is displayed in resource class. In the original teaching process, the general teaching method of word recognition is difficult for the student. Therefore, teacher Wu adopts multi-media teaching method of Chinese characters, which is different from the traditional style of learning, and uses pictures and sounds to help him memorize phrases. In this teaching and learning activity, the teacher introduces and demonstrates how to operate first, and then the student operates it on their own.

Because these students are very interested in computers, they intend to stay longer when seeing the pictures that interest them. Fig. 5.right shows teacher demonstrates how to put the appliance of infrared emitter on the word, the corresponding picture and pronounce of the word. This design of digital teaching material offers an intuitive way of learning for students when they learn new words.

Fig. 5. Course of Chinese

3.2 Case 2

Case 2 is to learn the concepts of third dimension. Students of teacher Chou, Shu-Ying have poor concepts of flexible, therefore, teacher Chou designs interactive effect with the application of objects that are often seen by children, and with the assistance of flash.

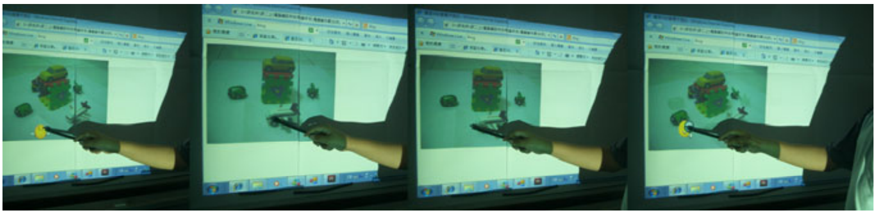


Fig. 6. The course of space training

The teacher makes use of the mighty interactive speciality of wiimote to design this lesson. The teacher is able to understand students' cognition of left and right, and concepts of extensity. Students show that it is very novel to use wiimote, as show in Fig. 6.

4 Conclusion

Digitized teaching material is also a trend, owning to its feature of interactive speciality that possesses immediate feedbacks. Moreover, with the support of wiimote, interactive whiteboard will be able to offer teachers and more students different opportunities to learn. For children, they can get immediate feedbacks because of the interactive speciality, which increases their motivation of learning.

Interface design can be mixed in the development of children, through the interactive speciality of information equipments, plus teachers' research and development for the design of teaching materials. Especially for special students, through the application of appliances, limits are decreased. Special thanks many teachers from elementary and junior high schools, who, for the arrangements of courses, have consideration for students' individual differences. Besides basic teaching aids, they propose many ideas and practice them in class.

As for teachers, in the design of teaching materials, this digitized learning process is certainly friendly. Therefore, it is easy to popularize digitized teaching materials to other teachers so as to accomplish teaching materials together.Interacting with computers reflects the movement within the interactive whiteboard. Some of the topics in designed teaching units make interactive interface design of assistive technology as their learning goals. Technology-based learning focuses on content learning in order to explore more user-centred and collaborative approaches in learning aspect. The importance of information and communication skills of interface design in the future has been asserted to support the development of these skills and tools in schools. Students need to possess the capability to use technology and information. Teachers therefore can integrate information technology into teaching activities. To efficaciously promote teaching effectiveness is the primary direction and the main goal for the school and teachers. Teaching materials as well as teaching methods have been continuously developed and innovated, especially in technological assistive tools.

Living in the era with the advance of technological assistive tools, students with special needs have a more convenient life when absorbing knowledge. By virtue of innovative ideas of serious teachers, students are able to receive a better opportunity to acquire knowledge. The student is concentrated most of the time during the operating process. Because children are very interested in computers, using multi-media to teach and learn is easier to give rise to the student's motivation of learning than general activities which use books or pencils and papers. And with pictures and sounds as hints, the scaffold of learning is built up. From the children used the interactive teaching materials, some children feel that the learning processes are no longer so difficult.

This multi-media teaching material is presented through multi-senses, which makes the content more vivid, and adds interest to the process of learning. Students can also choose learning method (visual approach, auditory approach) which is suitable for them to learn, this impression deepens more, and the effect of learning is better; moreover, it saves time. Furthermore, thanks to application information interface, students are inclined to absorb knowledge actively and aggressively. They can learn independently with no need of other's assistance. Interactive media interface for children with learning disabilities has become an important and helpful computer-aided design for teaching materials. The interface arrangement also gives students assistance when they attempt to learn different units. There is a close link among assistive technology, special education and communication design research as well as studies that examine how interactive design of teaching materials can influence learning. Human–computer interaction is a valuable issue for learning disabled students in the future.

Acknowledgments. Special thanks National Tainan school for the hearing impaired, this work was partially supported by the National Science Council, Taiwan, under the Grant No.97-2410-H-024-018-,98-2410-H-024-018.

References

1. Klatt, B.A., Goyal, N., Austin, M.S., Hozack, W.J.: Custom-fit Total Knee Arthroplasty Results in Malalignment. J. Arthroplasty 23, 26–29 (2008)
2. Shimizu, H., McDonough, C.S.: Programmed Instruction to Teach Pointing with a Computer Mouse in Preschoolers with Developmental Disabilities. Research in Developmental Disabilities 27, 175–189 (2006)
3. Hung, P.H., Lin, C.Y., Lu, C.C., Chang, Y.Y.: Development and Application of Online Assessment for Experimental Debugging Performance. In: The 7th conference of the international test commission, p. 102 (2010)
4. Kawate, K., Ohneda, Y., Ohmura, T., Yajima, H., Sugimoto, K., Takakura, Y.: Computed Tomography–based Custom-made Stem for Dysplastic Hips in Japanese Patients. J. Arthroplasty 24, 65–70 (2009)

5. Lin, C.Y., Hung, P.H., Lin, J.Y., Lun, H.C.: Augmented Reality-based Assistive Technology for Handicapped Children. Key Engineering Materials 439–440, 1253—1258 (2010)
6. Cho, V., Cheng, T.C.E., Lai, W.M.J.: The Role of Perceived User-interface Design in Continued Usage Intention of Self-paced e-learning Tools. Computers & Education 53, 216–227 (2009)
7. Kim, Y.J.: The Effects of Task Complexity on Learner–learner Interaction. System 37, 254–268 (2009)
8. Leaf, J.B., Dotson, W.H., Oppeneheim, M.L., Sheldon, J.B., Sherman, J.A.: The Effectiveness of a Group Teaching Interaction Procedure for Teaching Social Skills to Young Children with a Pervasive Developmental Disorder. Research in Autism Spectrum Disorders 4, 186–198 (2010)
9. Lopez, O.S.: The Digital Learning Classroom: Improving English Language Learners Academic Success in Mathematics and Reading Using Interactive Whiteboard Technology. Computers & Education 54, 901–915 (2010)
10. Warwick, P., Mercer, N., Kershner, R., Staarman, J.K.: In the Mind and in the Technology: The Vicarious Presence of the Teacher in Pupil's Learning of Science in Collaborative Group Activity at the Interactive Whiteboard. Computers & Education 55, 350–362 (2010)
11. Waite, S.J., Wheeler, S., Bromfield, C.: Our Flexible Friend: The Implications of Individual Differences for Information Technology Teaching. Computers & Education 48, 80–99 (2007)
12. Rosinski, P., Squire, M., Strange, B.: Human-computer Interaction, Interface Design, and Composition Pedagogy. Computers and Composition 26, 149–163 (2009)
13. Bower, M., Hedberg, J.G.: A Quantitative Multimodal Discourse Analysis of Teaching and Learning in a Web-conferencing Environment–the Efficacy of Student-centred Learning Designs. Computers & Education 54, 462–478 (2010)
14. Kotzé, P., Renaud, K., Biljon, J.V.: Don't Do This – Pitfalls in Using Anti-patterns in Teaching Human–computer Interaction Principles. Computers & Education 50, 979–1008 (2008)
15. Tataroglu, B., Erduran, A.: Examining Students Attitudes and Views Towards Usage an Interactive Whiteboard in Mathematics Lessons. Procedia Social and Behavioral Sciences 2, 2533–2538 (2010)
16. Lin, C.Y., Hung, P.H., Wang, L.C., Lin, C.C.: Reducing Cognitive Load Through Virtual Environments Among Hearing-impaired Students. In: PACCS 2010, pp. 183–186 (2010)
17. Lee, J.C.: Hacking the Nintendo Wii Remote. Pervasive Computing, 39–45 (2008)

Integration of a Spanish-to-LSE* Machine Translation System into an E-learning Platform

Fernando López-Colino, Javier Tejedor, Jordi Porta, and José Colás

Human Computer Technology Laboratory, Escuela Politécnica Superior,
Francisco Tomás y Valiente 11, 28049 Madrid, Spain
{fj.lopez,javier.tejedor,jordi.porta,jose.colas}@uam.es

Abstract. This paper presents the first results of the integration of a Spanish-to-LSE Machine Translation (MT) system into an e-learning platform. Most e-learning platforms provide speech-based contents, which makes them inaccessible to the Deaf. To solve this issue, we have developed a MT system that translates Spanish speech-based contents into LSE.

To test our MT system, we have integrated it into an e-learning tool. The e-learning tool sends the audio to our platform. The platform sends back the subtitles and a video stream with the signed translation to the e-learning tool.

Preliminary results, evaluating the sign language synthesis module, show an isolated sign recognition accuracy of 97%. The sentence recognition accuracy was of 93%.

Keywords: Machine Translation, Sign Language, LSE, Accessibility.

1 Introduction

During the last decade there has been an increasing interest in machine translation research from an oral language to a signed language. To obtain a complete system providing this functionality, it is necessary to integrate the results from three different research areas: Automatic Speech Recognition (ASR), Machine Translation (MT) and Sign Language (SL) Synthesis. Although the literature provides several examples of working prototypes, it is hard to find examples of these systems applied to education with representative results (see next section). Different statistics report that the 47% of the Spanish Deaf people have not finished any degree or even are illiterate, compared to the 21% of the whole Spanish population (INE[1] 1999 and MEC[2] 2000/2001). Such data claim that Deaf people have many drawbacks to achieve the educational, social and cultural levels of the general society. It is possible to use MT systems to contribute to alleviate this social problem.

---

* Spanish Sign Language.

1 National Statistics Institute.

2 Ministry of Education.

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 567–576, 2011.

2 Related Work

We have stated that a system that converts speech into a signed message requires using three different research areas: ASR, MT and SL synthesis. We will briefly describe recent works from the last two. Finally we will comment some MT translation systems that have been applied to the e-learning.

In 2003, a technical report about MT systems applied to ASL [1] concluded that existing systems were just work in process or simple demonstrators. In the recent years, different authors have reported improvements to their systems [2, 3] and applications to other languages [4, 5]. However, we believe this research direction has not reached the level of development achieved in other areas like MT between oral languages. These systems do not fulfill the required functionality to provide a real useful service to the Deaf community.

Although the SL-related MT systems do not provide complete functionality, the literature provides some examples of applications that enhance e-learning systems by means of synthetic signed contents. Some systems use the existing technology to provide e-learning tools to learn a SL [6, 7]. Although these systems only allow the visualization of signed contents, there are others that also allow the interaction of the student and provide feedback about his/her signing ability [8]. Particularly, for Spanish Sign Language (LSE), a tool was recently presented that aims at helping the study of new vocabulary [9]. However, our aim is to improve the experience of deaf users when using e-learning tools. The application of previous MT systems to e-learning tools allows deaf users the access to the speech-based contents in these applications. Some of these systems have only focused in a closed domain, such as mathematics [10, 11]. Finally, it is remarkable the efforts of the Greek researchers for generating an e-learning platform for Greek Sign Language [12, 13, 14].

3 System Overview

ISABEL [15] is a multipurpose video streaming platform, which allows easy modification of the viewer's layout and the management of video streams. We have used the ISABEL system due to its e-learning capabilities. The speakers' audio stream is both transmitted to every student and to our MT system (named MaTSyLSE, depicted in Fig. 1). The platform receives the audio stream and returns to the ISABEL platform both the transcribed text, used for subtitling, and the video stream of the signed translation. The ISABEL platform considers this signed video stream as another participant in the conference management and redirects the stream to the users that request for it. Fig. 1 shows both the architecture of the system and the final GUI of the working system.

3.1 ASR Module

The audio obtained from the video stream is transcribed. In so doing, the ASR module consists of a speech recognizer built from Julius tool [16]. It is a large vocabulary continuous speech recognition system that works on real time to produce the most

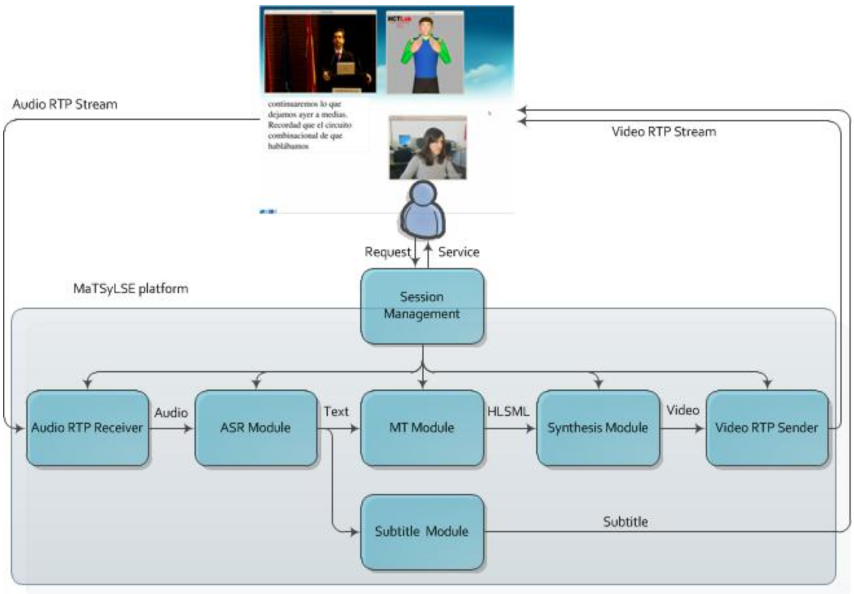


Fig. 1. GUI and diagram of the architecture

likely sequence of words corresponding to the input speech signal. It uses Hidden Markov Models to represent the acoustic space, a bigram language model and contains a vocabulary of 5K words. The output of this module is the transcription of the speech.

3.2 MT Module

When translating between similar and intelligible languages, as occurs between many Romance languages, quality and massive translation is possible to some extent. However, translation between typologically unrelated languages, as Spanish and LSE are, has another purpose: to break the so-called language barrier that makes communication between individuals difficult, if not impossible.

The MT module implements a transfer-based approach to translation. The main phases in this architecture are analysis, transfer and generation. In the analysis phase, a Spanish text like *Si llueve mañana no iré a trabajar* ("if it rains tomorrow I won't go to work") is first analyzed morphologically and then syntactically to obtain a constituency tree like the tree on the left side of Fig. 2. Then a dependency structure, depicted in the right side of Fig. 2, is built up from the constituency structure as an intermediary result for transfer. A dependency structure is a more abstract structure where constituency and word order have been removed and only functional relations between words are represented. This structure is transferred to LSE. During the process, some nodes representing words are removed, as is the case of definite articles, which have no correlate in LSE. On the other hand, other information can appear in the new structure, like the non-overt subjects in Spanish. Note that in the

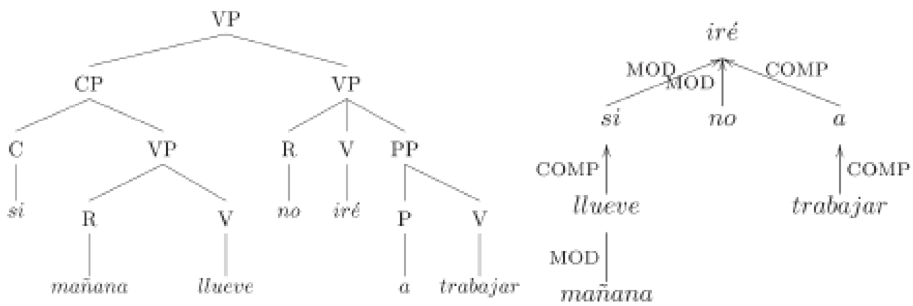


Fig. 2. Constituency tree (left) and its corresponding dependency tree (right)

transferred dependency structure of Fig. 3, a first person pronoun subject *yo* ("I") has been made explicit. Generation is formulated as a constraint problem. A precedence graph is built from the information about the relative order between a head and its dependent or between pairs of dependents with the same head. As an example: suppose a noun phrase with the structure "N_1 *de* N_2" (N_1 of N_2). When translating this expression to LSE, sometimes N_1 precedes N_2, as in *hermano de Luis* ("Luis' brother"), which is translated as LUIS HERMANO because *hermano* denotes a permanent relationship. In other cases the correct order is N_2 N_1, as in *mesa de estudio* ("study table"), because *estudio* is the function or purpose of *mesa*, and therefore, it is translated as MESA ESTUDIO. An algorithm based on topological sort produces the linear surface order in LSE. Consequently, the dependency graph of Fig. 3 will generate the LSE glossed sentence EJEMPLO / MAÑANA LLOVER / YO TRABAJAR IR NO ("EXAMPLE / TOMORROW RAIN / I TO-WORK TO-GO NO"). Finally, as a last step before the synthesis, signs are morphologically realized according to their description.

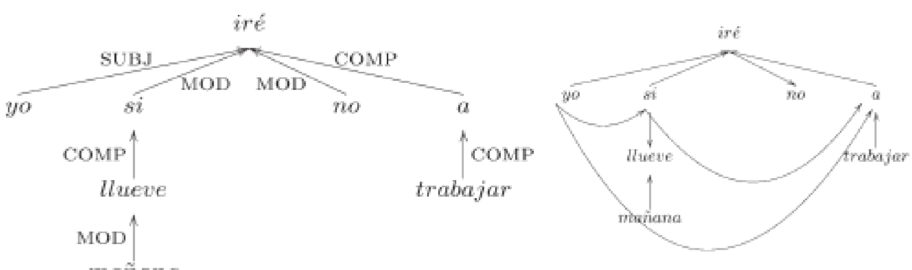


Fig. 3. Transferred dependency tree (left) and its corresponding precedency graph (right)

A wide-coverage unification-based grammar for Spanish has been implemented for the analysis phase. The linguistic phenomena dealt by the grammar include complementation, adjunction, pronominalization, relativization, etc. Grammar is inspired by the lexicalist approach of theories as the Head-driven Phrase Structure Grammar (HPSG) [17], where the lexicon is the locus of almost all the grammatical information and schemas can be implemented as context-free rules. However, for convenience, the accounting for complex groups of specifiers, modification or coordination is not expressed in the lexicon but in rules modeled under X-bar theory [18]. Apart from schemas, the grammar also includes lexical rules. A lexical rule is an input-output rule mainly used, in our case, to alter the argument structure of a lexical item. Therefore, phenomena like subject pro-drop, complement extraction, complement deletion, animated object marking or pronominalization, to name a few, are dealt with lexical rules. The parser is an active chart-based agenda-driven Earley's parser [19] implemented in Prolog but makes use of the language processing services implemented in FreeLing [20] for tokenization, named-entity recognition and PoS-tagging.

Parsing robustness is achieved through three techniques: relaxation, weighted phrase structure rules and a mechanism for the assembly of partial parses. Relaxation here is used mainly to handle frequently seen incorrect assignments of the PoS-tagger by means of casting rules. With these rules, for example, determiners and pronouns recover their original lexical determiner-pronoun ambiguity, so that the parser can proceed exploring, if necessary, a wider search space for a complete analysis. Semiring parsing [21] is a framework in which deductive systems for parsing can be extended to manipulate weights. Semirings generalize the representation and computation of weights for weighted languages. Examples of semirings are the counting semiring, defined as $(\mathbb{N}, +, \times, 0, 1)$, which is used to compute the number of derivations in grammars with unit rules, or the Viterbi semiring $([0..1], \max, \times, 0, 1)$, which is used to compute the probability of the best derivation for a probabilistic grammar. Here, we use the tropical semiring defined as $(\mathbb{N}, \min, +, -\infty, 0)$ to compute the minimum derivation weight of a parse tree. This decision pretends, as in other preference-based parsers, to favor structurally simple analyses. Speakers make mistakes, correct themselves during speaking, produce false starts and use ungrammatical constructions. Several methods were proposed for robust analysis. Partial parse is defined in [22] as the set of intermediate passive edges of the chart whose spans are not overlapping and cover the entire input sequence. For selecting the best sequence of partial parses, they apply the shortest-path algorithm to maximal projections, i.e., complete phrases. These phrasal edges are assigned a weight of one whilst the weight of lexical edges is two. Finally, solutions are re-ranked using a scoring function. Shortest-path algorithm is an adaptation of the general algorithm to the case of direct acyclic graphs. The algorithm runs with a worst-case time linear in the number of edges and vertices. In [23], different methods for partial parse selection are presented and evaluated for HPSG parsers producing much deeper analyses than the parser presented here. The criteria examined for selecting partial parsers include longest edges and shortest paths using several weight estimation functions based on probabilities derived from treebanks. Our heuristic approach here is simple: it uses the derivation cost and is based on the intuition that very complex partial analyses should be rejected as well as the simplest ones. So, arcs $[i,j,w]$ in the chart are assigned a distance $w/K(j-i)$ where K is a factor that controls the complexity of the analyses and can be estimated from the weights and lengths of partial analyses within complete analyzed utterances.

3.3 Synthesis Module

The synthesis module [24, 25] receives the translated message from the MT module using the HLSML notation. This notation allows representing glosses, phonetic definitions, morphologic variations, classifier predicates, fingerspelling sequences, etc. Although the synthesizer can manage all of these inputs, the first version of the MaTSyLSE platform is based only in glosses.

The system parses the HLSML input by obtaining the sequence of signs in the message. The phonetic definition of each sign is stored in a relational database. This definition is used for the generation of the animation sequence of the whole sentence. This animation is rendered using a real time rendering API [26] and the resulting video is streamed to the client.

3.4 Subtitle Module

The subtitle module is a simple text processing element. The ASR module generates different length transcriptions, from one word to a complete sentence. These transcriptions cannot be directly redirected to the users because they may collapse the subtitle window, making its reading impossible.

The subtitle module buffers the transcriptions and sends them to the user controlling the length and the pauses between consecutive text chunks. These chunks' average length is 18 characters sent every 700 ms; the final length is adjusted so no word is hyphenated.

4 Evaluation

The initial evaluations of the system have been focused on the evaluation of the user GUI required for deaf users and the avatar's signing performance. Evaluation of the ASR and the MT modules will be presented in a forthcoming work.

For the evaluations, a group of ten deaf native signers have been selected; four of these users work at the LSE linguistic department of the FCNSE. The ages of this group ranges from 20 to 62 years old, with an average of 37.1 years. All of the users in this group have a medium comprehension level for reading Spanish. The evaluations contained written instructions that users could understand.

4.1 Definition of the GUI

The first evaluation was oriented to the definition of the layout of ISABEL for this kind of application. We proposed a set of five different window distributions with different sizes. The unanimous selection was the set with the largest avatar window size. This is understandable because the signer users will omit the speaker's window and they prefer as much detail as possible in the avatar's window.

The preferred layout is depicted in Fig. 4. The final avatar dimension is 600×600 pixels. Although signing users preferred a larger avatar size, a compromise between the screen size and the transmission bandwidth was set. There is also a subtitle window for non-signer users. The font of the window can be adjusted by the user.

Fig. 4. Screenshot of the application

4.2 Evaluation of the Synthesis

Preliminary evaluations have measured the performance of the Synthesis module. These evaluations measured both the isolated sign recognition rates and the sentence understanding. The previously described group of native LSE signers performed both the evaluations.

We provided the users with the different URL to access the different web forms, which contained the questionnaires. These forms contained text that explained the experiment to the users and instructed them how to proceed.

The first questionnaire contained a set of twenty randomly-selected signs. Each question showed a video of the avatar signing each sign. After the visualization of each video, the users had to write the Spanish word that corresponded to that sign. For some of the selected signs there are multiple correct answers (e.g., LSE's sign for the noun "water" and the verb "to drink" is the same). We considered as correct all of them. The result of this evaluation presents a rate of 97% correctly recognized signs.

The second questionnaire contained a set of ten sentences in LSE. These sentences were proposed by an LSE interpreter and synthesized using the synthesis module. Hence, we ensured that these sentences were grammatically correct. The procedure of this second evaluation was similar to the one stated above. After the visualization of the sentences, the users had to choose one out of five different possible sentences. These sentences were different variations to the one represented by the avatar. The result of this second evaluation shows a 93% of sentences correctly recognized.

After the evaluations, we asked the users for their opinion. One of their concerns was the relatively inexpressiveness of the avatar, because the facial expression was only modified due to phonology and syntax requirements. Another concern was related to the intelligibility of some of the signs used in the sentences. Some users misunderstood few signs and selected a wrong answer.

5 Conclusion

This paper has presented a Spanish-to-LSE machine translation system integrated within ISABEL, an e-learning platform. On one hand, the system allows deaf people to access the speech-based information by means of a synthetic signed message using MT techniques. On the other hand, the system also provides subtitles for people with low degree of hearing loss.

As far as experimentation is concerned, this work has presented an initial evaluation of the system. We have focused on evaluating the visual layout of the accessible version of the e-learning tool. We have also provided initial evaluations of our SL synthesis module.

As future work, we will extend these evaluations, measuring the performance of the ASR and MT modules. We will also extend our research towards different educational domains (math, science, etc.) and towards different age groups (university, high school, primary).

The ISABEL platform is a general purpose videoconference tool. We have proposed using it for e-learning. However, we will also explore other environments such as VoIP online MT translation, broadcast news MT translation, etc. which can get benefit from our MT system.

Acknowledgements. Authors would like to acknowledge the FPU-UAM grant program for its financial support. Authors are grateful to the FCNSE linguistic department for sharing their knowledge in LSE and performing the evaluations. Many thanks go to María Chulvi and Benjamín Nogal for providing help during the implementation of this system. This work was partially supported by the Telefónica Móviles España S.A. project number 10-047158-TE-Ed-01-1.

References

1. Huenerfauth, M.P.: A Survey and Critique of American Sign Language Natural Language Generation and Machine Translation systems. Tech. Rep. MS-CIS-03-32. Computer and Information Science, University of Pennsylvania (2003)
2. Jennings, V., Elliot, R., Kennaway, R., Glauert, J.: Requirements for a Signing Avatar. In: Proceedings of LREC, 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies, Valletta, Malta, pp. 133–136 (2010)
3. Kennaway, R., Glauert, J., Zwitserlood, I.: Providing signed content on the internet by synthesized animation. ACM Transactions on Computer-Human Interaction 14(3), 1–29 (2007)
4. Dasgupta, T., Basu, A.: Prototype machine translation system from text-to-indian sign language. In: Proceedings of the 13th International Conference on Intelligent User Interfaces (IUI 2008), pp. 313–316. ACM, New York (2008)

5. San-Segundo, R., Barra, R., Córdoba, R., D'Haro, L.F., Fernández, F., Ferreiros, J., Lucas, J.M., Macías-Guarasa, J., Montero, J.M., Pardo, J.M.: Speech to sign language translation system for Spanish. Speech Communication 50(11-12), 1009–1020 (2008)
6. Agravat, B.B.: Inputted text to animated sign language, interactive interface, a self-learning with fun. In: ACM SIGGRAPH Educators Program. ACM, NewYork (2007)
7. Ohene-Djan, J., Naqvi, S.: An Adaptive WWW-based System to Teach British Sign Language. In: Fifth IEEE International Conference on Advanced Learning Technologies, pp. 127–129 (2005)
8. Aran, O., Ari, I., Akarun, L., Sankur, B., Benoit, A., Caplier, A., Campr, P., Carrillo, A., Fanard, F.X.: SignTutor: An Interactive System for Sign Language Tutoring. IEEE Multimedia 16(1), 81–93 (2009)
9. López-Colino, F., Colás, J.: New interactive tool for teaching new vocabulary in LSE. In: Proceedings of III National Conference on Spanish Sign Language. CNSE-UNED, Madrid (2009)
10. Adamo-Villani, N., Doublestein, J., Martin, Z.: The Mathsigner: An Interactive Learning Tool for American Sign Language K-3 Mathematics. In: International Conference on Information Visualisation, pp. 713–716 (2004)
11. Adamo-Villani, N., Carpenter, E., Arns, L.: An immersive virtual environment for learning sign language mathematics. In: ACM SIGGRAPH Educators Program. ACM, New York (2006)
12. Efthimiou, E., Sapountzaki, G., Karpouzis, K., Fotinea, S.E.: Developing an e-Learning Platform for the Greek Sign Language. In: Miesenberger, K., Klaus, J., Zagler, W.L., Burger, D. (eds.) ICCHP 2004. LNCS, vol. 3118, pp. 1107–1113. Springer, Heidelberg (2004)
13. Karpouzis, K., Caridakis, G., Fotinea, S.E., Efthimiou, E.: Educational resources and implementation of a Greek sign language synthesis architecture. Computers & Education 49(1), 54–74 (2007)
14. Sapountzaki, G., Efthimiou, E., Karpouzis, K., Kourbetis, V.: Open-ended resources in greek sign language: Development of an e-learning platform. In: Proceedings of the Workshop on the Representation and Processing of Sign Languages, LREC 2004, Lisbon (2004)
15. Quemada, J., de Miguel, T., Pavón, S., Huecas, G., Robles, T., Salvachúa, J., Acosta Ortiz, D., Sirvent, V., Escribano, F., Sedano, J.: Isabel: An Application for real time Collaboration with a Fexible Floor Control. In: Proceedings of the 1st International Conference on Collaborative Computing: Networking, Applications and Worksharing. San Jose, CA, USA (2005)
16. Lee, A., Kawahara, T., Shikano, K.: Julius – an Open Source Real-Time Large Vocabulary Recognition Engine. In: Proceedings of Eurospeech, pp. 1691–1694 (2001)
17. Pollard, C., Sag, I.A.: Head-Driven Phrase Structure Grammar. Studies in Contemporary Linguistics. University of Chicago Press, Chicago (1994)
18. Jackendoff, R.: X-bar Syntax: A Study of Phrase Structure. MIT Press, Cambridge (1977)
19. Earley, J.: An efficient context-free parsing algorithm. Communications of the ACM 13(2), 94–102 (1970)
20. Atserias, J., Casas, B., Comelles, E., González, M., Padró, Ll., Padró, M.: FreeLing 1.3: Syntactic and semantic services in an open-source NLP library. In: Proc. of LREC (2006)
21. Goodman, J.: Semiring Parsing. Computational Linguistics 25(4), 573–605 (1999)
22. Zhang, Y., Kordoni, V., Fitzgerald, E.: Partial Parse Selection for Robust Deep Processing. In: Proceedings of the Workshop on Deep Linguistic Processing (DeepLP 2007), Prague, Czech Republic, pp. 128–135 (2007)

23. Kasper, W., Kiefer, B., Krieger, H.U., Rupp, C.J., Worm, K.L.: Charting the Depths of Robust Speech Parsing. In: Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics, pp. 405–412. ACL, College Park (1999)
24. López-Colino, F., Colás, J.: Hybrid paradigm for Spanish Sign Language synthesis. Universal Access in the Information Society (in press)
25. López-Colino, F., Porta, J., Colás, J.: Linguistic Principles Applied to the Automatic Synthesis of Spanish Sign Language. Estudios de Lingüística de la Universidad de Alicante 24 (2010)
26. Java Community Process: JSR-184. Mobile 3D Graphics API for J2ME™

Towards Ambient Intelligence in the Classroom

George Margetis[1], Asterios Leonidis[1], Margherita Antona[1], and Constantine Stephanidis[1,2]

[1] Foundation for Research and Technology – Hellas (FORTH)
Institute of Computer Science
N. Plastira 100, Vassilika Vouton,
GR-700 13 Heraklion, Crete, Greece
[2] University of Crete, Department of Computer Science
{gmarget,leonidis,antona,cs}@ics.forth.gr

Abstract. This paper discusses an education-centric approach towards ambient intelligence in the classroom, raising fundamental requirements that should be taken into consideration, in order to efficiently provide genuine students' education enhancement. These requirements are addressed by an integrated architecture for pervasive computing environments, named ClassMATE, which facilitates all necessary mechanisms for context – aware ubiquitous computing in the classroom. Furthermore, a smart classroom prototype, incorporating the ClassMATE's infrastructure, is presented constituting the first test – bed for the study of the educational process in intelligent classrooms.

Keywords: ambient intelligence, education, smart classroom.

1 Introduction

Ambient Intelligence (AmI) has been anticipated by an increasing number of researchers and practitioners worldwide as an emerging field of research and development that is nowadays rapidly gaining wide attention, and in particular in Europe [9]. This wide interest has driven research towards diverse smart environments of everyday life activities. On the other hand, a general trend towards intelligent environments is also signified by the wide embracement of ubiquitous computing and pervasive interaction, which constitute today fundamental factors both in ICT research and industry.

An important everyday life activity that has been already permeated by ICT is that of education, which constitutes a promising domain for AmI, since ambient intelligence can play an important role by increasing students' access to information, enriching the learning environment, allowing students' active learning and collaboration, and enhancing their motivation to learn [5]. However, the majority of current research efforts in the field of AmI mainly focus on smart home or collaborative office environments, and few approaches address the need for an intelligent environment for students' education in the classroom.

An example of intial application of AmI approaches in the classroom environment is the Smart Classroom [17], which facilitates collaborative learning among college

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 577–586, 2011.

students using pervasive computing technology. In more detail, it integrates handheld devices, such as Personal Digital Assistants (PDAs), with fixed computing infrastructures (e.g., PCs, sensors, etc.) in a wireless situation-aware network in the sense that any device can capture different situations in a classroom dynamically to form ad hoc networks to facilitate both student-student and student-instructor collaboration. Features as the aforementioned are considered to be fundamental for developing intelligent environments that aim to augment the educational process in the classroom. However, existing approaches still lack the dimension of environment intelligence, as they require that the teacher decide for whatever is to performed in the classroom, and they also don't take into consideration the individual learner's needs.

This paper presents a systematic approach for ambient intelligence in the classroom, suggesting a set of "intelligent" facilities to enhance the educational process. The key feature that differentiates the presented approach from similar work is the education-centric approach that has been adopted. In more detail, an integrated architecture for pervasive computing environments, named ClassMATE (Classroom Multiplatform Augmented Technology Environment) [13] is described, which monitors the ambient environment and makes context-aware decisions in order to assist (i) the student in conducting learning activities, by simplifying everyday tasks and providing personalized content according to individual needs, and (ii) the teacher with administrative issues by automating common activities. The following sections discuss fundamental issues that should be taken into account by AmI-enabled classroom environments and provide an example of such an environment. The last section of this paper provides conclusions and introduces future work plans.

2 Towards Ambient Intelligence in Classroom

According to Cook et. al [3], any smart environment can be adequately decomposed in four fundamental layers: physical, communication, information and decision. Each layer performs a different role in the environment, facilitating diverse operations and addressing specific requirements. Following this approach, this paper discusses fundamental aspects and raises requirements to be addressed by a smart environment architecture in order to enable ambient intelligence in a classroom.

The analysis presented in this paper is based on the design and deployment of the ClassMATE framework, which was implemented in the context of the AmI Classroom Project of the ICS-FORTH AmI Programme [6]. ClassMATE is depicted in Fig. 1, where Fig. 1(a) illustrates the orchestrating role of ClassMATE in an intelligent classroom, emphasizing its monitoring and coordination infrastructure, while Fig. 1(b) depicts the overall architecture of ClassMATE. ClassMATE aims to provide a robust and open ubiquitous computing framework suitable for a school environment that: (i) provides a context aware classroom orchestration based on information coming from the ambient environment, (ii) addresses heterogeneous interoperability of AmI services and devices, (iii), facilitates synchronous and asynchronous communication, (iv) supports user profiling and behavioral patterns discovery and (v) encapsulates content classification and supports content discovery and filtering.

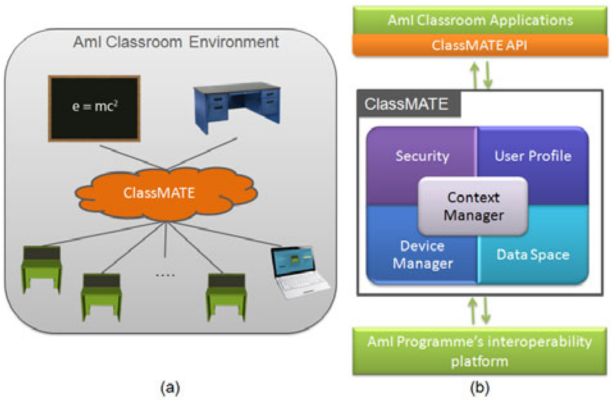


Fig. 1. a) ClassMATE provides a complete monitoring and coordination infrastructure for an AmI Classroom Environment. b) The ClassMATE architecture.

To achieve the above objectives, ClassMATE introduces various modules enclosed either in the (i) Ambient Environment Manager or in the (ii) Content Personalization Manager. The Ambient Environment Manager encloses two modules that monitor the environment and enable contextual awareness, namely the Context and the Device Manager. The Context Manager includes: (i) the Platform Expert that operates as an abstraction layer of the multiplatform environment that provides access to the wide-variety of platform-specific functions in a platform-independent manner, (ii) the ClassMATE Event System that defines a hierarchy with specialized event types forming the ClassMATE event type system and implements the essential mechanisms for their distribution, (iii) the Artifact Director that is a context aware module that orchestrates each artifact, (iv) the Class Orchestrator that controls every aspect of the classroom in a top-level, (v) the State Serialization Manager that manages application's state serialization and deserialization, and finally (vi) the Migration Processor that facilitates the application migration from the current local artifact to a remote node. The Device Manager includes: (i) the Multitouch Device Manager that enables multitouch interaction schemes and (ii) the Book Localizer that determines current context of use (e.g. currently studied course).

The Content Personalization Manager encloses two modules which are responsible for the delivery of personalized educational content based on the current needs of the individual learner: the User Profile and the Data Space Manager. The User Profile collects personal data associated with a specific user (both static and dynamic) and the Data Space Manager provides an abstraction layer between the applications and the physical storage layer and encapsulates a filtering mechanism for personalized content delivery based on user needs and preferences (available through the User Profile).

The next subsections discuss the basic requirements that need to be addressed by smart environments in the smart classroom perspective. Additionally, a short description of how these aspects are addressed by ClassMATE is also included.

2.1 Physical Components and Pervasive Computing

Physical components and pervasive computing constitute the perception ability of smart environments, which is a fundamental prerequisite for their existence. In this perspective, the recent evolution of sensors and actuators has driven to the implementation of sophisticated sensors-actuator networks (e.g., tinyOS [20], Mate [14], Zigbee Alliance [22], etc.), able to communicate either with each other in order to combine raw information gathered from the environment (e.g., user position and velocity, tactile/contact, biochemical metrics, etc.) or with smart environments' ubiquitous computing frameworks in order to transmit this information and initiate further computations.

The physical components and pervasive computing approaches of a smart classroom do not raise any special requirements different from other smart environments, since the information gathered at this level is primitive and eventually similar for any smart place.

In this respect, ClassMATE's intercommunication needs between sensors / actuators and primary services are addressed by a generic services interoperability platform, named FAMINE (FORTH's AMI Network Environment), which has been implemented in the context of the FORTH-ICS AmI Programme, and provides the necessary functionality for the intercommunication and interoperability of heterogeneous services hosted in an AmI Environment.

2.2 Natural Interaction in the Classroom

As stated in [18], natural interaction in smart environments requires the elaboration of new concepts that extend beyond the current desktop and menu driven paradigms. In addition, the key characteristic of these new interaction concepts is the use of physical artifacts as representation and interaction means, seamlessly integrating the physical with the digital world. In this respect, the educational applications of a smart classroom should be extended to a more pervasive nature, ensuring that they are able to be launched and manipulated in any smart artifact of the classroom. For example, a student in the smart classroom may start working on an exercise on his smart desk using a touch screen, transfer his work to the smart interactive blackboard because of his teacher's request and resume it from the point where he left off. The student interacts with each smart artifact in the environment naturally, and is able to transfer the application from one artifact to another seamlessly.

In order for ClassMATE to implement seamless pervasive interaction in the smart classroom, it encapsulates the Application Launcher core module, which is responsible to launch the appropriate application, in a valid state for its current context of use, addressing user's interaction needs according to the classroom artifact that the application has been launched. An application can be launched either directly by the classroom artifact coordination module, as a response to a native FAMiNE (context-oriented) event, or indirectly by its Mime Handler delegate when handling a Mime Command fired by an application. In both cases, for an application to be launched, ClassMATE should incorporate mechanisms to both resolve from the installed applications the preferred one(s), either by name or by mime type association, and ensure that the essentially security – related requirements are fullfilled. Fig. 2 illustrates the overall application interaction manipulation procedure.

Fig. 2. Application interaction manipulation procedure

2.3 Students and Teachers' Modeling

With the advent of smart and ubiquitous spaces, recent research efforts on user modeling have focused on models that support intelligent environments to capture and represent information about users and contexts, so as to enable the environment to adapt to both [11]. A detailed review of user modeling in the context of ubiquitous user modeling applications is provided by Jaimes [10].

In the area of pervasive computing there has been considerable work in modeling location. Zhou et al. [21] identified two essential steps in personal place acquisition: obtaining physical locations and obtaining labels. Extending further the notion of labels, PERSONAF (Personalised Pervasive Scrutable Ontological Framework) [16] is an abstract framework for pervasive ontological reasoning, supporting personalisation in pervasive computing environments and scrutable representation and reasoning. PERSONAF has two main parts, the three-layered PECO ontology and the reasoning engine, ONCOR.

However, all the aforementioned user modeling systems mostly model generic users' attributes, and do not address specific issues which are fundamental in educational systems. In [2] an interesting approach based on adaptive hypermedia is introduced, bringing to surface different forms of models that adapt the content and the links of hypermedia pages to the user's relevant characteristics. As outcome of this work, the student model matching technology is introduced, which can be used in order to analyze and match student models of many students at the same time.

Based on the above, the User Profile module of ClassMATE facilitates the classroom's users (students and teachers) behavior monitoring and assessment, in order to provide user related metadata to the ClassMATE's services and applications. According to the IEEE's Learning Technology Systems Architecture (LTSA) [7], the Users Profile module represents a learners' record repository, which keeps track of every individual student's learning status and behavior data. Additionally, the User Profile accommodates the knowledge-based library of students' behavior patterns, dynamically gathered via activity monitoring. Data gathered by the User Profile service through an iterative monitoring and evaluation process, constitutes the main feedback for the Context Manager, so that a learner's centric rational is applied for content delivery and interaction control, providing thus adaptation to individual student's needs.

2.4 Providing Educational Content

In order for educational content to be provided to students effectively, addressing their learning needs, it should be coupled with a number of educational metadata. IEEE Learning Object Metadata (LOM) [8] is the commonly used standard for learning object metadata, specifying a conceptual data schema that defines the structure and the data elements of a metadata instance for a learning object. A learning object is defined as any entity -digital or non-digital- that may be used for learning, education or training. A Metadata instance for a learning object describes relevant characteristics of the learning object to which it applies. Such characteristics may be grouped into general, life cycle, meta-metadata, educational, technical, rights, relation, annotation, and classification categories. Because of LOM's flexibility, it has received widespread support from major players in the educational technology industry, including international repository efforts such as MERLOT (merlot.org) and ARIADNE (ariadne.unil.ch), as well as the U.S. Department of Defense SCORM initiative (www.adlnet.org).

ClassMATE encapsulates a smart classroom's educational content classification and archiving through its Data Space core module. The role of the Data Space is threefold: a) it implements a centralized content repository, providing transparent content access and management by any ClassMATE application and service, as if it was a local resource, b) it encapsulates a content classification mechanism, based on IEEE's LOM specification, providing the necessary content-related rationale to data mining procedures, and c) it encapsulates a sophisticated filtering mechanism for personalized content delivery. For the latter to be accomplished, the Data Space strongly collaborates with the User Profile to collect essential static or dynamic user characteristics.

2.5 Decision Making

Fully automated decision making system for smart environments is a difficult objective. A few fully-implemented applications of decision-making technologies have been reported (e.g. [15], [4]) and most of them are based on state of the art neural – networks approaches, that in most cases cannot be easily applied in everyday life activities data.

However, simpler approaches mainly based on monitoring and simple rule based decisions, may prove to be, in practice, more efficient than sophisticated decision making systems. In this respect, ClassMATE encapsulates the Context Manager core module, which is responsible for monitoring and triggering the system's behavior in the smart classroom. The Context Manager is the orchestration component of the ClassMATE's architecture. It monitors the ambient environment and makes context-aware decisions. In more details, the Context Manager is responsible for making the decisions for every process workflow in the classroom's environment, and for controlling the operation and collaboration of ClassMATE's services and applications to address users' needs at any specific point in time. To this purpose, the Context Manager applies appropriate reasoning strategies to (i) user-, (ii) service- and (iii) application- related data in the classroom environment. Besides this general orchestration provided by the Context Manager, every AmI artifact in the classroom operates under the orchestration of a local Artifact Director, which is responsible for its robust operation. The Artifact Director at any time keeps track of what is currently running (applications or services) on the artifact, and according to the Context Manager's directions they initiate, stop or suspend the processes running on the artifact.

3 Smart Classroom in Practice

In this section an experimental smart classroom environment used for ClassMATE deployment and assessment is described. This smart classroom environment was implemented as a prototype for the AmI classroom to be deployed in the FORTH-ICS' AmI facility [19].

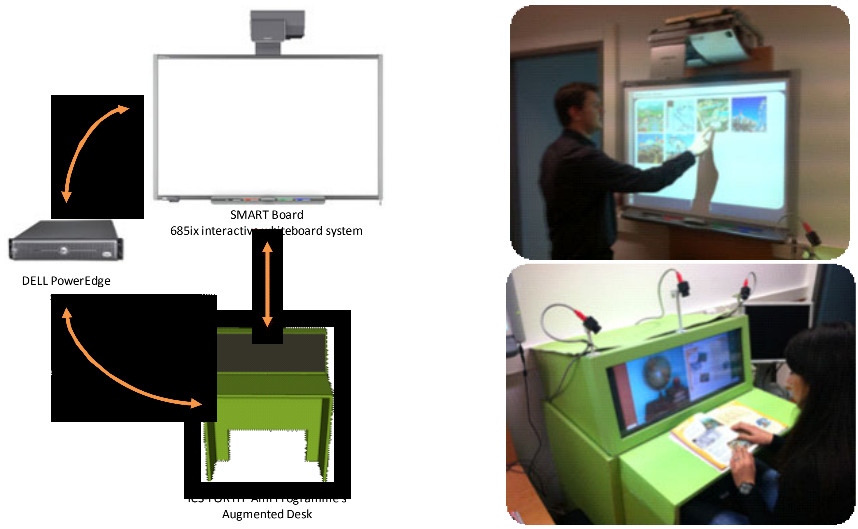


Fig. 3. Smart classroom prototype architecture and photos

Fig. 3 illustrates the overall architecture of the smart classroom prototype which is comprised of a SMARTBoard interactive whiteboard running Windows 7 Professional, the FORTH-ICS AmI Programme's augmented desk [1] and a DELL PowerEdge server which hosts the ClassMATE core framework. For the execution of educational experiments an educational front-end, named PUPIL [12], has been developed and deployed on the smart classroom prototype devices. In brief, PUPIL: (i) promotes the design of usable educational applications though a library of "intelligent" widgets, (ii) equips classroom artifacts with flexible workspaces and enables application migration among them, (iii) supports reusability of common interface patterns and minimizes artifact specific design decisions and (iv) frees designers from building the same interface for various platforms as the single version automatically transforms to the current context to ensure optimal display.

In order for an assessment of the educational process in smart classroom to be elaborated, a number of educational applications have been developed and are presented in this section, namely the ClassBook, the Multimedia, the Multiple-Choice Exercise and the Hints application.

ClassBook Application

The ClassBook Application is the electronic version of the physical book, displaying the currently open book page. The images and exercises displayed on any page are selectable, and upon selection the appropriate applications are launched to display relevant content (e.g., the Multimedia Application is launched if an image is selected or the Multiple-Choice Exercise if an exercise is selected).

Multimedia Application

The Multimedia Application as implied by its name displays multimedia content. The student can choose to view images or videos about a preselected topic. For example, the student can select an image displayed on the ClassBook Application to view relevant multimedia.

Multiple-Choice Exercise Application

This application is the online representation of a multiple-choice exercise. Through this application the student can solve the exercise electronically by just selecting one of the possible answers. A hint button is offered next to each sentence, in case the student requires help to find the correct answer. As soon as the hint button is selected, the Hints application is launched to display available information.

Hints Application

The Hints Application is launched only when the student explicitly asks for help about a specific exercise. It supports three kinds of hints that are presented to the student gradually in order to assist the development of critical thinking skills. If the first hint is insufficient for the student to solve the exercise, the he can ask for a second and a third one.

These applications constitute the front-end of the ClassMATE back – bone and aim to provide:

- Natural interaction with the smart classroom artifacts (desk and board). For example, students interact through touch and gestures with the artifacts, without the need for traditional obtrusive input devices.
- Adaptable UIs for all the smart classroom artifacts
- Personalized content according to the student(s) profile. For example, the multiple-choice exercise application presents exercises according to the student's skills and preferences.

4 Conclusion

This paper discussed an education-centric approach towards ambient intelligence in the classroom, raising fundamental requirements that should be taken into consideration. These requirements have been addressed by an integrated architecture, named ClassMATE, which facilitates all the necessary mechanisms for context – aware ubiquitous computing in the classroom. Furthermore, a smart classroom prototype, incorporating the ClassMATE's infrastructure, was presented constituting the first test – bed for the study of the educational process in intelligent classrooms.

The next steps for the presented work involve a real scale smart classroom, which will constitute one of the real life simulation spaces of FORTH-ICS' AmI facility The implementation of the simulation space will allow the long – term evaluation of ClassMATE with actual students and teachers, aiming not only to assess the usability of the proposed system, but its actual impact in the educational process as well. The consolidation of evaluation results will provide useful feedback for further elaboration and development.

References

1. Antona, M., Margetis, G., Ntoa, S., Leonidis, A., Korozi, M., Paparoulis, G., Stephanidis, C.: Ambient Intelligence in the classroom: an augmented school desk. In: Karwowski, W., Salvendy, G. (eds.) Proceedings of the 2010 AHFE International Conference (3rd International Conference on Applied Human Factors and Ergonomics), Miami, Florida, USA, July 17-20 (2010)
2. Brusilovsky, P.: Adaptive and Intelligent Technologies for Web-based Education. In: Rollinger, C., Peylo, C. (eds.) Künstliche Intelligenz 4, Special Issue on Intelligent Systems and Teleteaching, pp. 19–25 (1999)
3. Cook, D.J., Das, S.K.: How smart are our environments? An updated look at the state of the art. Journal of Pervasive and Mobile Computing 3(2), 53–73 (2007)
4. Cook, D.J., Youngblood, M., Heierman III, E.O., Gopalratnam, K., Rao, S., Litvin, A., Khawaja, F.: MavHome: an agent-based smart home. In: Proceedings of the First IEEE International Conference on Pervasive Computing and Communications, PerCom 2003, pp. 521–524 (2003)
5. Cook, D.J., Augusto, J.C., Jakkula, V.R.: Ambient intelligence: Technologies, applications, and opportunities. Pervasive and Mobile Computing 5(4), 277–298 (2009)
6. FORTH-ICS Ambient Intelligence Programme, `http://www.ics.forth.gr/ami` (accessed in April 2011)
7. IEEE Learning Technology Standards Committee: Draft Standard for Learning Technology-Learning Technology Systems Architecture (LTSA). IEEE Computer Society, IEEE 1484.12.1 (2002)
8. IEEE LOM: Draft Standard for Learning Object Metadata. IEEE Learning Technology Standards Committee, IEEE 1484.12.1 (2002)
9. IST Advisory Group: Ambient Intelligence: from vision to reality, (2003), Electronically tp://ftp.cordis.lu/pub/ist/docs/istag-ist2003_consolidated_report.pdf
10. Jaimes, A.: Data Mining for User Modeling and Personalization in Ubiquitous Spaces. In: Nakashima, H., Aghajan, H., Juan Carlos, A. (eds.) Handbook of Ambient Intelligence and Smart Environments, pp. 1015–1038. Springer Science+Business Media, Heidelberg (2010) ISBN 0-387-93807-3
11. Jameson, A., Krüger, A.: Preface to the special issue on user modelling in ubiquitous computing. User Modeling and User-Adapted Interaction 15(3-4), 193–195 (2005)
12. Korozi, M.: PUPIL: Pervasive UI development for the ambient classroom (Master's thesis) (2010), e-Locus `http://elocus.lib.uoc.gr/dlib/a/e/2/metadata-dlib-81a07682706c2163d8f582245fd9edfd_1288689489.tkl`

13. Leonidis, A., Margetis, G., Antona, M., Stephanidis, C.: ClassMATE: Enabling Ambient Intelligence in the Classroom. World Academy of Science, Engineering and Technology (66), 594–598 (2010), http://www.waset.org/journals/waset/v66/v66-96.pdf (retrieved February 16, 2011)
14. Levis, P., Culler, D.E.: Maté: a Tiny Virtual Machine For Sensor Networks. In: Architectural Support for Programming Languages and Operating Systems (2002)
15. Mozer, M.C.: Lessons from an adaptive home. In: Cook, D.J., Das, S.K. (eds.) Smart Environments: Technology, Protocols, and Applications, pp. 273–298. Wiley, Chichester (2004)
16. Niu, W.T., Kay, J.: PERSONAF: framework for personalised ontological reasoning in pervasive computing. User Modeling and User-Adapted Interaction 20(1), 1–40 (2010)
17. Shi, Y., Xie, W., Xu, G., Shi, R., Chen, E., Mao, Y., Liu, F.: The smart classroom: Merging technologies for seamless tele-education. IEEE Pervasive Computing, 47–55 (April-June 2003)
18. Stephanidis, C.: Human Factors in Ambient Intelligence Environments. In: Salvendy, G. (ed.) Handbook of Human Factors and Ergonomics, 4th edn. John Wiley and Sons, USA (to appear, 2012)
19. Stephanidis, C., Antona, M., Grammenos, D.: Universal Access Issues in an Ambient Intelligence Research Facility. In: Stephanidis, C. (ed.) Universal Access in Human-Computer Interaction: Ambient Interaction, Beijing, P.R. China, July 22-27. The Proceedings of the 12th International Conference on Human-Computer Interaction (HCI International 2007, vol. 6, pp. 208–217 (2007)
20. TinyOS Project, http://www.tinyos.net (accessed in April 2011)
21. Zhou, C., Frankowski, D., Ludford, P., Shekhar, S., Terveen, L.: Discovering personally meaningful places: an interactive clustering approach. ACM Trans. Inf. Syst. 25(3) (2007)
22. Zigbee Alliance, http://www.zigbee.org/ (accessed in April 2011)

Learning Styles and Navigation Patterns in Web-Based Education

Jelena Nakić, Nikola Marangunić, and Andrina Granić

Faculty of Science, University of Split, Nikole Tesle 12, 21000 Split, Croatia
{jelena.nakic,nikola.marangunic,andrina.granic}@pmfst.hr

Abstract. Researchers agree that it is possible to diagnose student's learning style and that learners with a dominant preference for certain learning style may have difficulties in knowledge acquisition in conditions where teaching strategy is not compatible with it. This paper presents an experimental work conducted in order to determine if students with different learning styles have different navigational needs while using web-based learning environment. Correlations between learning styles and students' learning performance are also observed and discussed. A group of 102 graduate and postgraduate students were involved in the study. Learning styles according to Felder-Silverman learning style model have been explored in the context of an e-learning course delivered through a Learning Management System. The main results show that the course supports global learners to some extent. The methods for meeting the needs of sequential learners are proposed.

Keywords: learning styles, learning behaviour, navigation patterns, web based environment, empirical study.

1 Introduction

Learning is a process that engages perceptual and cognitive capabilities of students in various ways. Students perceive, process and represent learning material differently; they have different preferences for types, number and order of learning resources. Generally, a set of attitudes and behaviours which determine an individual's preferred way of learning is considered as a learning style [1]. While there is a number of learning style theories, researchers agree that it is possible to diagnose student's learning style and that learners with a dominant preference for certain learning style may have difficulties in knowledge acquisition in condition where it is not compatible with the teaching strategy [2]. The thesis that incorporation of learning styles in learning environment enables more pleasant learning experience and higher performance of students has inspired development of many adaptive educational systems such as CS388 [3], INSPIRE [4], eTeacher [5], INDeLER [6].

This paper presents an experimental work conducted in order to determine if students with different learning styles have different navigational needs while using a Learning Management System (LMS). Correlations between learning styles and students' learning performance are also observed and discussed.

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 587–596, 2011.

The paper is structured as follows. Background section briefly introduces the theories of learning styles with an emphasis on Felder-Silverman learning style model [2] and stresses the efficiency issues of adaptation to learning styles in web-based education. Main part of the paper brings detailed experimental procedure along with obtained results. Discussion offers interpretation of findings and critical review of the methodology. The last section summarizes the results and outlines future work.

2 Background and Motivation

2.1 Learning Styles

Exploration of learning process has emerged a variety of models of learning styles, each relying on diverse concepts and proposing distinctive descriptions and classifications of learners' tendencies. Kolb [7] propose that learners can be distinguished into *convergent learners*, *divergent learners*, *assimilators*, and *accommodators.* Honey and Mumford [1] offer another interpretation of Kolb's theory of experiential learning [7] and classify learners into four types: *activists*, *pragmatists*, *reflectors* and *theorists*. This learning style model is frequently implemented in existing adaptive systems, for example INSPIRE [4]. However, one of the most popular models of learning styles in adaptive education is Felder-Silverman learning style model, FSLSM [2], implemented for example in CS388 [3], SAVER [8], eTeacher [5], INDeLER [6], as well as an add-on for the LMS Moodle [9]. FSLSM places student's learning tendencies along discrete scales on four dimensions: *sensing/intuitive*, *visual/verbal, active/reflective* and *sequential/global*. This model provides very precise quantitative estimation of learner preference for each dimension. To date, FSLSM is considered as the most suitable learning style theory for application in adaptive systems design and development [10; 11].

The sensing/intuitive dimension of FSLSM classifies learners according to the type of information they preferentially perceive: sensing learners prefer concrete information with lots of facts and examples, while intuitive learners learn better from theories and principles. Visual/verbal dimension reflects students' preferred perceptual tendencies: visual learners like to see pictures and graphs; verbal learners learn better what they hear and the best what they hear and discus out loud. Verbal learners also deal better with written representation of data than visual learners. Active/reflective dimension considers students' way of processing information, i.e. converting it into knowledge: active learners prefer to be engaged in physical activity, collaborative discussion or any kind of experimentation, while reflective learners benefit from introspection and quiet observation. Finally, sequential/global dimension describes the way students make progress towards comprehension of subject: sequential learners proceed through the material in a logical order, usually in the manner the material is presented. Opposite to them, global learners prefer to glance through the whole material and then select the topics to grasp more deeply. They usually master the material by jumping to more complex issues, filling the gaps after.

In empirical study that will be presented in the paper, learning styles have been studied in the context of the e-learning material developed as Moodle lessons for learning basics of human-computer interaction (HCI). The material offers fundamentals and principles of the domain; it is rather theoretical, accompanied with relatively small number of illustrations and few examples. Along these lines we believe that such lessons are already adjusted to intuitive, verbal and reflective students, so the experiment deals only with the sequential/global dimension of the FLSLM.

2.2 Efficiency of Adaptation to Sequential and Global Learners

Despite an increasing number of existing adaptive and adaptable learning systems which consider learning styles, empirical studies on an efficiency of the developed systems are rarely conducted. Furthermore, results of such studies, even only in respect to the sequential/global dimension, are rather inconsistent. For example, Bajraktarević et al. [12] found significantly higher learning outcomes for students learning in conditions that matched their learning styles. Similar results were found in several studies regarding other models of leaning styles comparable to the sequential/global dimension [13; 14]. On the other hand, there are studies that have not found any differences concerning learning outcomes between matched and mismatched group. For example, Graf and Kinshuk [9] found that adaptation is effective only in terms of time students spent on learning activities and of the number of requests for additional learning objects, but not in terms of learning outcomes. Furthermore, Brown et al. [15] found no correlations of sequential/global dimension and learning performance in any of relevant aspects. Their study has imposed further investigation of potential educational benefits of adaptation to sequential and global learners.

Taking into account related work and the fact that user performance considerably depends on a particular system, we have studied learning styles in the context of the e-learning material developed at the University of Split, Croatia. If an outcome of the study suggests that the learning process would benefit from the lessons adapted to sequential and global learners, once created learning material could be easily restructured in terms of learning objects sequences and their availability for students with different learning preferences.

3 Empirical Study

The paper presents an experimental work aimed at finding out more about learning styles, particularly sequential/global dimension and the relationship to students' navigation patterns when learning in the web-based environment. An empirical study was carried out in order to identify and validate individual sequential/global dimension, as well as to detect its correlations with the results of selected objective variables. Learning outcome and learning behaviour in the chosen web-based learning environment have been measured and defined as objective variables.

3.1 Research Questions

This empirical study attempted to provide answers to the following questions: do learners with different learning styles employ different navigation patterns when using the web-based e-learning system; does it affect their learning performance and if does, in what way? Consequently, we have postulated the following:

H_0: There is no difference in the learning performance, i.e. the sequential/global learning does not affect learning outcomes and learning behaviour if the students are provided with enough time to pursue the whole learning material.

H_1: Learners classified into one learning style according to the sequential/global dimension accomplish higher learning outcomes.

H_2: Differences related to the sequential/global dimension result in different navigation behaviour of learners, i.e. sequential learners prefer system controlled navigation, while global learners use a number of various navigational tools which enable learner controlled navigation.

3.2 Instruments and Measures

In order to identify learning styles, particularly the sequential/global dimension, the Index of Learning Styles questionnaire (ILS) [16] was used, which aims at assessing learning styles based on the FSLSM. The ILS questionnaire is commonly used instrument which contains 44 questions distributed along four dimensions. FSLSM determines learning preference for each dimension by values ranged from 11 to -11. For the purpose of the study, 11 questions related to the global/sequential dimension have been extracted.

E-learning material developed as Moodle lessons and used in several HCI courses at University of Split was involved in the study. Each learning page is accompanied with a set of navigation tools to support various learning behaviour of students. Figure 1 illustrates the layout of a typical page: learning content is placed in the centre of the screen along with navigation buttons located below, while lesson menu map is positioned on the left side of the screen, offering chapter headings as links.

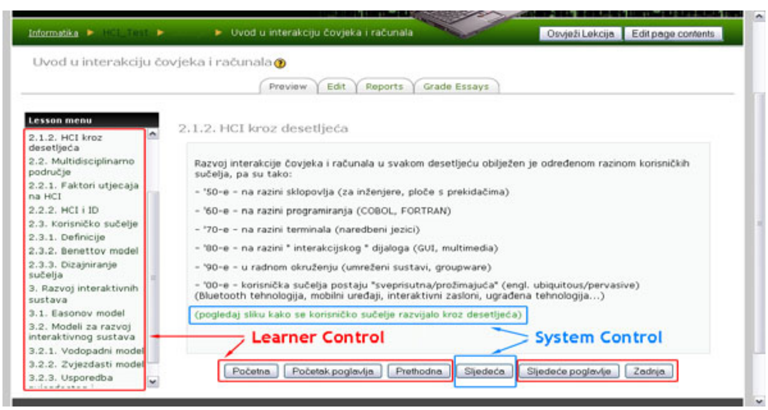


Fig. 1. Screenshot of the e-learning material with navigation tools

Analysis of the conducted learning sessions enabled us to identify student's learning behaviour thus providing measures for a number of variables: navigation steps and usage of system control versus learner control tools, together with navigation paths of each student.

Navigation steps are represented by two variables called *Next* (relative number of system control tools used by students, i.e. the ratio of system control tools number and total number of clicks) and *Map* (relative number of clicks on the lesson menu map). Navigation path was determined through the total number of clicks (variable *Clicks*), the relative number of *Jumps* (usage of learner control buttons) and the number of *Passes* through the whole learning material which students used in their learning session.

Learning outcomes, expressed as variable *Quiz,* are calculated from scores on a post-test which concludes learning session. The post-test is structured as an on-line quiz via Moodle with 20 multiple choice questions. The experimental design of the conducted empirical study is illustrated in Figure 2.

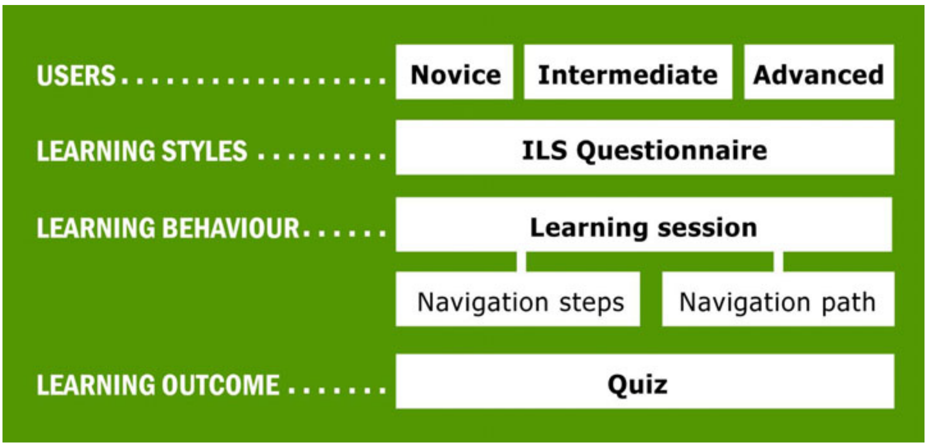


Fig. 2. Design of the experiment

3.3 Participants and Procedure

A group of 102 participants from two faculties was involved in the study. They were all computer science students, sharing nearly equal experience in web-based learning. Experimental procedure comprised three major steps. In the first step a pilot study with ten participants was performed in order to test learning material adopted for the experimental session, duration of the session and post-test quiz, as well as to determine quantitative parameters for the measurement of the student's learning behaviour.

Remaining 92 participants joined the study in the second and the third step. They have been allocated to two experimental groups: 61 participants to the "learners" group and 31 to the "assistants" group. Learners were selected from three different university courses and have been classified according to their background knowledge related to the chosen subject matter into novice, intermediate and advanced learners. In the second step all learners completed ILS questionnaire. A twenty minutes learning session followed by the on-line post-testing, as the third step of the experiment, concluded the procedure. The group consisted of 31 assistants was randomly chosen to help in observing and recording learners' work.

In the end, 53 participants have successfully completed procedure of the main study. The group was consisted of 27 male (51%) and 26 female (49%) students. The age of all participants varied from 18 to 23 years, with a median value of 22. The distribution of gender, age and study group is presented in Table 1.

Table 1. The distribution of age, gender and study group in the sample

Participants	Study group	Gender			Age	
		Male	Female	Total	< 22	> 22
Novice	5. sem. graduate	9	13	22	21	1
Intermediate	1. sem. postgraduate	13	7	20	14	6
Advanced	3. sem. postgraduate	5	6	11	0	11
Total		27	26	53	35	18

3.4 Results

With the aim of testing our hypotheses, data analysis was performed using SPSS 16.0 software statistical package.

Descriptive statistics for all measured variables, along with the overall number of participants (N), the minimum and the maximum scores, arithmetic means (Mean) and standard deviations are presented in Table 2.

Table 2. Descriptive statistics of the measured variables

Variables	N	Minimum	Maximum	Mean	Standard Deviation
Quiz	53	3,13	7,93	5,92	1,170
Clicks	53	8	116	36,68	17,414
Passes	53	1	2	1,55	0,503
Next	53	0,13	1	0,73	0,204
Map	53	0	0,74	0,11	0,179
Jumps	53	0	0,64	0,10	0,131

Table 3 shows descriptive statistics of measured variables for the participants divided into two groups of learners according to their ILS score. The number of participants per group (Ng), arithmetic means (Mean) and standard deviations are presented.

Table 3. Descriptive statistics of the measured variables for participants divided into two groups of learners

Variables	ILS	Ng	Mean	Standard Deviation
Quiz	sequential	25	5.64	1.244
	global	28	6.18	1.060
Clicks	sequential	25	3540	15.610
	global	28	37.82	19.094
Passes	sequential	25	1.56	0.507
	global	28	1.54	0.508
Next	sequential	25	0.69	0.219
	global	28	0.66	0.191
Map	sequential	25	0.12	0.197
	global	28	0.11	0.165
Jumps	sequential	25	0.12	0.148
	global	28	0.08	0.113

Factors that could have an influence on web-based learning were identified from the acquired results of measured variables. Learning outcomes and perceived navigation of users were explored with the analysis of differences of arithmetic means and by the correlation analysis. Statistically significant difference in the results of measured variables for the two groups of learners was tested with a number of t-tests for independent samples. T-test scores (t) along with degrees of freedom (df) and levels of significance (p) are shown in Table 4.

Table 4. Findings on testing differences in results of measured variables for the two groups of learners

Variables	t	df	p
Quiz	-1.70	51	0.096
Clicks	-0.50	51	0.618
Passes	0.17	51	0.863
Next	0.67	51	0.507
Map	0.09	51	0.927
Jumps	1.21	51	0.230

Bivariate correlation method was carried out. Pearson's correlation coefficients for participants' results in the relevant variables are shown in Table 5.

Table 5. Pearson's correlation coefficients of measured variables
* significant correlation at level of p < 0.1
** significant correlation at level of p < 0.01

Correlation of results	r
Quiz - Jumps	-0.23*
Next - Map	-0.89**
Next - Jumps	-0.42**
Map - Jumps	0.60**

4 Discussion

Statistically significant difference in the *Quiz* scores ($t = -1.70$; $df = 51$; $p < 0.1$) between the two groups of learners indicates distinction in learning outcome in favour of global learners. Furthermore, analysis of results shows no significant differences in the measures for navigation behaviour of learners. For that reason H_0 hypothesis is only partially supported, while H_1 hypothesis is fully accepted.

The finding that supports H_1 hypothesis could be a result of the structure of the learning material: the lessons are presented first, and students take quiz after they have learned all the lessons related to the subject matter. Although not created with an intention to support any learning preference, it seems that such structure supports global learners, the result which is in line with suggestions from the literature, e.g. [11].

A statistically significant correlation ($r = -0.23$; $p < 0.1$) was obtained between variables *Quiz* and *Jumps*. Negative correlation implies that in post-testing users with higher usage of learner control tools score slightly lower result. No difference was found in the relative number of jumps which have been made by our two groups of users in their learning session. Our findings suggest that both sequential and global learners had jumped to the similar extent. While free browsing is, according to FSLSM, typical behaviour of global learners, sequential learners usually follow navigation path controlled by the system. For that reason, previous result suggests that sequential learners who have made a lot of jumps, scored lower in post testing. It seems that the structure of the learning material with no navigational restrictions in a way "discriminates" sequential learners.

Considering navigation patterns for the two groups of learners categorized by the ILS, there were no statistically significant differences found to support hypothesis H_2, thus H_2 is rejected. However, analysis of the results has shown clear distinction between users who prefer the system control over the learner control. High negative correlation between variables *Next* and *Map* ($r = -0.89$; $p < 0.01$) in addition to correlation between *Next* and *Jumps* ($r = -0.42$; $p < 0.01$) reveals strong differences in learners' navigation behaviour; regular users of system control tools rarely used learner control tools. Positive correlation between *Map* and *Jumps* ($r = 0.60$; $p < 0.01$) is also in line with our findings, supporting the structure of the learning material and offered navigation tools.

Obviously, there is a vast distinction among learners in their navigation behaviour, but it seems that the ILS questionnaire as an instrument does not offer sophisticated way of their distinction. To investigate this hypothesis, we have tested internal consistency reliability of the ILS. Cronbach's alpha coefficient for the sequential/global dimension was 0.45, while literature suggests that an acceptable value for attitude-assessing instruments is at least 0.5 [17]. Research shows that the scores for the reliability of the instrument could be improved if the weakest item of the instrument is removed [18]. In our study, such elimination resulted only in a minor increment of the alpha coefficient, that is 0.48. Although a majority of studies reported slightly higher values of the internal consistency reliability of the ILS, as reviewed for example in [19], researches agree that ILS reliability is generally weak and that there is a need for its refinement.

Besides using insufficiently reliable measuring instrument, another possible limitation of the conducted experiment should be mentioned. There were a number of "assistants" involved in the empirical study, i.e. 31 students were engaged for observing and recording user's navigation patterns. Although they have been precisely instructed, still the data for 8 of the total of 61 users have been dismissed due to the irregularities in the observation of their learning behaviour. Therefore, in similar experiments the usage of log files or adequate software for screen monitoring would be much more appropriated.

On the other hand, a strong feature of the conducted empirical study comes from careful operationalization of learning behaviour which has resulted in the definition of the precise metric for learners' navigation patterns. Defined variables for measuring navigation steps and navigation path are conformed to the structure and layout of the learning material, and at the same time applicable for the identification of navigation patterns in comparable learning environments.

5 Conclusion

Summarizing the results of the conducted experiment we conclude that the structure of the learning material for the selected course to some extent supports global learners. This result indicates that sequential learners could benefit when knowledge testing is more frequently provided (best at the end of each chapter) and from the partial restriction of available navigation tools, thus enabling more strict system control. The additional finding of the study is related to the analysis of the learners' navigational patterns and their possible connections with the learning outcomes. However, such relationship along with learning styles implications on the web-based learning requires additional thoughtful research. Currently, the opinions about possibilities of an employment of the learning styles in adaptive systems are rather controversial [20; 21]. Thus our future work will also encompass investigation of other factors that could contribute to the better learning performance in the web-based environment, in accordance with the framework for user individual differences proposed in [22].

Acknowledgments. This work has been carried out within project 177-0361994-1998 *Usability and Adaptivity of Interfaces for Intelligent Authoring Shells* funded by the Ministry of Science and Technology of the Republic of Croatia.

References

1. Honey, P., Mumford, A.: The Manual of Learning Styles. Peter Honey Publications, Maidenhead (1992)
2. Felder, R.M., Silverman, L.K.: Learning and Teaching Styles in Engineering Education. Engineering Education 78(7), 674–681 (1988)
3. Carver, C.A., Howard, R.A., Lane, W.D.: Addressing different learning styles through course hypermedia. IEEE Transactions on Education 42(1), 33–38 (1999)
4. Papanikolaou, K.A., Grigoriadou, Kornilakis, H., Magoulas, G.D.: Personalising the Interaction in a Web-based Educational Hypermedia System: the case of INSPIRE. User-Modeling and User-Adapted Interaction 13(3), 213–267 (2003)
5. Schiaffino, S., Garcia, P., Amandi, A.: eTeacher: providing personalized assistance to e-learning students. Computers & Education 51, 1744–1754 (2008)
6. Jovanović, D., Milošević, D., Žižović, M.: INDeLER: eLearning Personalization by Mapping Student's learning Style and Preference to Metadata. International Journal of Emerging Technologies in learning (iJET) 3(4) (2008)
7. Kolb, D.A.: Experiential learning: experience as the source of learning and development. Prentice-Hall, Englewood Cliffs (1984)
8. Garcia, P., Amandi, A., Schiaffino, S., Campo, M.: Evaluating Bayesian networks' precision for detecting students' learning styles. Computers and Education 49, 794–808 (2007)
9. Graf, S., Kinshuk: Providing adaptive courses in learning management systems with respect to learning styles. In: Richards, G. (ed.) Proceedings of the World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education (E-learn), pp. 2576–2583. AACE Press, Chesapeake (2007)

10. Kuljiš, J., Liu, F.: A comparison of learning style theories on the suitability for e-learning. In: Hamza, M.H. (ed.) Proceedings of the IASTED Conference on Web Technologies, Applications, and Services, pp. 191–197. ACTA Press (2005)
11. Graf, S., Liu, T.-C., Kinshuk, Chen, N.-S., Yang, S.J.H.: Learning Styles and Cognitive Traits - Their Relationship and its Benefits in Web-based Educational Systems. Computers in Human Behavior 25(6), 1280–1289 (2009)
12. Bajraktarević, N., Hall, W., Fullick, P.: Incorporating Learning Styles in Hypermedia Environment: Empirical Evaluation. In: de Bra, P., Davis, H.C., Kay, J., Schraefel, M. (eds.) Proceedings of the Workshop on Adaptive Hypermedia and Adaptive Web-Based Systems, Nottingham, UK, pp. 41–52. Eindhoven University, Edinhoven (2003)
13. Ford, N., Chen, S.Y.: Individual differences, hypermedia navigation and learning: An empirical study. Journal of Educational Multimedia and Hypermedia 9(4), 281–311 (2000)
14. Liegle, J.O., Janicki, T.N.: The Effect of Learning Styles on the Navigation Needs of Web-based Learners. Computers in Human Behavior 22, 885–898 (2006)
15. Brown, E., Fisher, T., Brailsford, T.: Real Users, Real Results: Examining the Limitations of Learning Styles within AEH. In: Proc. of the Eighteenth ACM Conference on Hypertext and Hypermedia (2007)
16. Felder, R.M., Soloman, B.A.: Index of learning styles questionnaire (1997), `http://www.engr.ncsu.edu/learningstyles/ilsweb.html` (retrieved September 15, 2010)
17. Tuckman, B.W.: Conducting Educational Research, 5th edn. Wadsworth Group, Belmont (1999)
18. Litzinger, T.A., Lee, S.H., Wise, J.C., Felder, R.M.: A Study of the Reliability and Validity of the Felder-Solomon Index of Learning Styles. In: Proceedings of the 2005 American Society for Engineering Education Annual Conference & Exposition (2005)
19. Felder, R.M., Spurlin, J.: Applications, reliability and validity of the index of learning styles. International Journal on Engineering Education 21(1), 103–112 (2005)
20. Brusilovsky, P., Millán, E.: User Models for Adaptive Hypermedia and Adaptive Educational Systems. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) Adaptive Web 2007. LNCS, vol. 4321, pp. 3–53. Springer, Heidelberg (2007)
21. Brown, E., Brailsford, T., Fisher, T., Moore, A.: Evaluating Learning Style Personalization in Adaptive Systems: Quantitative Methods and Approaches. IEEE Transactions on Learning Technologies (Special Issue on Personalization) 2(1), 10–22 (2009)
22. Granić, A., Nakić, J.: Enhancing the Learning Experience: Preliminary Framework for User Individual Differences. In: Leitner, G., Hitz, M., Holzinger, A. (eds.) USAB 2010. LNCS, vol. 6389, pp. 384–399. Springer, Heidelberg (2010)

Phynocation: A Prototyping of a Teaching Assistant Robot for C Language Class

Akihiro Ogino, Haruaki Tamada, and Hirotada Ueda

Faculty of Computer Science and Engineering, Kyoto Sangyo University, Japan
{ogino,tamada,ueda}@cc.kyoto-su.ac.jp

Abstract. This paper shows a system that supports students who learn programing language in a class using a robot. We found out that it would help students to solve the error that the robot talks hints with gestures.

Keywords: Human-Robot Interaction, Educational Technology.

1 Introduction

As robots become familiar to people, friendly human-robot interaction is an important research topic. A human have a friendly feeling to a robot by its gesture and manner of speaking even if the robot makes uniform response [1]. As one of applications concerning friendly human-robot interaction, we are developing a teaching assistant robot system. As a study similar to us, Sartatzemi et al. reported the research of teaching the programming fundamentals to Greek secondary school student using LEGO Mindstorms technology [5, 6]. It is important to learn programing language for students in faculty of information. They cannot avoid studying it.

Our robot system supports understanding of beginner students in a programing language class by talking to a student an error place. Our system also collects information of student's error and provides the information to the teacher to infer student's understanding. In our previous paper [7], we have proposed a framework to support a teacher and students in a class of programing using a communication robot.

As for support of students, students can use simply our system without being conscious of the usage of the system and can get hints from a robot. For example, in a programing class for beginner, students usually make errors such as a syntax error (missing semi-colon and unmatched brackets) and a linkage error (misspell of symbol names). Therefore, the compiler and linker in a computer show insipid error messages that the beginner students should fix. However, the students almost cannot understand the cool-headed message, because actual causes are too far from causes he/she thinks. They go off programming before they learn programing language deeply. Therefore, our robot system provides hints by voice to each student and entertains them with gesture when he/she has compiled his/her programing codes and has made errors. Students who troubled in simple errors can have a great time solving their error by themselves. Hence, the teacher can take time to the solution of complex error.

Meanwhile, it is rare case that the programming education is conducted one-on-one. Generally, one teacher lectures to students more than 60. It is difficult for the teacher to have a time to explain all problems of each student and to know weak

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 597–604, 2011.

points of individuals. As for support of the teacher, our system collects errors of students to support that the teacher knows weak points of students at the class and makes assignments of programing.

In this paper, we present a result of evaluation concerning support of students that is one of two support functions to have of our system. We had psychological experiment that investigates degree of satisfaction concerning a robot suggestion to a student. We evaluate the potential of our robot system to beginner students.

This paper is organized as follows: Section 2 explains the system design of our system. Section 3 explains an experiment to evaluate whether a robot of our system is useful to support that a student corrects programing error. Section 4 summarizes this paper.

2 System Design

In this section, we present the design of our robot system to support a teacher and students in a programing class.

Figure 1 shows an overview of our system. In the practical course, the teacher usually gives assignments of programing to students. Students try to solve given assignments, and the teacher educates some techniques through assignments. Our system, which is surrounded by the dotted oval of the figure, supports the teacher and students.

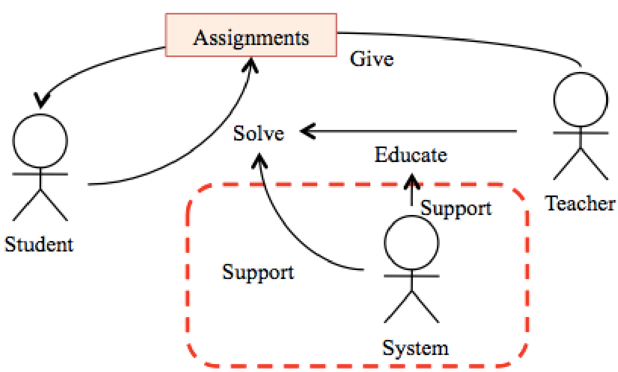


Fig. 1. Overview of Our System

2.1 Function to Support Students

Our system has two functions to support beginner students that solve assignments of programing by a robot.

- First function is that students use simply our system without being conscious of the usage of the system
- Second function is to keep motivation of students concerning exercises of programing

As first function, we developed a function for which students can use our robot system without being conscious of the usage. It is difficult for beginners to learn programing skills and usage of our system at the same time because beginners do not understand operations for programming. Therefore, it disturbed students to learn usage of support system in order to get programing skills. Then the efficiency of learning would be greatly decreased. We designed our system that has not the special interface such as a web interface and a special command. Beginners can concentrate on learning the work of programming through usual interfaces with our robot support.

As second function, we developed a function for which students keep their motivation to learn programing by the robot that has a game element to increase the learning efficiency [2].

The practical course is that the student tries to solve given assignments by making trial and error. Unfortunately, the beginning of programming student makes many mistakes because of misunderstanding and incomprehension of compilation error messages. The student might give up the programming. Our system provide hints and praises to students to elicit the appetite for learning through a robot. The robot gestures based on error of students (e.g. failure of compilation) and cheer them up. We induce that students have interest in programing through amusing gestures of the robot. We choose Phyno as the robot [3]. The phyno is a child robot, which modeled a 3 years old child, and is designed to perform pretty gestures. We guess that the cuteness of gestures makes the students that try to find out various gestures and keep practice of programing.

2.2 Function to Support a Teacher

Our system supports education by showing the statistics information of the data based on actions of the students. The teacher needs to educate many things to students. Also, the teacher usually wants to change the level of education by understanding of students.

However, the teacher can measure understanding of students in only particular opportunity such as exam, and report. The teacher can change the level of education if the teacher can know understanding of the students in the class. For example, if many students finish an assignment by taking hints of the assignment, the teacher could open the new assignment. On the other hand, if not, the teacher can explain the assignment, or give more hints.

In this work, we focus on the compilation of source code in actions of the learner for educational support. The programming beginner does not have enough knowledge to fix the compilation error. Therefore, he/she cannot understand compilation errors and does not read them. It is hard task for programming beginners to fix compilation errors. The teacher wants to comprehend the condition of the learner in each situation for more effective education.

Our system collects error messages of compilation of learners. Our system automatically categorizes and summarizes collected compilation errors. Our system presents the resultant data to the teacher.

2.3 Scenario Using Our System

We show a typical scenario of a programing class using our robot system in Figure 2. The teacher supports students by two ways that are hints through the teacher and hints with the robot.

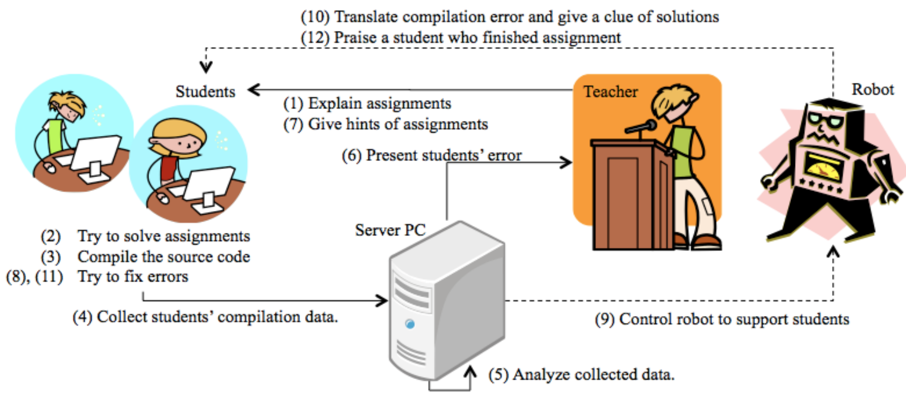


Fig. 2. Scenario using Our System

A teacher gives assignments to students and the students try to solve it:

(1) A teacher explains assignments of programing.
(2) Students try to solve given the assignments.
(3) The students compile a source code of an assignment.
(4) Our system collects compilation data of students
(5) Our system analyzes amount of statistics from them.
(6) The way to support students by the teacher:
(7) The system presents a statistics of errors of students to the teacher.
(8) The teacher gives hints of assignments to students by taking the statistics of errors of the students.
(9) The students try to fix errors by getting hints or the clue of solution from the teacher.
(10) The way, which is drawn the dotted line, to support students by the robot:
(11) Our system controls the robot to support students by his gestures and voice.
(12) The robot translates the compilation error and gives a clue of solution.
(13) Students try to fix errors by getting hints or the clue of solution from the robot.
(14) The robot praises them when students fix errors.

3 Experiment to Evaluate Robot Support

In this study, we have developed two functions to support students among functions to support a teacher and students. First function is that students use simply our system without being conscious of the usage of it. Second function is that a robot keeps motivation of students to learn a programing language. We evaluated two functions through basic comparative experiments which are an experiment using a robot and not using a robot.

3.1 Development of a System with Simply Usage

We exploit the Mac OS X for programming environment in order to develop the first function. The Mac OS X server has the NetBoot service which enables to boot client PC from server-side disk image. We can provide same environment to the laptop PC of each student by the NetBoot.

We designed the interface of terminal that is almost same as normal terminal and set up it in the server-side disk image. For example, a student compiles his/her source code of C language using "gcc" such as "gcc -o sample sample.c". In our system case, the student compiles the source code using "phynoc" such as "phynoc -o sample sample.c". Therefore, students do not need to be conscious of the usage concerning our system.

3.2 Evaluation of Robot Support

We had basic comparative experiments to investigate that a communication robot can support students to keep motivation for programing.

- Experiment A: A student as an examinee corrects errors in a sample source code A of C language with a robot.
- Experiment B: The examinee corrects errors in another sample source code B of C language without the robot.

We made the basic comparative experiments to 11 examinees. We divided two groups in order to avoid the order effect. An examinee of group 1 carried out experiment B after experiment A. Group 2 is the reverse of group 1. We performed the experiment A and B by turns. Figure 3 shows one scene of experiment A. An examinee looks at the robot and hears a hint of it. We made two sample source codes as almost same. Number of errors is same.

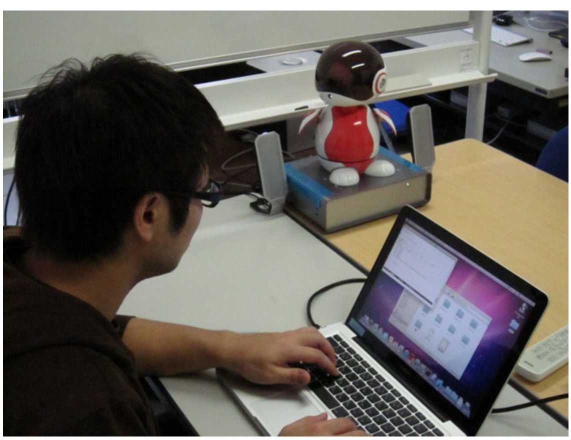


Fig. 3. A Scene of Experiment A

We have evaluated a support of robot using data that examinees wrote impressions of the experiment with the robot and without it. We carried out two experiments in separate rooms. Figure 4 shows positions of equipment in the room A and B. There is one laptop PC, the robot and one teacher in the room A. The examinee tries to solve the sample source code A in this room A. The robot looks at the examinee and talks the type of the error and the line number of the source code that the error occurred. When an examinee made one error in his computer, the robot talks a hint concerning the error. The robot continues to talk hints until he resolves all errors. For example, the robot talks "a semi-colon is missing on line 3" to him with gestures. The teacher in the room A helps when the examinee cannot solve the source code. There is one laptop PC and one teacher in the room B. The examinee tries to solve the sample source code B in this room B. The teacher in the room B also helps when the examinee cannot solve the source code.

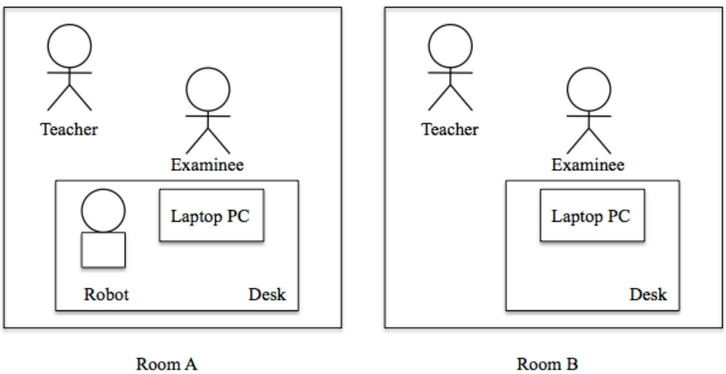


Fig. 4. Environment of Comparative Experiments

Table 1. A result of Questionnaires Concerning Support by a Robot

Questionnaires	Number of Yes
Was the support with the robot useful to correct errors?	5 / 11
Do you think the support with the robot useful for the person who gives up correcting errors on the way?	11 / 11
Do you think the support with the robot useful for a person without motivation for learning?	10 / 11
Do you think the robot needs to talk detailed explanation of errors more than this experiment?	5 / 11

In first questionnaires, we asked them whether the support with the robot was useful to correct errors, when you compare the case that there is not a robot and there is a robot. 5 examinees responded that the robot is useful to correct errors.

They wrote the reasons as follows:

- I could find easily the place in which the error occurred because the robot talks the type of the error and the line number of the source code that the error occurred.

- I could easily understand meaning of errors because the robot translates to Japanese from English.

Other examinees wrote that I do not need the robot because it was easy for them to correct errors in the sample source. In this experiment, we used errors of entry-level, such as "a semi-colon is missing" or " a variable j is not defined". They did not need hints to correct errors. We have to change the source code to match the level of examinee in next experiment.

In second questionnaires, we asked them whether the support with the robot was useful for the person who gives up correcting errors on the way, when you compare the case that there is not a robot and there is a robot. In this questionnaire, all examinees responded that the robot would help weak-willed person concerning learning the programing language. They wrote the reasons as follows:

- The robot uses a hint and can teach what the weak-willed student should begin with.
- The robot repeats hints by Japanese until the student corrects errors.

We found out that it is difficult for them to learn programming language and English at the same time. Therefore, it is good for them that the robot talks a hint with the mother tongue of them in order to concentrate to learn programing.

In third questionnaires, we asked them whether the support with the robot was useful for a person without motivation for learning, when you compare the case that there is not a robot and there is a robot. Almost examinee responded that the robot is useful. One of examinees wrote the reasons as follows:

- The student would have interest in learning of programing through actions of the robot.
- I look at the robot with amusement when it talks a hint to him/her.

We think that gestures of the robot can generate an element of entertainment and can bring out interest in learning.

In fourth questionnaires, we asked them whether the robot needed to talk detailed explanation of errors more than this experiment. More than half of examinees responded that it did not need. They wrote the reasons as follows:

- In order that the student learns a programing language by himself, the robot only gives simple hints to students and the teacher has better to explain detail of the errors.

We found out that students try to learn by themselves if we can think effective level of hints.

4 Conclusion

In this paper, we have presented a system for support of students who learn programing language, using a robot. Our system has two functions to help student.

One function is for simply usage. When they learn programing language, it is hard to learn the usage of our system at the same time. We developed an interface of system without changing a method of compiling it. Students can use our system in the same way as a common system and get a support by a robot.

Another function is for a learning support of student using a robot. We use a robot to help students to learn a programing language in an exercise class. The robot talks a hint to a student when they have made an error in their source code using mother language. The robot indicates the line number on which the error occurred and the type of the error as a hint. For example, the robot talks that "Hello. Mr. A, please check that a semicolon may be missing on the line 3". The hints of robot would be useful because beginner students cannot understand the meaning of the error by English that he/she made.

We have presented a result of psychological experiment that investigates degree of satisfaction concerning the robot suggestion to a student. The result of the experiment showed that it would help students to solve the error that the robot talks hints. Gestures of the robot would keep motivation of them to learn programing language.

In the future, we would try to find a timing of gestures of a robot in order to help students who correct errors in a source code of a programing language. We would investigate aspects of amusement concerning gestures of a robot to keep motivation to learn a programing.

References

1. Matsumoto, N., Ueda, H., Yamazaki, T., Tokosumi, A.: The cognitive characteristics of communication with artificial agents. In: Proc. Joint 3rd International Conference on Soft Computing and Intelligent Systems and 7th International Symposium on advanced Intelligent Systems (SCIS & ISIS 2006), pp. 1269–1273 (September 2006)
2. Feldgen, M., Clua, O.: Games as a motivation for freshman students to learn programming. In: Proc. of 34th Annual Frontiers in Education, FIE 2004, pp. 11–16 (2004)
3. Ueda, H., Chikama, M., Satake, J., Sato, J., Kidode, M.: A prototyping of a dialog interface robot at the ubiquitous-home. In: IPSJ Ubiquitous Computing Systems (UBI). Information Processing Society of Japan (IPSJ), vol. 2005, pp. 239–246 (March 2005)
4. Apple Computer Inc.: Netboot and Network Install, `http://www.apple.com/server/macosx/netboot.html`
5. Sartatzemi, M., Dagdilelis, V., Kagani, K.: Teaching Programming with Robots: A Case Study on Greek Secondary Education. In: Bozanis, P., Houstis, E.N. (eds.) PCI 2005. LNCS, vol. 3746, pp. 502–512. Springer, Heidelberg (2005)
6. The LEGO Group, Mindstorms Education, `http://www.legoeducation.jp/mindstorms/`
7. Tamada, H., Ogino, A., Ueda, H.: Robot Helps Teachers for Education of the C Language Beginners. In: Jacko, J.A. (ed.) HCI International 2009. LNCS, vol. 5611, pp. 377–384. Springer, Heidelberg (2009)

A Generic OSGi-Based Model Framework for Delivery Context Properties and Events

Jaroslav Pullmann[1], Yehya Mohamad[1], Carlos A. Velasco[1], and Stefan P. Carmien[2]

[1] Fraunhofer Institute for Applied Information Technology FIT,
Schloss Birlinghoven, 53757 Sankt Augustin, Germany
{carlos.velasco,yehya.mohamad,jaroslav.pullmann}@fit.fraunhofer.de
[2] Tecnalia Research, Paseo Mikeletegi 7, Parque Tecnológico, E-20009 San Sebastián, Spain
stefan.carmien@tecnalia.com

Abstract. Content adaptation systems rely on standards-based modeling of user needs and preferences, rendering platforms, assistive technologies and other relevant aspects of the overall delivery context. Despite their differing domains, these models overlap largely in respect of their digital representation and handling. We present hereby our work on a generic model framework exhibiting a novel set of features developed to tackle commonly found requirements in the area of user and delivery context modeling.

Keywords: delivery context model, user preferences, content adaptation, OSGi, reactive systems.

1 Introduction

The recurrent requirements for user, device and delivery context modeling has led to the development of a generic model framework based on the OSGi module runtime environment for Java. The OSGi platform gained a wide acceptance as flexible deployment model of dynamic modular applications (bundles) ranging from mobile appliances to enterprise deployment scenarios. Our framework comprises a set of OSGi bundles exposing services for model maintenance, querying and dynamic management of data sources (sensors) used to populate the models. The underlying model representation allows for storage of hierarchical, ordered, highly structured data independently of their underlying data-model – RDF triples or XML info sets.

2 Generalized Model Representation

2.1 Abstract Data-Model

There are various data structures commonly used to encode modeling vocabularies: XML trees (acyclic directed graphs), RDF triple graphs[1] in different serializations,

[1] RDF bindings for UAProf; Device Profile Evolution Architecture, etc.

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 605–611, 2011.

JSON[2], simple property-value mappings (Java property files), etc. To relieve model engineers and users from considering the peculiarities of either data structure and to allow an uniform management and query interface, a *generalized, graph-based storage architecture* has been developed, capable of capturing hierarchical, ordered, highly structured data and adherent metadata. It considerably simplifies the import and aggregation of external profile sources.

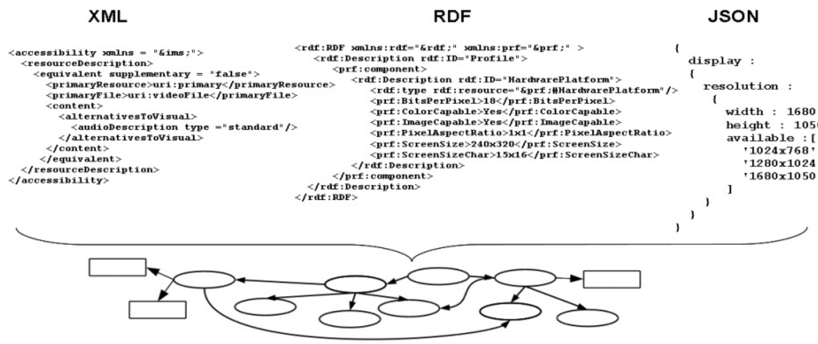


Fig. 1. Generalization of different input data structures

Orthogonal to technical aspects, the model representation is considered generic in respect to data purpose, granularity and repository organization. Thus user preferences, location data, ambient contexts or home appliance properties are equally well captured by named model instances within the same framework supporting cross-referencing and joined queries. No organizational paradigm is enforced. In contrast to traditional single containment hierarchies (collections, directories) individual model instances are assignable to arbitrary number of functional groups that cover a particular shared "aspect":

- common data and metadata (group name, access rules, provenience, etc.) with various insertion semantics (merge, append, replace) and conflict resolution rules
- validity constraints, computed values and reactive rules via *model item properties*
- named queries mapped to virtual top-level nodes of a model

A subset of the XPath 2.0[3] language has been adopted to navigate and query the repository. Path expressions are the means of selecting nodes, defining model item properties and linking them to individual models or groups. They leverage extension functions e.g. to retrieve model instances:

```
sys:model("pc")/display/resolution/width
```

The repository management internally employs and exposes a broader set of events. In addition to monitoring life-cycle stages (created, updated, deleted), value change and (in)validation of nodes, their usage context is considered as well. When external clients subscribe, repository events or input sources (sensors) are added, removed or send they heartbeat signals, and corresponding events are triggered. There

[2] http://www.json.org/
[3] http://www.w3.org/TR/xpath20/

are dedicated path functions to test the existence (a) or to access payload of events raised on a context node <path> (b). These functions are used in trigger conditions of rules (see below):

```
a) boolean on:value-changed(<path>)
boolean on:subscription-added(<path>)
b) String get:subscription-added-by(<path>)
<T> get:value-changed-from(<path>)
<T> get:value-changed-to(<path>)
ValueChangeEvent get:value-changed-event(<path>)
```

2.2 Model Item Properties

The validation and processing of model nodes has been extended in spirit of XForms model item properties[4]. They pertain to a node set of an individual data instance. By means of model groups we extended their applicability to a set of comparable model instances as well. Models inherit group's item property definitions and may additionally define local ones.

A schema-based validation is not enforced. It depends on the original data model (XML Schema, RDF Schema, JSON Schema[5]), whereas validity conditions might by specified in a generic way by *constraint expressions*. They further benefit from the use of functions, node's context and its recent value, which makes them appropriate to express assumptions on value's correctness and plausibility:

```
Node user = registry.getModel("userId29");
user.select("//age").addConstraint(". > 18")
```

Nodes of one or multiple model instances might be correlated or have assigned values by a *calculation expression*. These declaratively state a dependency between the computed node, its input nodes and scalar arguments, which is automatically maintained across model updates.

Finally, reactive behavior can be associated with nodes by means of *rule expressions*. This comprises model updates and invocation of internal methods or external services whenever rule's trigger condition becomes satisfied. In contrast to approaches like [1] restricted to pattern matching on node's value the whole range of event-test and retrieval functions is available for the definition of trigger conditions.

```
Node device = registry.getModel("deviceId6");
          model.addRule("abs(
          get:value-changed-
event('/monitor/temperature')/from-     get:value-
changed-event('/monitor/temperature')/to)
          ge 20",// suspicious temperature
          notify("admin@mydomain.com")
          );
```

4 http://www.w3.org/TR/xforms/#model-xformsconstraints
5 http://tools.ietf.org/html/draft-zyp-json-schema-03

Since this approach operates on recent values only we plan to introduce support for recognition of their continuous change over a time frame by means of complex event processing[6].

3 Architecture of the Model Framework

The generic model repository is the core service provided by the framework. Thanks to the R-OSGi[7] platform extension it is transparently exposed to local and external clients. In addition to location transparency there are obvious benefits of using an OSGi runtime container for application development:

- emphasis of the service interface: depending on the deployment scenario and the capabilities of the underlying device the service might be backed by different implementations while retaining the same interface.
- modularity, dependency management and hot deployment: explicit statement of dependencies at code (package) and service level along with their automatic management through the Service Component Runtime (SCR) yield to a compositional approach of building scalable, layered applications.
- mature standard service interfaces and various open source implementations.

Figure 2 outlines the architecture elements and their inter-dependencies. Within the upper layer reside framework clients like content adaptation services leveraging a graphical user interface to retrieve and manage their profile data and various sorts of sensors. The data exchange services map between the internal graph representation and the exchange formats. The generic model service employs an event bus service for internal distribution and processing of repository events. The subscription service manages event subscriptions and routes events to their respective recipients.

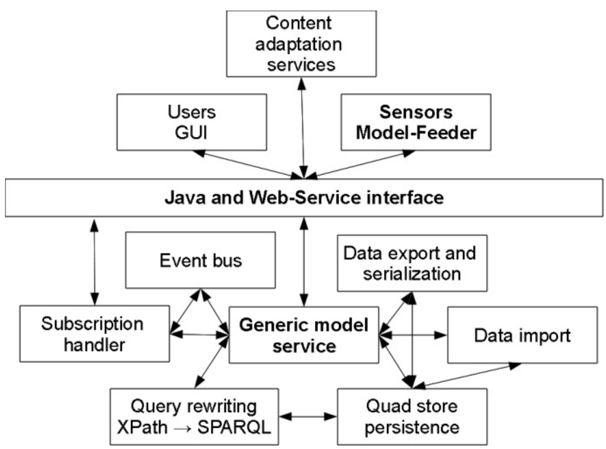


Fig. 2. Architecture components

[6] http://en.wikipedia.org/wiki/Complex_event_processing
[7] http://r-osgi.sourceforge.net/

The remaining sections focus on two practical aspects of model management that were of particular importance in our projects – implementation of context-aware input sources for (device) models and light-weight remote interfaces to model services.

3.1 Dynamic Property Sources

A novel aspect of recent technologies like DPE[8] and DCCI[9] is their explicit coverage of dynamic properties of the delivery context [2]. Some examples are:

- system parameters: volume, display geometry and orientation
- system resources: memory, CPU usage, network bandwidth
- active processes of assistive technologies: screen readers, magnifiers
- attached hardware: Braille keyboard, USB devices

The actual sources of this information, their implementation, interaction with the host environment or considerations on quality of service parameters (sampling rate, overall reliability) seem to be out of the scope and are not discussed at all.

For the purpose of acquiring runtime parameters of host devices we leverage the OSGi runtime and a set of detection bundles. These consist of Java code, shell scripts, native libraries and other resources. Every bundle registers an implementation of the *ModelFeeder* interface which is expected to update model locations listed in the configuration property "exported-properties" by invoking methods of the *Repository* service. The component manifest may specify dependency on further *ModelFeeder*s providing a filter on their exported properties.

The OSGi Service Component Runtime takes care of enabling any component which mandatory dependencies are satisfied and injected via the named "bind"-method or disabling it, when these become unavailable. The code snippet illustrates a component for detection of the open-source screen reader Orca, that states its dependency on another component supplying a list of active OS processes.

```
public interface ModelFeeder(){
     void setRepository(Repository r);
     void releaseRepository(Repository r);

     // life-cycle methods of this component
     void doActivate();
     void doDeactivate
     …
   }
<scr:component name="component-example"
   xmlns:scr="http://www.osgi.org/xmlns/scr/v1.1.0">
```

8 http://www.openmobilealliance.org/Technical/release_program/dpe_V1_0.aspx
9 http://www.w3.org/TR/DPF/

```
<implementation
class="com.imergo.modelservice.impl.readers.Orca"/>
<property name="model-id" value="localhost"/>
 <property name="exported-properties"
           value="device/sound/volume"/>
 <reference
  interface="com.imergo.modelservice.Repository"
  cardinality="1..1"
  bind="setRepository"
  unbind="releaseRepository/>
 <reference name="linux-helper"
  interface="com.imergo.modelservice.ModelFeeder"
  target="(&(platform.os=linux)(exported-
   properties=device/processes))"
  bind="setModelFeeder"
  unbind="releaseModelFeeder"
  cardinality="1..1"
  policy="dynamic"
  activate="doActivate"
  deactivate="doDeactivate"/>
 />
</scr:component>
```

3.2 Client Interfaces to the Model Framework

When considering remote client interaction with the model framework following interface requirements were tackled:

– simplicity and loose coupling: this requirement has been satisfied by provision of a REST service interface. This allows for retrieval and filtering of models:

```
GET <server>/model/<id>
GET <server>/model/<id>/path/within/model
```

– scalability: the client should not be forced to retrieve data at the level of storage entities (model instance). Either a path-based query should be supplied or a named query stored on the server should be invoked in order to filter the result. Client queries relying on literal path expressions are brittle, since they require a close knowledge of the data model and might accidentally break when the model has structurally changed on the server side. To circumvent this dependency we introduced a named query interface. A query expression is stored on the server, associated with a model or model group and named using the CURIE notation. The client invokes this query as a virtual step within the associated model:

```
GET <server>/models/<id>/my:epxr
```

– proactive (PUSH-oriented): the client should not be forced to maintain a local model cache in order to circumvent network round-trips. Instead it should be able to subscribe for particular event(s) notifications.

```
POST <server>/models/<id>/path/within/model
subscribe=get:value-changed-
event&notify=http://myepr...
```

4 Conclusions

Our generic model framework significantly reduces the effort of managing delivery context models and their dynamic data sources. The data-structure agnostic storage and support for navigational and property-based queries has proven successful in deployment scenarios involving a multitude of modeling vocabularies and use cases. Further developments target at an extension by event and rule-based reasoning via additional OSGi Bundles in order to evaluate streams of user interaction and system events.

References

1. Assad, M., Carmichael, D.J., Kay, J., Kummerfeld, B.: Active Models for Context-Aware Services. Technical Report 594. University of Sydney (2006)
2. Timmerer, C., Jaborning, J., Hellwagner, H.: A Survey on Delivery Context Description Formats – A Comparison and Mapping Model. Journal of Digital Information Management, 8 (1) (2010)

Inclusive Scenarios to Evaluate an Open and Standards-Based Framework that Supports Accessibility and Personalisation at Higher Education

Alejandro Rodriguez-Ascaso, Jesus G. Boticario, Cecile Finat, Elena del Campo, Mar Saneiro, Eva Alcocer, Emmanuelle Gutiérrez y Restrepo, and Emanuela Mazzone

aDeNu Research Group, Artificial Intelligence Department, Computer Science School, UNED, c/ Juan del Rosal, 16. 28040 Madrid, Spain
{arascaso,jgb}@dia.uned.es, cecile.finat@bec.uned.es, {mcampo,masterdiscap}@psi.uned.es, ealcocer@bec.uned.es, {egutierrez,emazzone}@dia.uned.es

Abstract. An extensive number of students with disability who are increasingly choosing either distance or blended learning are in the need of accessible learning contents and services, as well as of psycho-educational support. To cope with their needs a general framework implemented through a standards-based open architecture focused on supporting inclusive learning at Higher Education has been developed in the EU4ALL project. The framework is implemented through an open service oriented architecture that can be integrated into current learning platforms. Services involving stakeholders (students and professionals) to attend users' needs are provided through standards-based framework components, which combine user-centred design tools with modelling techniques based on the state-of-the-art on semantic web technologies. The framework, the architecture and its components are being evaluated at large, medium and small universities to show their flexibility and appropriateness. In this paper, after introducing the framework we focus on the evaluation scenarios at one large university and their first evaluations outcomes.

Keywords: Accessibility, personalisation, adaptation, learning, higher education, information and communications technology, ICT, learning materials, virtual learning environments, VLE.

1 Introduction

Education should be able to attend specific learner's needs. However, students and professionals with disability have problems in accessing learning because of the diverse barriers they may encounter. In fact, while many physical barriers have been removed in face-to face education institutions, Information and Communications Technology (ICT) services are still not fully accessible to an extensive number of students who are increasingly choosing either distance or blended learning. This is

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 612–621, 2011.

especially worrying in countries such as Spain, in which roughly 50% of students with disabilities adopt this form of learning.

To tackle these issues, and provide the necessary support from the technological and psycho-educational viewpoint, the EU4ALL European project (IST-FP6-034778) has developed a general and flexible framework to support the needs of inclusive learning scenarios [1]. This framework defines and implements an open and extensible architecture of services for Accessible Lifelong Learning (ALL) in terms of standards, whose flexibility and reusability (e.g, three different Virtual Learning Environment, VLE, interoperate with the architecture: Moodle, Sakai, dotLRN) is being evaluated at four different sites: the two largest distance learning universities in Europe, OU (Open University in the UK) and UNED (National University for Distance Education in Spain) with roughly 10.000 students with disabilities each; a medium-size site, UPV (Polytechnic University of Valencia) and a smaller site, UPLeiria (Polytechnic Institute of Leiria).

The framework consists of several standards-based components, designed according to a service oriented architecture (SOA, defined through WSDL specifications and which can be integrated into different VLEs): The user model component (UM) follows a standard way, via ISO/IEC 24751, see [2], to define the "user with special needs" profile, which addresses users' interactions needs. This component is involved in most framework adaptive features and is integrated into the VLE and most services. The Content Personalisation component (CP) is in charge of finding alternatives to a given resource and evaluating which one is the best available alternative (according to ISO/IEC 24751-3) to suit user's interaction needs (expressed in compliance with ISO/IEC 24751-2). The eServices Server (ESS) provides a general and flexible approach to support workflows and tailor services for diverse environments. Thus, processes present in institutions of distance learning are modelled as executable workflows (eServices) allowing for interaction between automatic services and end users, thanks to generated multi-modal user interfaces. The Metadata Repository (MR) isolates metadata from contents themselves thus supporting the CP in a general and standard-based way. Finally, the Recommender System (RS) selects and delivers the appropriate recommendation to a given learner in a given context to provide a personalized guidance based on educational criteria following a rule-based approach [3].

From the outset, EU4ALL lifecycle was designed with the premise of producing outcomes that really meet the needs of targeted users. With this objective in mind, and consistently with the principles proposed by the ISO 9241-210 standard [4], EU4ALL has focused on users since the very beginning of the project in two ways: a) Design was based upon an explicit understanding of users, tasks and environments. Stakeholders who have a significant role in HE accessibility have been identified. Such information was collected in EU4ALL through different methodologies including questionnaires for students and staff, interviews and focus groups, and reviews of online discussion forums and mailing lists relating to people with disabilities in universities. User's requirements have been concretized in a set of use cases, which have lead to specification of services. Furthermore, a conceptual model was designed based on the analysis of the data gathered during that initial phase of the project. Such data model describes how people working in HFE institutions think about accessibility and what are the broad areas of barriers and solutions in current practice. b) The design was driven and refined by user-centred evaluation through an iterative process. Both experts in accessibility and usability as well as end users have taken part in the evaluation of EU4ALL prototypes. Accessibility and usability problems in the first project prototype have been identified and then fixed, leading to the second version of EU4ALL services. Second version of the prototype is currently going through a large-scale evaluation with end users with different types of impairments in the four EU4ALL pilot sites.

There are related projects focused on adaptive and accessible learning, see [5, 6] and other service-oriented architectures, such as the e-Framework for Education and Research [7], the Open Knowledge Initiative [8], and the Fluid Project [9]. However, they do not implement HE services that are both (1) being evaluated in real-life HE settings with hundreds of students and (2) supported by a standards-based service-oriented architectures as in the EU4ALL project.

In this paper we discuss the services provided at UNED's evaluation scenarios and the preliminary results of the evaluation activities. This paper is organized around three main topics. First, we introduce the problem and the framework along with the user centred approach are introduced. Second, we describe the scenarios of the evaluation are described showing the services and the stakeholders involved. Third, we present the consecutive rounds of the evaluation. To conclude, a brief discussion on the main issues and conclusions are provided.

2 Evaluation Scenarios at UNED

Following a scenario-based approach [10], UNED has illustrated how end users will interact with EU4ALL services in the context of real, web-based courses. Such scenarios are based on the set of EU4ALL's use cases, which have been adapted for reflecting UNED's real context. These scenarios have informed the process of designing the evaluation activities to be undertaken at UNED's pilot site, where students and professionals make use of EU4ALL services. More than one hundred users are to take part in the pilot site in total.

Stakeholders taking part in the UNED evaluation scenarios include students and professionals: Lecturers, who are supported in producing accessible materials and are provided with tools for supervising course's accessibility before it is offered to students; Students, who are allowed to express their accessibility needs through the system as well as to provide their feedback on system's behaviour; Disability officers who support students by assessing their needs, and serve as a liaison between students and other university professionals to address any problems caused by inaccessibility of activities or materials; Transformation officers who work on the adaptation of materials, in co-ordination with lecturers and librarians; Librarians that support the tagging and management of learning materials in electronic repositories.

2.1 Needs Assessment Service (NAS)

David has just registered for his first year at university to read Physics. He is asked to fill-in a questionnaire in the VLE preferences section about his preferences to access learning content and about the assistive products he may use in daily work activities.

Within the accessibility section there is a question about assistive technologies used and David selects Screen Magnifier. The NAS also offers the possibility of being personally assessed by professional experts. When asked about accessibility preferences for learning contents, he does not know what DAISY is, and contacts the Disability Unit to get information on this.

Enric is a student with hearing impairment who has just registered at UNED. Through the accessibility section of the VLE preferences section, he chooses subtitles as text alternatives to audio contents in multimedia material. The Disability Unit also informs him of the existence of RF transmitters at the Study Centres, and how to request one of them when attending face-to-face tutoring sessions.

2.2 Authoring Support

Sarah is a UNED lecturer authoring new material. In order to create new material for next year's course, Sarah runs eXe (an Open Source authoring application to assist teachers and academics in publishing web content without the need to become proficient in HTML or XML markup) on her Desktop PC and creates a new project. For each new section of the material, she selects a template (iDevice, in eXe terminology) and does the editing. For one of the sections, she searches for some multimedia objects, i.e. TV and radio programmes produced by UNED, and includes them in the project. She also searches for an image to be included in the project. She is asked to enter alternative text description of this image. Authoring support also includes resources to assist authors for searching/creating other types of accessible resources (e.g. mathematical content, multimedia materials).

2.3 Resource Accessibility Information Service (RESACCINFO)

After adding any new material to the course space at dotLRN, Sarah may request the RESACCINFO service. This service provides a report on the accessibility of the material. Another way to offer this service to lecturers is by reminding them the need to ensure the accessibility of materials for an approaching course. The reminder includes a recommendation to evaluate the accessibility of these materials through the RESACCINFO service. The report provided by the RESACCINFO service includes the list of transformations the materials need to go through, according to their accessibility features and the students' preferences available at the system (e.g. DAISY alternatives, audio-description, subtitles). For those materials for which there is no accessibility information available within the system, the service prescribes a full accessibility evaluation to be performed by accessibility officers.

2.4 Resource Adaptation Management Service (RESADAPT)

The report obtained from the RESACCINFO service may indicate that there are some pending transformations. Sarah, the lecturer, requests the transformations needed by calling the RESADAPT service. The corresponding transformation officers receive a notification indicating where the transformation is requested. This notification contains information such as the URI of the original learning object, the transformation/s needed, comments issued by the lecturer, contact information of the lecturer requesting the transformation, suggested date - when the transformed material is expected to be ready, etc. Transformation officers work on their transformation (e.g. transforming a whole SCORM course into a DAISY book, adding subtitles to a video, writing the transcript of a UNED radio programme, etc.). If any clarification issue arises, they may contact the lecturer requesting the transformation. When the transformation is finished, each of them go to the URI they received in the request and follow instructions to upload the transformed material. Then, the material is tagged by the librarians and uploaded to the university digital repository. Lecturer Sarah, the lecturer, is notified about the availability of the materials.

2.5 Material Feedback Service

Through UNED's VLE, Enric (a hearing impaired student) accesses a SCORM course with some videos in it. Although Enric's accessibility preferences included captions for videos, they are not available in this course. At any time during the SCORM session, the student can report back on accessibility problems he has encountered while browsing the course. Then, the module's lecturer receives an email where she is informed of the accessibility problem and of the student's comment. The email includes a link to the RESACCINFO service, where all the module's materials are evaluated and the pending transformations reported.

David has very low vision. He is using a DAISY book he obtained from the VLE. While reading it, he has found that there is a non-intelligible description of a diagram contained in this book. He goes to the VLE space of this course where there is a link to the material feedback service, through which he reports the problem. This is notified via email to the lecturer in charge of the course. The email includes a link to the RESADAPT service through which the lecturer requests a revision of the DAISY book. The RESADAPT service allows the lecturer to include text with a reference to the diagram in the DAISY book that does not have an appropriate text description. Furthermore, at the end of the term the system sends an email to all the students inviting them to provide feedback on the accessibility of materials they have used through that period.

2.6 Adaptive Psycho-educational Support Service

Jon is a tutor for the "Discovering UNED's learning platform" course. When the enrolment process is over, Jon receives a notification with specific information about accessibility preferences and psycho-educational needs provided by the students and collected by the NAS service. One of the students has a hearing impairment. The APES service issues a recommendation containing useful information and support links to assist the tutor in order to address the psycho-educational needs of this learner. The information consists of a description of the main difficulties that the learner might find when using a Virtual Learning Environment (e.g. reading difficulties). The tutor also receives guidance on how to produce adapted learning materials and to optimise the student's performance (e.g. creating a glossary, creating alternative activities, adapting text in an easy reading format, using pictures to illustrate concepts, etc). He is also provided with a contact address for the UNED Applied Psychology Service to get more specific support if needed.

Enric is student with hearing impairments. He is provided with specific support as the course begins (glossary, dictionary of synonymous, tutor supporting him on comprehension and writing difficulties through a forum, learning materials enriched with diagrams). When users encounter an impasse that has not been covered by the design, dynamic support can be provided through a recommendation system (e.g. the system can look for students with a similar profile who overcame difficulties when performing the same activity and recommends the current student to do the action that students in similar conditions previously did) or through direct student-tutor communication.

3 Evaluation of the Needs Assessment Service

A small case user study was carried out to evaluate the user interfaces (UIs) designed in the UNED first prototype of the NAS service. The study consisted of a controlled experiment where each participant performed three tasks within the dotLRN based EU4ALL prototype. Methodology proposed in [11] guided the design of the evaluation methodology.

3.1 Description of the Procedure and of the Pool of Users

Several individual sessions were conducted during two weeks, in December 2009. A session consisted of a pre-study questionnaire which included –in addition to standard demographic questions– questions about computer literacy and ATs daily used, an introductory talk to EU4ALL project, an extensive explanations on each of the tasks to be realized during the session, the performance of 3 tasks (each one followed by a questionnaire about ease, amount of spent time and information included related to the task). The 3 tasks to be performed by the participant on the prototype were:

- T1–Preferences for accessible content. Participant fills in a web-form to set the adaptations she needs when accessing different types of learning content objects by the LMS (UI is presented in Figure 2).
- T2–Assistive technologies. Participant completes a four-page web survey to report which assistive technologies she normally uses with her computer.
- T3–Learning style. Participant answers a web test based on Felder's model [12] to know which kind of learner she is in four dimensions: active/reflective, sensing/intuitive, visual/verbal and sequential/global.

The pool of users included 8 students (5 male, 3 female), who were mainly recruited through the Community for Attention to Disability at UNED [13]. Their studies included law, business, psychology, spanish philology, education and PhD. 3 participants (37,5%) were under 30s, 2 (25%) were between 30-45, and 3 (37,5%) were between 45-65. Their computer literacy vary among poor (25%), average (25%) and excellent (50%). When asked how many hours they use internet per week, 1 participant (12,5%) reported between 1 and 5, 4 participants (50%) reported between 6 and 15, and 3 participants (37,5%) reported more than 15. All of them present some kind of impairment: visual (50%), auditory (12,5%) or upper-limbs mobility (37,5%). Most of them (87'5%) normally use one or more assistive technologies – SW and/or HW – with their computers.

Half of the students came to our lab to participate in the experiment, while the other half chose to have our team going to their houses. During the experiment, 1 participant used enlarged fonts, 2 participants used a screen reader (one of them in combination with a braille display), 1 participant used a screen magnifier/reader and high contrast color schemes, and 1 participant used an adapted keyboard and an adapted mouse.

3.2 Measures and Results

During the controlled experiment, for each task we measured three objective indicators and four subjective ones. The objective indicators were: (a) Number of errors leading to adaptation mismatch (i.e., mismatches between user's actual adaptation and those adaptations selected through the system); (b) Number of times the user asks for help on how to proceed with the interaction in order to complete the task; and (c) Time to complete the task. The subjective indicators were based on the user's opinion about: (a) the easiness to complete the task, (b) the appropriateness of the amount of time spent in performing the task, (c) the suitability of the information provided; and (d) the suitability of the system for allowing the user to describe his/her adaptations (content enhancements/alternatives) and ATs. The results of the data analysis for objective and subjective measures are presented correspondingly in Table 1 (in terms of average, standard deviation, minimum and maximum value) and in Figure 1 (where the user rated each topic on a 7-point Likert scale, where 1= "Totally Disagree" and 7= "Totally agree").

Table 1. Objective Measures Summary

	Errors				Help				Time (seconds)			
	AM	SD	Min	Max	AM	SD	Min	Max	AM	SD	Min	Max
T1	0,88	1,36	0	4	0,38	0,7	0	2	102,13	74,44	17	229
T2	0,38	0,48	0	1	0,38	0,48	0	1	238,25	105,17	123	444
T3	0	0	0	0	0,25	0,43	0	1	421,6	237,09	200	865

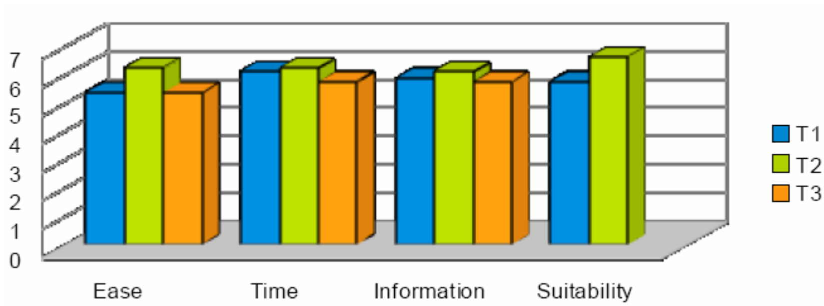


Fig. 1. Subjective Measures Summary

3.3 Lessons Learned and Ongoing Evaluation Activities

Although the number of users taking part in the first round of evaluation sessions was not statistically relevant to identify specific problems for specific AT users, we believe some qualitative conclusions can be drawn. Users were more familiarized with ATs (at least they recognized those that they use everyday) than with content enhancements/alternatives. During the experiment, users encountered problems to describe their preferred accommodations as far as content adaptation is concerned. The design of the form in its initial version followed closely the structure and the options contained in the ISO / IEC 24751-2. However, neither ease of understanding nor effectiveness were enough cared as far as such a form was going to be used by students - generally not technical experts.

Final evaluation of all dotLRN's prototypes is being carried out currently at UNED's pilot site. Regarding the NAS service, lessons learned from the evaluation of the first prototype were used to improve the user interface of the second prototype. The former included just three categories of adaptations: (1) for visual content (audio-description and long-description as selectable options), (2) for textual content (Daisy as selectable alternative), and (3) for auditory content (transcript or sign-language as selectable options). The latter includes five categories, terms explanations and examples as follows:

- Adaptations for text (books, articles and text documents available as Word or PDF files, etc.): Spoken texts (texts that can be heard, such as Daisy or talking books).
- Adaptations for audio content (e.g. radio programs): Transcript (text document which reproduces the audio content) or sign language.
- Adaptations for images (e.g., diagrams or figures in a document): Alternative text (description of the image through a text, so text and image are equivalent).
- Adaptations for animations (sequence of images allows the presentation of a moving scene, for example an animation of the solar system): Alternative text (description of the animation with a text, so text and animation are equivalent).
- Adaptations for film and video (for example, videos or television programs): Audio-description (audio description by a voiceover of what happens in the picture at times when there is silence or pauses in the dialogue) or Subtitles (text that appears in the bottom of the image, transcribing the dialogues, describing the sounds of the film, etc.) or Transcript (text document which reproduces the audio from a video).

Furthermore, the new experiment aims to investigate on how the original learning object and their adaptations should be presented by the system. For that purpose, two web-based SCORM courses have been adapted from existing UNED learning materials and are being used in the evaluation of the second prototype. The courses include a variety of learning contents and objects, such as text (including mathematical notation), audio and multimedia, together with their corresponding adapted contents:

- Transcripts and sign language interpretations as adaptations for audio content (e.g. radio programs).
- Text as adaptation for diagrams (e.g., diagrams or figures in a document).

- Subtitles and sign language interpretation as enhancements for multimedia contents; Transcripts as alternatives to multimedia contents.
- Daisy as an alternative for text (in fact, Daisy is used as an alternative for a whole SCORM course, with any audio and multimedia being replaced by their corresponding transcripts).

All the contents comply with WCAG 2.0 in conformance level Double-A +. Contents requiring mathematical notation have been coded through MathML presentational.

4 Conclusions and Future Work

Authors presented the UNED's setting of the project EU4ALL, which consists of scenarios evaluated within the large-scale evaluation process involving over a hundred users covering all the required stakeholders (students, teachers, disability officers, technical and administrative personnel…).

We developed services of diverse nature and complexity, from workflows to those providing immediate response. Innovation issues draw on personalisation features (i.e. attending user's needs in a given context and time) based on user modelling techniques and standards, which require interoperability between different architecture components (Recommender System, Content Personalisation, Metadata Repository, User Model, Device Model, VLE, etc.). For instance, a semantic-oriented recommendation system is able to capture psycho-educational and inclusion aspects to attend the needs of the students and a standards-based personalised learning design scenario includes different users sharing some learning difficulties (e.g. students with dyslexia, attention deficit, auditory impairment, foreign students) [13]. The scenarios created for evaluating the services involve all the required stakeholders (i.e. students and professionals) and take into account the different tasks and technical/educational needs that were previously identified through one of the largest known users' requirements elicitation efforts in the field [14].

In this paper we have introduced the evaluation scenarios at UNED and the preliminary results of the evaluation activities. Following the project life-cycle we are currently involved in the large-scale evaluation, which is expected to provide valuable outputs on the solutions provided and related open issues.

Acknowledgments. The authors would like to thank the European Commission and the Spanish Government for funding the projects of aDeNu group that have supported this research work. In particular, EU4ALL (IST-2005-034778) and A2UN@ (TIN2008-06862-C04-01/TSI). The authors would also like to thank both EU4ALL and A2UN@ consortia.

References

1. Santos, O.C., Boticario, J.G., Raffenne, E., Granado, J., Rodriguez-Ascaso, A., Gutierrez y Restrepo, E.: A standard-based framework to support personalization / adaptation and interoperability in inclusive learning scenarios. In: Lazarinis, F., Green, S., Pearson, E. (eds.) Handbook of Research on E-Learning Standards and Interoperability: Frameworks and Issues, IGI Global (2010)

2. ISO/IEC 24751-1:2008 Information technology – Individualized adaptability and accessibility in e-learning, education and training
3. Santos, O.C., Boticario, J.G.: Modeling recommendations for the educational domain. In: 1st Workshop on Recommender Systems for Technology Enhanced Learning (RecSysTEL 2010) In conjunction with 4th ACM Conference on Recommender Systems (RecSys 2010) and the 5th European Conference on Technology Enhanced Learning (EC-TEL 2010), Barcelona, Spain (September 2010)
4. ISO 9241-210:2010 Ergonomics of human-system interaction – Part 210: Human-centred design for interactive systems
5. GRAPPLE project (2011), http://www.grapple-project.org (retrieved January 27, 2011)
6. FLEXO project (2011), http://www.ines.org.es/flexo (retrieved January 27, 2011)
7. E-Framework (2011), http://www.e-framework.org/ (retrieved January 27, 2011)
8. OKI (2009), http://www.okiproject.org/ (retrieved January 27, 2011)
9. Fluid (2011), http://fluidproject.org/ (retrieved January 27, 2011)
10. Carroll, J.M.: Making Use: scenario-based design of human-computer interactions. MIT Press, Cambridge (2000)
11. IBM Accessibility Center. White paper: conducting user evaluations with people with disabilities, http://www-03.ibm.com/able/resources/userevaluations.html (retrieved January 27, 2011)
12. Felder, R.M., Linda, K.: Learning and Teaching Styles In Engineering Education Silverman. Engr. Education 78(7), 674–681 (1988)
13. del Campo, E., Saneiro, M., Santos, O.C., Boticario, J.G.: Psycho-educational support for students with disabilities in higher education, applied through a recommender system integrated in a virtual learning environment. International Journal of Developmental and Educational Psychology 3, 237–247 (2010)
14. Petrie H., Power C.: Swallow EU4ALL Deliverable. D1.1.3 - Report on ontology of users and their requirements (April 2010) (a final revised version will be delivered based on large scale evaluations)

Relationship between BPM Education and Business Process Solutions: Results of a Student Contest

Werner Schmidt

University of Applied Sciences Ingolstadt, Esplanade 10,
85049 Ingolstadt, Germany
Werner.Schmidt@haw-ingolstadt.de

Abstract. In order to optimize academic and outside-university education in BPM it is necessary to know how educational aspects influence the quality of BPM solutions. Information regarding this topic can rarely be found. In this paper we report on a student contest which was carried out to address this problem and gather related information. An introduction is followed by information on the contest setting, the results and findings, before concluding with a summary and a description of research to be done in the future.

Keywords: BPM, S-BPM, academic education, professional experience, BPM methods, BPM tools.

1 Introduction and Motivation

There seems to be a lack of information about the correlation between the extent of BPM education people have and the quality of BPM solutions they build. Due to economic reasons it is hardly possible to evaluate such relationships in a test environment in daily business. The expenses to have several groups of people work on the same BPM issue and then compare the results usually are too high. So we initiated a student contest with 22 participants as an approach to look at how different levels of BPM education might influence the quality of work results. We hoped to discover positive correlations, but we were also aware of the fact that some factors might have little or no impact on the quality.

As the number of participating students (22) who answered the questions was too low to represent a sound statistical base, the results were interpreted more from a qualitative perspective than from a quantitative. Consequently, we neither computed variances nor correlation measures like covariance, but just compared absolute values and evaluated means.

2 Contest Setting

2.1 Participants

The contest participants consisted of seven teams of two to four members each with a total number of 22 students. We did not select the teams, but recruited them on a first come, first served basis after promoting the contest within the academic community.

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 622–631, 2011.

The teams represented seven German and Austrian universities and different study programs extending from a Bachelor level in Information Science (Inf), Business Informatics (Winf) and Business Engineering (Wing) to a specialized Master level course called 'ERP & Process Management (ERP&PM)'. The students differed in age, in their progress within the programs (semester/year of study), in their professional experience and in their status (either full-time or part-time students).

2.2 Case

The teams' task was to work on a process of the **Intelligence Sonic Systems Corporation**, a fictitious manufacturer and vendor of highly sophisticated time measurement gear (in the picosecond area). This company had developed a revolutionary technology and was continuously expanding. Therefore there was a need to implement an effective and efficient order-to-ship process. The students received a short presentation about the company with some basic information and key performance indicators, as well as a rough process description.

The **process** described the development, production, quality assurance, and shipment of a pico clock device in accordance with the customer needs and agreed upon in the contract management phase (see figure 1). The device as a whole product consisted of a circuit board as the major part, the housing and the documentation.

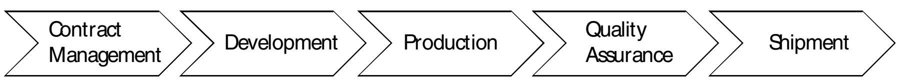


Fig. 1. Contest process

Each step was illustrated in more detail by one slide The process showed real-life characteristics. There was an ideal path if a suitable circuit board was available. There was an exception path if a board needed to be newly developed. In this instance there was another case distinction, depending on whether the new circuit board could be derived from an existing one or whether it had to be developed from scratch. All paths required a multitude of different activities to be carried out with different people involved.

The **assignment** was to analyze, design, validate and implement the process in such a way that it would be executable in a real-life organizational and IT environment. For comparability reasons all teams were to use the same methodology and software tool environment.

The **methodology** chosen was S-BPM as a relatively new, but widely discussed paradigm. S-BPM is characterized by a straightforward approach towards the analysis, modeling, implementation, execution and management of interaction patterns with an explicit focus on stakeholders (subjects, actors) and the interactions they take in order to achieve the process goals. As S-BPM refers to organizations as socio-technical and socio-economic systems, it recognizes current developments in social computing and stakeholder orientation, (re-)aligning human needs and capabilities with artifacts [1], [2], [3], [4]. Thus S-BPM stands for Subject-oriented, Stakeholder-oriented and Social Business Process Management.

The decision for S-BPM also determined the **BPM software** used for the contest. At present the Metasonic Suite is the only software environment that supports the S-BPM methodology. It consists of three major components with the Build module for modeling, the Proof module for interactive online validation and the Flow module as a process engine [5].

As **guidelines** for accomplishing the assignment the teams received the following hints and tips:

- Just as in a real-world situation the given process description might be incomeplete, inconsistent, unclear, or even wrong (or at least partially wrong).
- As a consequence there might be the need for making assumptions (and documenting them) and again questioning them.
- Apply clear methodology (and document steps)
- Proceed step by step, setting well-defined and realistic milestones, save milestone results before continuing.
- Providing an executable process weighs the most among the evaluation criteria (other criteria were not published).

2.3 Time-Frame

Preparation phase. All participants had the opportunity to prepare for the contest by participating in webinars on S-BPM methodology and on the Metasonic Suite (modeling, validation, execution and business objects) used as a BPM tool. The webinars were offered by the tool vendor both as synchronous interactive sessions (5 hours in total) and asynchronous e-learning content, the latter of which included the recorded live sessions as well as a large amount of download material such as method and tool manuals, exercises etc. The teams could install the software on their own computers and then work on exercises.

Contest phase. The contest lasted 1.5 days and took place at SAP Research in Walldorf, Germany. It started on day 1 at noon with an introduction to the overall event (welcome, organizational issues, assignment briefing). The deadline for the solution was announced to be on day 2 at 12:30 p.m. After the introduction the teams worked on the project from 1:00 p.m. until 6:30 p.m., and, if they chose to do so, also after dinner in the hotel for as long as they wished (even the entire night). Official working hours on day 2 were from 8:30 a.m. until 12:30 p.m. After the lunch break the teams explained their solutions in 30-minute presentations to the jury.

While working on the assignment the teams had the opportunity to use wildcards in order to receive support. Wildcards could be used by each team as follows:

- one wildcard for a face-to-face workshop (30 min.) with the Intelligence Sonic Systems project leader as an expert for the process,
- three wildcards for consulting or support sessions (30 minutes each) on methodology and software tool (by phone or on the internet),
- wildcards could be used on day one until 10:00 p.m. and on day two until 12.30 p.m. Usage of a wildcard had to be prescheduled at least 1 hour in advance.

2.4 Assessment

Evaluation of contest solutions. A jury of 4 BPM professionals (2 consultants, 1 CIO, 1 CEO) evaluated the solutions based on the team presentations and on more detailed individual inspection of the solutions afterwards. Criteria were grouped into two main areas: process execution and process and model layout (see section 3.1).

Participant survey. At the end of day 2 we asked the students to fill out a questionnaire consisting of 5 areas of interest. They were asked to answer questions on their academic and professional experience, their BPM education, their BPM knowledge and associated practical experience (concepts, methods, software tools), their S-BPM knowledge and associated practical experience and also on some general aspects of the contest itself (see section 3.2).

3 Results and Findings

3.1 Evaluation Results

Table 1 summarizes the evaluation of the contest solutions by the BPM experts.

Overall result (A). Although points scored were not very high, except for teams on ranks 1 and 2 (A2), all teams did a fairly good job, considering the complexity of the process net of the assignment and the limited time given. According to statements of most participants the process net was by far more complex than what they had ever dealt with before (e.g. small examples in classes like processing a vacation request of an employee). Therefore, modeling and implementing the contest process certainly was a real challenge.

Process Execution (B). In total, the leading teams (ranks 1-2) scored almost three quarters of the points available in the execution section (B11). Beyond a huge gap there were two teams who achieved one quarter of the available points, and these were followed even more distantly by the remaining teams with between 5 and 12 percent. The most successful teams scored 100% in the area of validation and execution of the ideal path (B1-B4). They followed the recommendation to first make the main path work before dealing with exceptions. Teams on the other ranks did not clearly distinguish the cases and generally muddled through all eventualities. As a consequence, they frittered away their time and their solutions were not fully executable, neither in the validation step, nor as a final workflow.

Process and model layout (C). In this section the results did not vary as much as for process execution. Rank 1 as an exception obtained almost 50%, the other six teams between 13 and 27% (C7). For the teams on ranks 1-4 the amount of points achieved was significantly lower than in execution (B2), for the teams on ranks 5-7 it turned out to be the other way round.

This is somehow surprising because one could assume that the better the model is laid out and structured, the easier it is to achieve process execution. A possible explanation could be the special features of Metasonic Proof which allow the instant

Table 1. Selected results of the evaluation of team solutions

Point scores by team on rank …		1	2	3	4	5	6	7	Avg.	Max.
A1	**Total**	51	48	17	16	10	7	6	22.14	75
A2	**Percent**	68	64	23	21	13	9	8	30	-
B	**Process execution**								Avg.	Max.
B1	Ideal path execution in validation module successful	6	6	2	2	0	0	0	2.29	6
B2	Ideal path execution in workflow module successful	12	12	4	4	0	0	0	4.57	12
B3	Exception path execution in validation module successful	1	3	1	1	1	0	0	1.00	3
B4	Exception path execution in workflow module successful	2	6	0	2	2	0	0	1.71	6
B5	Issues like pico engineer, board design etc. solved	6	6	3	1	2	3	0	2.86	6
B6	Input/output mismatch solved	2	0	0	2	0	0	0	0.57	6
B7	Process overlaps solved	6	2	2	2	2	0	0	2.00	6
B8	Database and web integration	6	0	0	0	0	0	0	0.86	6
B9	Overall impression	3	9	3	0	0	0	3	2.57	9
B10	Subtotal	44	44	15	14	7	3	3	18.43	60
B11	Percent	73	73	25	23	12	5	5	31	
C	**Process and model layout**								Avg.	Max.
C1	Clear process structure	2	1	1	0	1	1	1	1.00	3
C2	Short, straightforward process	1	1	0	0	1	1	1	0.71	3
C3	Clear model layout (colours etc.)	2	1	0	0	0	1	0	0.57	6
C4	Clear role model	2	1	1	2	1	1	1	1.29	3
C6	Subtotal	7	4	2	2	3	4	3	3.57	15
C7	Percent	47	27	13	13	20	27	20	24	-

validation of a model in an interactive role play in a distributed computer environment [5]. Thus, model errors can easily be recognized and corrected before validation is run again. This kind of efficient trial-and-error procedure probably helped the teams on the way to process execution, even though their model structure was not clearly defined.

3.2 Survey Results

The following tables and statements give information about the answers the students gave in the questionnaire and relate them to the ranking of the evaluation.

Table 2 contains the questions on **academic and professional experience** of the participants. The content of the study program itself cannot really give an explanation for the ranking results. Only the program of the team on rank 3 (ERP & Process Management) has a dedicated focus on BPM, in all other programs BPM is usually only one subject among a multitude of others.

Number of semesters and age positively correlate with the program level to what we call academic experience. This seems to be a good indicator for the level of quality of the solution. The team on rank 2 as an outlier in this area might have compensated this with professional experience, an area we regarded with D5 and D6. Values here are very high for teams on ranks 2 and 3, and comparably low for rank 1. And although rank 7 does not really fit into the picture, the amount of professional experience also seems to have a positive impact on the contest solution.

Table 2. Selected survey results – Academic and professional experience

Values for team on rank …		1	2	3	4	5	6	7	-
D	**Academic and professional experience**								Avg.
D1	Study program	Wing	Winf	ERP& PM	Winf	Inf	Winf	Wing	-
D2	Program level (Bachelor, Diploma, Master; no.)	D3, BA1	BA3	MA4	BA3	BA2	BA3	BA3	-
D3	Semester	12.3	3.0	10.0	5.0	2.5	5.0	6.3	6.3
D4	Age	25.0	22,0	26.0	22.0	23.5	21.3	22.3	23.2
D5	Professional experience before college (y/n; no.)	n4	n1, y2	n1, y3	n3	n2	n3	n3	-
D6	Part-time study? (hours per week)	2.5	10.3	28.1	0.0	0.0	0.0	6.0	6.7

While table 2 more or less shows hard facts, the following charts mostly express the results of a self-assessment by the participants. We start with the **BPM education** section (see table 3), where we first asked for classes and trainings on BPM inside and outside of the university environment. Then the students were asked to assess themselves with regards to their knowledge and practical experience in BPM in general, established BPM methods and associated software tools. At the end of this section they were asked to state how many processes they had already modeled and how many of these they had implemented as executable workflows.

On an average, the contest participants had attended 1.4 classes with about 40 full hours (E1, E2). The values pretty much vary, ranging from double the mean for the teams on ranks 3 and 4, while two teams reported zero.

BPM hands-on practice shows a quite similar picture with a mean of 0.6 trainings with about 18 hours (E3, E4). Here three teams almost double the average, while two groups did not have any training hours at all. None of the participants have had BPM classes or trainings outside of the university.

Regarding lines E5 and E6, the hours spent in class and in hands-on trainings do not seem to have a clear impact on how the students rated their BPM knowledge and practical experience. The average here was 3.2 and 3.4 (satisfactory to some degree). The exception is the team on rank 3, probably due to the program dedication. There is also no obvious relation between hours of education, self-assessed know-how and the quality of the contest solution (compare values for teams on ranks 1, 3 and 7).

Table 3. Selected survey results – BPM education (knowledge and practical experience)

Values for team on rank …		1	2	3	4	5	6	7	-
E	**BPM education (grades 1-6[1])**								Avg.
E1	No. of BPM classes	1.3	1.0	3.3	3.0	0.0	1.0	0.0	1.4
E2	No. of class hours (full hours)	37.5	30.0	90.0	90.0	0.0	30.0	0.0	39.6
E3	No. of BPM hands-on trainings	0.3	1.0	0.5	0.0	0.0	1.0	1.7	0.6
E4	No. of BPM trainings (full hours)	7.5	30.0	18.8	0.0	0.0	30.0	37.5	17.7
E5	Knowledge in BPM as a concept	3.8	2.7	1.8	3.0	5.0	3.0	3.0	3.2
E6	Pract. Exp. in BPM as a concept	4.0	4.0	2.5	3.0	4.5	3.3	2.3	3.4
E7	Knowledge on EPC/eEPC	3.3	2.0	1.5	5.0	5.5	2.7	5.3	3.6
E8	Practical experience in EPC/eEPC	3.8	3.3	2.0	5.3	6.0	3.0	5.3	4.1
E9	Knowledge on UML	4.3	2.7	2.8	3.3	2.0	3.7	2.3	3.0
E10	Practical experience in UML	4.7	3.3	3.3	3.3	6.0	3.7	2.8	3.9
E11	Knowledge on BPMN	5.0	2.3	3.3	2.3	3.5	4.7	4.7	3.7
E12	Practical experience in BPMN	5.0	3.3	4.0	2.7	6.0	4.7	4.7	4.3
E13	Knowledge on Petri nets	4.5	5.7	5.3	5.7	5.0	6.0	5.0	5.3
E14	Practical experience in Petri nets	4.5	5.7	5.5	6.0	6.0	6.0	5.0	5.5
E15	Knowledge on Powerpoint[2]	3.8	2.7	3.0	5.0	3.5	2.0	1.8	3.1
E16	Practical experience in Powerpoint[2]	3.8	3.7	3.5	5.3	6.0	2.7	1.7	3.8
E17	Knowledge on Visio[2]	2.5	3.7	2.0	4.7	5.5	3.7	4.7	3.8
E18	Practical experience in Visio[2]	2.5	4.0	2.0	5.0	6.0	3.7	4.7	4.0
E19	Knowledge on ARIS[2]	5.0	4.0	3.0	5.7	5.5	3.0	6.0	4.6
E20	Practical experience in ARIS[2]	5.0	4.3	3.5	6.0	6.0	3.0	6.0	4.8
E21	No. of processes modeled already	5.0	6.3	45.5	8.7	5.0	11.7	2.7	12.1
E22	among them: brought to IT-supported execution	4.8	4.0	3.3	1.5	0.0	0.0	1.3	2.1
E23	Ratio E22/E21 in %	95.0	63.2	7.1	17.3	0.0	0.0	49.8	-

In the **BPM method** area our questions covered (extended) Event-driven Process Chains (EPC/eEPC), Unified Modeling Language (UML), Business Process Management Notation (BPMN) and Petri nets. The answers document a relatively low average of knowledge ranging from 3.0 for UML, 3.6 for EPC/eEPC, 3.7 for BPMN and 5.3 for Petri nets (E7, E9, E11, and E13). A peculiar fact is that the rank 2 team rated itself good or satisfactory in the first three methods, the rank 3 team in the first two methods. All other groups stated to be good or satisfactory only in one distinct method, with the rank 1 team as an outlier, stating 3.3 as the best grade for one method and stating less than 4.3 for the others. The students graded their average practical experience with the methods worse than their knowledge of such, but in same order (E8, E10, E12, and E14). There was only one "good" (2.0), reported by the rank 3 team for EPC/eEPC. An impact on the ranking is not visible.

In our experience organizations often not only use **BPM tools** like ARIS for modeling, but also apply presentation and drawing tools like MS Powerpoint and MS Visio for the same purpose. So we asked for the students' skill level with regards to these applications, as well as for ARIS and others specifically named. According to E15 to E20 only two teams (ranks 6 and 7) were quite familiar with MS Powerpoint for modeling, another two groups (ranks 1 and 3) with MS Visio. For ARIS only teams on ranks 3 and 6 described their knowledge and practical experience as satisfactory to some degree, their respective averages of 4.6 and 4.8 indicate limited know-how in general. The participants did not name other software tools they had already used. The averages for the three tools mentioned generally show limited know-how, the individual values do not show any significance for the ranking.

Regarding the contest assignment, we assumed that the number of processes the students had already **implemented as IT-supported workflows** could be an indicator for the quality of their solutions. So we asked for this information as well, in addition to the number of **processes modeled**, and then computed the **ratio**. Line E21 shows an average of 12.1 for the number of processes modeled. Taking into consideration the team on rank 3 as an exception (45.5; dedicated program) brings down the mean to 6.6, with many teams fairly close to average. The small absolute numbers from zero to 4.8 in E22 show little overall experience in implementing processes in IT environments which of course limits the meaningfulness of the quotient in E23. Nevertheless the ratio clearly distinguishes between teams who mainly just modeled processes and those who also implemented a larger number of these in IT. As examples the rank 1 team brought almost 100% to execution, the rank 2 team still almost two thirds. Rank 7 has a ratio of almost 50%, but with only 1.3 processes implemented as workflows.

As the students were required to use the **S-BPM method** and the **Metasonic Suite as** the associated **software tool,** we also asked about their knowledge, experience and education in these areas (see table 4).

Table 4. Selected survey results – S-BPM education (knowledge and practical experience)

Values for team on rank …		1	2	3	4	5	6	7	-
F	S-BPM education (grades 1-6, full hours)								Avg.
F1	Knowledge on S-BPM methodology	3.0	3.3	2.8	2.7	5.0	3.0	2.0	3.1
F2	Pract. Exp. in S-BPM methodology	3.0	3.3	3.5	3.0	5.0	2.7	2.0	3.2
F3	Gained through classes, hands-on trainings and self-study activities	38.5	65.0	16.0	2.7	50.0	1.3	67.5	34.4
F4	Knowledge on Metasonic Suite	3.5	3.0	4.0	3.7	4.0	3.7	2.3	3.5
F5	Pract. Exp. in Metasonic Suite	3.5	3.7	4.3	3.7	4.5	5.0	2.2	3.8
F6	Gained through classes, hands-on trainings and self-study activities	24.0	68.0	25.3	50.0	2.0	71.8	35.4	35.4

The average knowledge on S-BPM was rated satisfactory (3.1), equal to that of UML. The mean for practical experience (3.2) was significantly better for S-BPM than for UML and the other methods referred to. In both categories the team on rank 7 reported a grade significantly above average (2.0), the team on rank 5 significantly below average (5.0). The individual values have no meaning for the contest results.

F3 sheds light on the amount of education on S-BPM the participants have previously had. We aggregated the numbers reported for class hours, training hours and self-study hours into one number. The average of 34.4 compares to 39.6 for BPM

(E2). The individual values generally vary from 2.3 to 67.5 hours with no significance for the ranking. Means for knowledge and practical experience with regards to the Metasonic Suite were 3.5 and 3.8, which means slightly below the values for the S-BPM method and within the range for the other tools, but approximately one grade better than for ARIS. Here again the education hours range from 2.8 to 71.8 and do not reveal a traceable correlation to the quality of the contest's resulting solutions.

Finally, we wanted to know how much time the students actually spent working on the contest assignment, including the hours worked by choice during the night. Also we asked for the participants' estimate of how their know-how of the Metasonic Suite had improved through the interactive application of such during the contest (table 5).

Table 5. Selected survey results – Contest specific data

Values for team on rank …		1	2	3	4	5	6	7	-
G	**Contest-specific data (full hours, grades 1-6, after the contest)**								Avg.
G1	Time spent on solution as a team	76.0	39.0	52.0	52.5	26.0	50.0	48.0	49.1
G2	No. of team members	4	3	4	3	2	3	3	3.1
G3	Average time spent per member	19.0	13.0	13.0	17.5	13.0	16.7	16.0	15.6
G4	Time worked during night	9.5	3.5	3.5	8.0	3.5	7.2	6.5	6.0
G5	Knowledge on Metasonic Suite	2.0	3.0	2.3	2.7	4.0	2.7	1.8	2.6
G6	Pract. exp. in Metasonic Suite	2.0	2.7	2.5	2.7	3.5	2.3	1.7	2.5

While the rank 1 team worked 76 hours in total, which is significantly more than all the others, the team on rank 5 forms the bottom of the table with only 26 hours (G1). The quotient in G3 takes into account the different number of respective team members. This however doesn't substantially change the order. As an indicator of additional effort, G4 shows the time per person spent voluntarily working on the solution during the night (G3 minus 9.5 hours as the formal working time). The results emphasize the fact that the amount of time invested in working on the solution does not seem to correlate with its resulting quality.

In the case of the team on rank 1, the high amount of effort invested in the solution might have helped to compensate for some of the lower values achieved with some of the other criteria where the team only reached average or even below average values.

Spending approximately 16 hours on the contest assignment using the Metasonic Suite seemed to be a good investment for the students. As visible in G5, this led to a significantly improved knowledge (by almost a whole grade), while G6 shows that practical experience was also noticeably increased (by as much as 1.3 grades).

4 Summary and Further Research

Table 6 provides a summary of the findings mentioned in Section 3. Moreover, there are some other noticeable facts: The team on rank 1 was below average in many areas (e.g. least knowledge and practical experience on methods). There is no explanation other than probably academic maturity and time invested. On the other end a team which was above average in many areas ended up on rank 7 team, without reasonable explanation. The only inherently consistent result was rank 2 for a team which reported individual values above average in many areas.

Table 6. Summary of Findings

Aspect	Impact on solution quality
Academic experience	Positive impact
Professional experience	Positive impact
Knowledge on BPM methods	Positive impact
No. of processes already brought to IT-supported execution	Positive impact
Hours of academic BPM education	No obvious impact
Experience in BPM methods	No obvious impact
Knowledge on/practical experience in BPM tools	No obvious impact
Hours of S-BPM education, knowledge on S-BPM methodology	No obvious impact
Hours of Metasonic Suite education	No obvious impact
Time invested in solution development as a team	No obvious impact

Even though the results are not statistically proven or representative they give some indications for further research. For example it seems to be worthwhile to further detail the evaluation criteria for the contest solutions as well as some survey questions. This would enable a more differentiated comparison of educational aspects to the associated criteria.

As the self-assessment for BPM education revealed low average grades it could make sense to ask students where they identify potential for improvement (e.g. content, didactics). Another question would be what kind of knowledge and skills had been necessary to build a better solution.

References

1. Fleischmann, A.: What is S-BPM? CCIS, vol. 85, pp. 85–106. Springer, Heidelberg (2010)
2. Schmidt, W., Stary, C.: Establishing an Informed S-BPM Community. CCIS, vol. 85, pp. 34–47. Springer, Heidelberg (2010)
3. Schmidt, W., Fleischmann, A., Gilbert, O.: Subjektorientiertes Geschäftsprozessmanagement, HMD – Praxis der Wirtschaftsinformatik, vol. 266, pp. 52–62. dPunkt, Heidelberg (2009)
4. Institute of Innovative Process Management, `http://www.s-bpm-one.org`
5. Metasonic AG, `http://www.metasonic.de/en/metasonic-suite-overview`

Design of a Multi-interface Creativity Support Tool for the Enhancement of the Creativity Process

George A. Sielis[1], George A. Papadopoulos[1], and Andrina Granić[2]

[1] University of Cyprus, Department of Computer Science
P.O. Box 20537, 1678 Nicosia, Cyprus
{sielis,george}@cs.ucy.ac.cy
[2] Department of Computer Sciences, Faculty of Science, University of Split
Nikole Tesle 12, 21000 Split, Croatia
andrina.granic@pmfst.hr

Abstract. This work examines the influence of the different working environments during the execution of a creativity process. The selection factor used for the decision is each user's learning style. For the examination of how this is perceived by the users and to test how this influences the creativity process within a Creativity Support Tool a prototype has been developed. The paper describes the prototype as well as its pilot testing by a group of users.

Keywords: Learning Styles, Creativity, Creativity Support Tools, Context.

1 Introduction

Creativity Support Tools (CSTs) are designed to facilitate users in generating innovative ideas related to existing open problems, or to new problems that appear on the scene. In the existing literature, we can find CST design approaches that simulate the creativity while supporting some of the most known Creativity Techniques, like brainstorming. The research interest in the plurality of these approaches focuses on finding methods that process the content, and use AI techniques to produce innovative ideas or creative outcomes automatically. The current work attempts to examine the creativity process from a different point of view, and this is, the examination of how different interfaces and different types of interactions influence the user's creativity effort during the creativity process.

Creativity is an ability carried by all people, but the level of creativity that each person has is different. Levels of creativity depend on how the person's creativity has been cultivated and amplified. Plucker and Beghetto [1] define creativity as "*the interplay between ability and process by which an individual or a group produces an outcome or product that is both novel and useful as defined within some social context*". In this work we use the latter definition for the development of a multi-interface context aware CST prototype which enhances the creativity process. The proposed approach elaborates the selection of a working environment based on the user's learning style with the belief that this can enhance the user's creativity. The prototype of this work aims to examine the influence of the user's learning style as a

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 632–641, 2011.

decision factor for the most suitable interface that the user can work with. We also describe the design methodology used for the implementation of a context aware CST, adapted to the user's learning style.

The exploration of learning process has inspired development of a variety of learning style models and categorizations. Kolb's learning style theory [2] proposes that learners can be classified into convergent learners, divergent learners, assimilators and accommodators. Honey and Mumford [3] present another interpretation of Kolb's theory and categorize learners into activists, pragmatists, reflectors and theorists. Gardner's Multiple Intelligences Theory [4] proposes seven intelligences: verbal, logical, visual/spatial, musical, interpersonal, intrapersonal and bodily. Felder-Silverman Learning Style Theory (FSLSM) [5] characterizes each learner according to four dimensions: active/reflective, sensing/intuitive, visual/verbal and sequential/global.

Although the learning style theories have been applied in educational environment widely, only recently researchers have started to undertake them in the computer-supported ones. The learning styles proposed by Gardner [4] in the work for Multiple Intelligence Theory and the related tests are considered as the most relevant to this work in relation to the use case scenario which is described in the following sections. Gardner proposes seven learning styles which can be simulated within a Creativity Support Tool representing a real case scenario. In our belief, for the implementation of a CST prototype which simulates a Creativity environment in a realistic way, the Gardner's tests appear to be the most simplified but at the same time accurate tests for the determination of a user's learning style.

In particular the document is organized as follows: Section 2 elaborates the motivation for the current study through a use case scenario; Section 3 describes a proposed context aware multi-interface Collaborative Creativity Support Tool (CCST); Section 4 describes the pilot testing of the prototype and finally the paper ends with the conclusions in Section 5.

2 Motivation

The perception of creativity in real life is correlated to innovation and exploration. A very common use of creativity in real life is the conversation. When two or more people exchange opinions on a subject, then this conversation consists of a creative process. Therefore, creativity is the set of ideas, opinions and conclusions that arise from stimuli that environ the participants of a creative process. In real life, stimuli are defined by the environment where a creative process is taking place. For example, if a group of people walk in the street, they will change their discussion subjects according to what they see during their walk. An artist who might see an extraordinary object will inspire her future work, while a group of scientists will use images, texts or figures to solve a difficult problem. In any of the aforementioned scenarios, the environment where the creative process takes place influences how each person perceives creativity, and how she participates in a creative process.

This study aims to explore how realistic creative processes can be supported by means of ICT and more specifically, to examine how the use of multiple interfaces during a creative process influence the creative productivity of the user. The motivation for this work emerges from the examination of the existing Creativity Support Tools from our previous work [6] from which none of the examined tools was supporting or enhancing creativity methods. The real life creative examples mentioned at the beginning of this section lead to a more precise examination of how the use of more than one interface can influence the creative ability of a person who uses a Creativity Support Tool to solve a problem or to find an innovative idea.

In the rest of this section we will present a real life scenario of a creative process that takes place in a classroom, and then we will describe how the same use case scenario can be applied with the use of a Creativity Support Tool.

2.1 Real Life Use Case Scenario

A group of 15 students attending a university course in applied mathematics are asked to solve a difficult mind-bending problem. The professor who teaches the course wants to examine the way of thinking of his students in problem solving processes. He initially describes the problem orally and asks the students to either collaborate in finding the solution or to work individually. He provides the following problem:

> *"You have a balance scale and 12 coins, 1 of which is counterfeit. The counterfeit weighs less or more than the other coins. Can you determine the counterfeit in 3 weightings, and tell if it is heavier or lighter"* [7].

Some of the students are taking notes while the professor describes the problem, and others listen carefully to the description. After the professor asks them to start working on the problem, one of the students asks him to write it on the white board, so that they can read it again and catch details that they probably missed. The professor writes the problem on the white board and asks them to work for 10 minutes on the collection of ideas related to the solution.

Six of the students start working individually and the rest of the students are working in groups of three. The professor observes the students´ working processes and their reactions. The students are working as follows:

- The students who work individually are writing their ideas for the solution methodology on a paper. For every idea they write, they review it crosschecked to the problem and either they continue writing a new one or delete the last. One of the students creates twelve paper balls and marks one of them as the fake one. He uses the paper balls as the coins referred to by the problem and he divides them into groups while he is taking notes. A third student is continuously reading the problem loudly. He makes some pauses to think and sometimes he writes some notes and diagrams. One other student is trying to solve the problem using mathematical equations and symbols.
- The three groups are working in a similar way. Their difference with the students who work individually is the fact that they collaborate with each other and they provide ideas after discussion. An interesting point is that each individual member of each group uses a different method of idea collection, similar to the students working individually. Each collects his ideas individually using either diagrams, or notes or paper balls, and when she has a good idea puts it in a common pool of ideas for discussion. In most of the cases when the group members explain ideas

for possible solutions to the other members of the group, they use different methods of demonstrating the ideas, such as design on the white-board, mathematical equations or oral speech.

When the time is over the professor asks the students to present their ideas and the problem solution, if they have one. The professor collects all ideas on the white-board and he begins an interactive conversation with the students for each methodological idea provided. Through the discussion they select the best ideas that seem closest to the solution, and they test the solution using figures and diagrams on the whiteboard.

2.2 The Use Case Scenario Using a Multi-interface Creativity Support Tool

The scenario described in sub-section 2.1 describes the case of a problem solving process that employs the use of conventional methods. In this paragraph we will attempt to adjust the same scenario into problem solving process, using a Creativity Support Tool and taking into account the different users´ learning styles. For better understanding we assume that the CST of this scenario has two different user groups with different user rights. These are the “moderators” and the “learners”.

Assumptions:

- The professor of the class belongs to the first user group which has more user privileges since he is the person who moderates the students/learners in their participation of the problem solving process.
- The students belong to the “learners” user group.
- The Creativity Support Tool we refer to is a web based Collaborative Creativity Support Tool (CCST) which can adapt the content according to the learning style that each user has.
- The CCST supports interfaces for 5 different learning styles; literal (**L**), musical (**M**), visual (**V**), interpersonal (**I**) and intrapersonal (**Intr**) [4].

The moderator uses the CCST to create a new creativity session for finding a solution for the 12 coins problem. He registers to the system and creates a new creative project by writing the problem description. After that, the moderator selects the participants/learners and he sends them an online invitation to participate in the creativity session for finding a solution for the problem. From the problem setup screen, he also defines options that reflect to the overall creativity process like “limited time=10 minutes”, “synchronous activities”, “Begin time and day”, etc.

The learners receive the invitations and they register onto the system at the defined date and time that the session is scheduled by the moderator to start. The first time the users register to the CCST, the system provides them with a learning styles questionnaire [8] which is mandatory to complete. The results of this questionnaire are saved to their profiles for the systems computations by the adaptation mechanism that the system supports. The adaptation mechanism is used for the selection of the content presentation method taking into account each learner’s learning style.

Taking into consideration the conventional case scenario from the previous sub section, the participants of the creativity session can be categorized according to their learning styles using the following pattern.

- Learners who repeat the problem loudly or ask the professor to repeat the problem description, belong to the musical learning style group.
- Learners who write the problem, take notes and write down their ideas, belong to the literal learning style group.
- Learners who design diagrams or create paper balls to experiment with, belong to the visual learning style group.
- Learners who work in groups and try to find a solution in collaboration with other learners, belong to interpersonal Learning style group.
- Learners who work individually belong to the intrapersonal learning style group.

The CCST classifies the users in the groups they belong to based on the above pattern and, with the use of a recommender system, it recommends to the users to accept the interface that corresponds to their learning style as a way to work. For the (**L**) group, the presentation of the content is text based. The problem description, the ideas surrounding the problem, tips for the user actions and all context information correlated to the problem is presented as text. The input is also text and it is supported by provision of word synonyms aiding the literal user to express an idea using the correct words. For the (**M**) group the interface contains a text to audio transformation system which offers to the learner the possibility of hearing what it is written in the screen. For the (**V**) group the problem description and the ideas provided for the problem are presented in a diagram. There is also a recommendation mechanism which provides images related to the problem and the user is able to select the images and add them in the diagram. The (**I**) group supports all the above interfaces for the users, but additionally offers the possibility for real time collaboration with other users. The collaboration method allows the synchronous collaboration between the learners as well as communication methods such as chat. The interface provided in each group member adapts the input content following the aforementioned adaptation pattern. For the (**Intr**) group the system does not allow the collaboration between the learners but it offers the learner the option to change into an (**I**) group.

When the session's time is over all ideas provided by the learners are collected into a diagram which is visible by all learners and the moderator. All learners vote on the best ideas and finally, the moderator collects the ideas with the highest scores as the problem's solution approach. With these ideas, each learner is able to understand the problem solution methodology.

3 Description of the Multi-interface CCST Prototype

In order to explore the ways in which adaptation of the working environment of a CST based on the learning styles of users who participate in a creativity session, we developed a multi interface Collaborative Creativity Support Tool prototype. The proposed prototype implementation that we present in this work consists of a simplified version of a CCST. For the aims of this work, we are using one creativity technique, brainstorming; and we implement two interfaces for two learning styles users, the literals and the visuals.

3.1 System Architecture

The prototype is a web based application, and it is designed following the three tier architecture principles. It is written in the PHP scripting language and it uses mysql as the backend database engine. It includes middle tier modules such as a synonyms suggestion mechanism, and a rule based context aware mechanism for identifying the user's actions and deciding the proper working environment and the content's adjustment to it. In particular, the system's architecture consists of three parts, the database, the application logic and the presentation. The most important part is the application logic which includes all the partial modules of the application. The modules of the application are independent sub-applications that are designed for specific system functioning. All modules are managed and triggered by the application logic part with the use of the context aware module. Each module as they are shown in Figure 1 is described in the continuation of this section.

- *Data transformation module* is the mechanism responsible for the transformation of data between the different interfaces. The existence of more than one interface presumes the existence of such a module. When data is input via an interface, this data must be available for use by other interfaces too. This is because users that belong to different learning styles groups and therefore work with different interfaces must have access to this data if they need it. Thus the transformation mechanism can recognize the type of presentation that must be used and adjust the contents in the correct form. The correct functioning of this module also depends on the database schema and the data storage structure used. The transformation mechanism contains several queries which they are adjusted in each particular case according to the interface they have to fill.
- *Interface Selector module,* in combination with the *context aware module*¸ identifies the current browsing location of an active user, and by taking into account the user's learning style, it loads the corresponding content to the interface styles and graphics to setup the interface.
- *Context aware module* tracks the user's actions and loads all the profile context information for every user. The context aware module consists of a rule based reasoning method on which the rules are determined according to the user's action, her activities and her learning style.
- *Learning styles classifier module* contains the learning styles questionnaires [8] and it is triggered the first time a user logs in to the system. The module contains also a scoring method and the necessary queries to classify the new user into a learning style group and update the user's profile.
- *Synonyms module* consists of a method of synonyms propositions, for the user, while she is typing text input to the interface. It uses an existing thesaurus [9] and with the use of ajax technology when a new word is entered into the system the synonyms suggestions appear in the screen. The user is able to select a synonym with a single click and continue the typing.

For the implementation of the prototype, some existing technologies have been used. As already mentioned at the beginning of this section, the prototype is written in PHP and mysql. The ajax prototype framework [10] is also used as well as the MobyThesaurus [9] taxonomy and the MindWeb [11] mind-map tool. The MindWeb

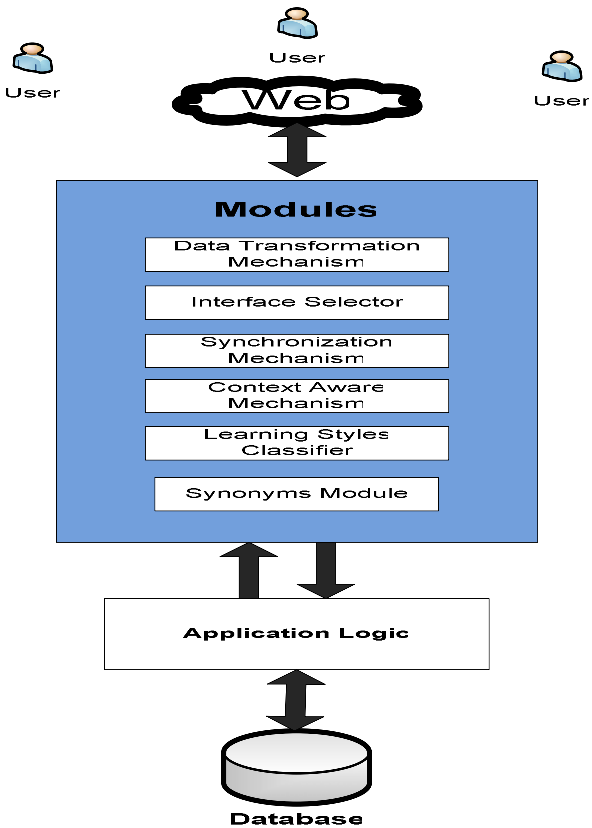


Fig. 1. System Architecture

mind map tool is modified for the purposes of this work in order to connect with the database and build its mind maps for each defined problem dynamically, on the fly. The connection with the database and the dynamic mind map build is not supported by other known mind map tools.

Due to lack of space, we provide a text based description of the implemented interfaces, rather than a set of screenshots.

- *Interface Approach for Literals***.** Users who belong to the literal learning style category are more comfortable in working with texts. They prefer to read text as well as write text. For this reason the proposed prototype supports a working environment based on text. The problem description, the ideas input from other users and the active user's ideas are presented on the screen in text form having a top to the end sequence of appearance. Literal users are also competent in selecting the proper words when describing a problem or an idea, thus the input forms of this interface are supported by a word proposition mechanism. This mechanism suggests to the user synonyms of the last entered word helping in writing well defined sentences and at the same time providing help in generating new ideas based on the suggested synonyms.

- *Interface Approach for Visuals***.** Users who belong to the visual learning style group are more comfortable in working with images. They prefer to design diagrams and create a visual effect of what they have in mind. Therefore the visual working environment is based on a mind-map diagram on which the root node is the problem description and the ideas surrounding the problem are nodes connected between each other with edges. The interface contains small icons that the user can add to the idea nodes and images that can be used as ideas. In the case of adding an image as idea, the user must also add a textual description of how the idea is related to the image regarding to the problem so this can be used by the literal working environment too.

4 Evaluation

For the pilot testing of the prototype, a group of 15 postgraduate students from the Computers Science Department of University of Cyprus with working experience in the IT sector, were asked to participate in an evaluation session and asked to register to the platform and work on a specific problem together. Six of them were female and nine were male. The purpose of this evaluation was to determine how the users perceive the use of an interface adapted to their learning style and proof that each user can work more comfortably with an interface that corresponds to the learning style they belong. For the aims of this evaluation we selected the 8 users from the overall group of 15 based on their results after they completed the learning styles test. The selected users were those who classified to the learning styles that the prototype supports, visuals and literals. Five of them were belonging to the visuals learning style group and 3 of them were literals.

Before the beginning of the evaluation all users received three common tasks to complete on the prototype, and a post task questionnaire to complete after each task. The aim of this pilot testing was the subjective assessment of the users so the 8 testers received a post test questionnaire to complete after they finish the three tasks. The three tasks given to the users were:

- *Task 1***.** Given the 12 coins problem if you think it is not well described please rewrite it in your own words using the Synonyms module.
- *Task 2***.** Provide three to five ideas for the defined problem**.**
- *Task 3.* Select the ideas that can be used as suitable solutions for the problem.

The first task aimed at the observing how helpful the synonyms module is, in describing a problem clearly and in examining how useful the module is for each learning style users.

The second task aimed at the examining how the users were able to provide new ideas by viewing ideas coming from other users in the two forms of presentation they were using. The visuals were viewing the ideas as a mind-map and the literals as a list of numbered sentences.

The third task aimed at examining how the users, literals or visuals, could manage the overall list of ideas using the two interfaces for identification of the most valuable ideas that could be used as solutions.

The post-task questionnaire included four 7-level Likert items and was taken from [12]: 1. *Overall, I am satisfied with the ease of completing this task;* 2. *Overall, I am* *satisfied with the amount of time it took to complete the task;* 3. *Overall, I am satisfied with the support information;* 4. *The interface I was working on was understandable*.

The mean value for answers in Task 1 was 5.22, for Task 2 it was 5.18 and for Task 3 it was 4.82. The third task had the lowest mean because of the difficulties faced by the literals on the selection of the accepted ideas as solutions of the problem.

The summary of the questions and the answers given by the completion of the post test questionnaire are depicted in the following table.

Table 1. Post test questionnaire summary of results

Question	Answers By Visuals	Answers By Literals
Do you think that the working interface assigned to you by the system was the suitable one for you?	All 5 users answered "yes". They believe that if they had the option to choose between the two interfaces they would choose the same interface.	One of the users answered "yes" and two answered "maybe". The users felt more comfortable working with textual representation of the content but they believe that the design needs several improvements. The structure of the presented ideas was not well organized and they would prefer to have more options to edit and change the presentation sequence for the ideas.
Did the interface facilitate you to understand the given problem?	The use of mind map helped in understanding the given problem. The use of the mind map helped in recognizing the important nodes with one look and focusing the interest at specific nodes which they contained large amount of information.	The interface helped in understanding the problem through the use of the problem refinement with the use of the synonyms module. Through this, the users managed to refine the problem using their own words.
What did you consider as weakness of the prototype?	The lack of an intelligent recommender system which would appear in the screen ideas in several forms such as images, texts or graphs. Others answered that they would like to use a drawing tool within the interface.	The ideas were presented in unstructured and the connections between new and previous ideas are not recognized. Some of the users also proposed the use of a recommender system for the recommendation of books and relevant resources to study.
What do you prefer to work with, images or text?	All users answered both, but with the flexibility to interact with images through a drawing tool.	All users answered both.

5 Conclusions

In this work we attempted to identify the learning styles as significant contextual factors for the enhancement of the creativity process. In real life creativity is an attribute that all people have. Its expression depends on the environment that people live work and act. Based on this, we examined how a different computer programming environment may influence the productivity of new ideas by a group of users. The selection of the working environment that corresponds to each user was taken based on the learning style group that each user belongs to. Through the preliminary testing results of the prototype that we presented in this work, it is obvious that the working environment based on the learning styles helps the users to feel more comfortable during a creativity session but it is also not enough. The subjective evaluation of the prototype leads to the conclusions that learning styles can be used as a decision factor for the selection of a creativity session's interface, but in some cases this can become more complex. This is a motivation for future work where more contextual factors of creativity must be found and used for the development of more successfully adapted to the users working interfaces.

References

1. Plucker, J.A., Beghetto, R.A.: Why creativity is domain general, why it looks domain specific, and why the distinction does not matter. In: Sternberg, R.J., Grigorenko, E.G., Singer, J.L. (eds.) Creativity: From Potential to Realization, pp. 153–167. American Psychological Association, Washington, DC (2004)
2. Kolb, D.A.: Experiential Learning: Experience as the Source of Learning and Development. Prentice-Hall, Englewood Cliffs (1984)
3. Honey, P., Mumford, A.: The Manual of Learning Styles, 3rd edn. Peter Honey, Maidenhead (1992)
4. Gardner, H.: Multiple Intelligences: The Theory in Practice. Basic Books, New York (1993)
5. Felder, R.M., Silverman, L.K.: Learning and Teaching Styles in Engineering Education. Engineering Education 78(7), 674–681 (1988)
6. Sielis, G.A., Tzanavari, A., Papadopoulos, G.A.: Enhancing the Creativity Process by Adding Context Awareness in Creativity Support Tools. In: Stephanidis, C. (ed.) UAHCI 2009. LNCS, vol. 5616, pp. 424–433. Springer, Heidelberg (2009)
7. Anonymous: Twelve Coins Problem, `http://www.iwriteiam.nl/Ha12coins.html` (retrieved on December 2010)
8. Anonymous, `http://www.businessballs.com/howardgardnermultipleintelligences.htm` (retrieved on December 2010)
9. Thesaurous, M.: `http://www.thesaurus.babylon.com/moby` (retrieved on December 2010)
10. Prototype, `http://www.prototypejs.org/` (retrieved on December 2010)
11. MindWeb, `http://sourceforge.net/projects/mindweb/` (retrieved on December 2010)
12. Lewis, J.R.: IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use. International Journal of Human-Computer Interaction 7(1), 57–78 (1995)

Shadow Expert Technique (SET) for Interaction Analysis in Educational Systems

Christian Stickel[1], Martin Ebner[1], and Andreas Holzinger [2]

[1] Social Learning / Computing and Information Services, Graz University of Technology, Steyrergasse 30/I, A-8010 Graz, Austria
tickel@tugraz.at, martin.ebner@tugraz.at
[2] Institute of Medical Informatics, Statistics and Documentation Medical University Graz, Auenbruggerplatz 2/5, A-8036 Graz, Austria
andreas.holzinger@meduni-graz.at

Abstract. This paper describes a novel usability method called Shadow Expert Technique (SET), which was applied on the learning management system of the Graz University of Technology in two different trials, with focus on consistency and visual complexity. This is the summary of the development of this new method and the approach to generalize it as a new way to get deeper insight into interaction processes.

Keywords: Shadow Expert Technique, Usability Test, LMS, Methods.

1 Introduction

This paper describes the work done on developing a technique for analyzing tiny bits of user interactions, in order to improve the interface as well as the interaction processes within our self-developed university wide Learning Management System (LMS) at Graz University of Technology. We will show how this technique was applied at two full scale usability tests and then discuss the lessons learned, thereby generalizing our so called Shadow Expert Technique (SET) [5]. Studies have shown that the consistency of user interfaces in LMSs plays a significant role on the learning performance [16]. The application of SET on our LMS is among these lines, as we try to reduce learner's cognitive load, by examining potential influential variables. So far the presented work was focused on improving consistency and visual complexity. At this point it might be important to mention that the development of SET as well as the experiments took place as part of a lecture course in advanced chapters of human computer interaction at Graz University of Technology during two consecutive years. The Shadow Expert Technique (SET) consists of several steps with two groups who strive to decompose the interaction processes of an end user, in order to derive suggestions for improvement. The primary advantage of SET is thereby the in-depth understanding of the end users behavior and actions by mirroring his emotional state, anticipating his thoughts as well as his next actions and discussing the observations and estimations within the group.

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 642–651, 2011.

2 Theoretical Background and Related Work

This chapter will roughly outline some topics which are related to the SET and provide some insight for a better understanding of the technique. SET is based on three foundations. The first is that emotions are elicited during the interaction of the user with the system. Emotions are a fast and fluid process, but if they can be detected by the analyst, they provide good markers for usability issues. The second foundation is the theory of mirror neurons, which are said to provide a human being with the ability to simulate the inner state of others, thereby providing also a basic level of empathy. This makes it even for 'untrained' persons possible to detect basic emotions like frustration or happiness. The third foundation is the dynamics of decision making and solution generation in moderated groups. Beside these three topics, this chapter provides a short description of the UE methods, Performance Testing, Thinking Aloud and Focused Heuristic Evaluation, as they were used to provide the necessary input material for the SET. A general description of SET can be found in chapter 3.

2.1 Emotions and Usability

Emotions are fast and an ever changing floating process during an interaction. They are created every time when an important change in the environment is perceived. Emotions provide us with the readiness to act in a certain way [4]. They are a psychological state that helps managing the achievement of goals, such that the relevance of events towards these goals is evaluated and eventually rewarded. Positive emotions occur when the goal is advanced and negative emotions appear when the goal is impeded. In usability engineering especially negative emotions provide valuable markers for issues. However they are far more important as they influence and interrupt ongoing interactions and even tend to change the course of action towards positive emotions. No user likes to achieve a goal with software that repeatedly plays with his frustration tolerance.

There are several approaches of labeling and categorizing basic sets of emotions [15]. Following a base like facial expressions and the well renowned coding system FACS [3] would provide six distinct categories, which are happiness, sadness, anger, fear, disgust and surprise. For the studies described within this work especially the negative emotions are of interest. In particular the detection of frustration is an interesting topic. Frustration is related to anger and disappointment, which is a negative form of surprise. It can be considered as problem-response behavior and will eventually lead to further negative behavior like narrow-minded problem solving approaches. The level of frustration can task wise be measured by using a combined recording the subject's arousal and valence [19] or using visual self-reports like SAM [10] or PrEmo [2]. However in the discussed studies we used a straight forward approach for detection by letting the observers 'mirror' the subjects, thereby becoming the users 'shadows'.

2.2 Mirror Neurons and Empathy

Mirror Neurons are a certain class of neurons for understanding the actions, intentions and emotions of other people, as well as the social meaning of their behavior. They were found by Rizzolatti [18] in 1995 and are since then subject of much controversial discussions. Rizzolatti states that mirror neurons allow human beings to grasp the mind of others by direct simulation instead of conceptual reasoning [17]. As this mechanism helps also to understand the emotions of others, it can also be seen as neuronal basis for empathy. A simple example might be that when a person observes another person pulling back an arm, as if to throw a ball, there would be a copy of this behavior in the observer's brain that helps understanding the goal. Additionally the facial expression gives clues about the inner state of other person, e.g. to solve the question if the ball is thrown to hurt the observer or just to play. In this way the mirror neurons help to anticipate the next steps by simulating the actions and emotions of those we observe.

For SET the step of anticipation is interesting in particular, as we want to find out what happens to a person, who clicks a button for the tenth time, or just gets cryptic error messages. Finally this leads to the conclusion that the (screen record) playback observation of a subject's actions on an interface together with a video of the face is sufficient to give insight into the interaction process as well as identify important emotional states like frustration or happiness, without the explicit knowledge of an emotion coding system like FACS or EmFACS [3]. The efficiency of this simulation can be even increased by several replays of the situation and by asking the observers to physically mimic the subject.

2.3 The Power of Groups

Group decision making refers to a process in which multiple individuals act together in order to analyze problems and generate solutions. The efficiency of groups is influenced by their size, demographic makeup, structure, contextual settings, goals and actions. In such a process the group utilizes the diverse strengths and expertise of its members, thereby generating a greater number of alternatives than an individual. A simple example for this is a brainstorming session. In order to work effective there's the need of a leader for a group. In case of the SET this role is taken by the investigator respectively moderator, who determines tasks and structures the discussion. The moderator clearly frames the problem and encourages the group to develop several creative options, without indicating preferences for any particular solution [11]. In our experience the structured discussion within the group generates a collective understanding of issues on the one hand and diverse solutions on the other hand. The exact role of the moderator will be explained in chapter 3.

2.4 Focused Heuristic Evaluation (FHE)

The Heuristic Evaluation is an informal usability inspection method that allows identifying potential problems in the design of user interfaces by following recognized usability principles [13, 9]. In this method, a list of pre-defined heuristics is provided to a small team of evaluators. The number of evaluators is still debated; however, Nielsen [14] recommends using three to five evaluators. Each evaluator works independently and tries to find out potential problems and positive aspects of the interfaces according to the checklist of provided heuristics. The purpose of this independence is to ensure an objective and unbiased evaluation. In simple Heuristic

Evaluations, these heuristics usually focus on discovering various general and diverse issues regarding the design and behavior of the system. However, in the case of a Focused Heuristic Evaluation, only a single issue is chosen, and a list of appropriate heuristics is generated accordingly for evaluation [5, 8].

2.5 Performance Test (PT)

Performance testing helps to identify issues that influence the performance of an end user while using a working system [16]. The performance of users are usually evaluated in three dimensions i.e. time, effectiveness and user efficiency. In order to gather reliable and precise information about the system for a performance test, the test must take place under realistic conditions with real end users. As suggested by Virzi [20], usually from 12 to 20 test persons are sufficient to gather reliable data. Each test is usually video and audio taped to help developers to compare and evaluate different system designs in an iterative development cycle. The core indicators of a user's performance such as Task Effectiveness, User Efficiency, Relative User Efficiency and User Satisfaction can be obtained as defined in ISO 9241-11 [1, 7].

2.6 Thinking Aloud (TA)

The "Thinking Aloud Method" was introduced by Lewis in 1982 to understand cognitive processes [12]. It encourages test users to speak verbally about whatever they are looking, feeling, and thinking, while performing any task. Test sessions are usually recorded on video such that the actions of participants can be tracked back to see what participants did, and how they reacted. This method helps observers to understand the cognitive processes involved in completing any task. By engaging actual or intending users as participants, the thinking aloud method provides a closer view of how users use the system and reveals several usability problems in performing any particular task [6].

3 Methods and Materials

This chapter provides a general description of the Shadow Expert Technique (SET) with detailed information on the three phases. It will further discuss the experimental design and argue the changes for optimizing the technique.

3.1 General Description of the SET

The Shadow Expert Technique (SET) is an experimental multilevel usability method that utilizes the input of well renowned methods like Heuristic Evaluation (HE), Thinking aloud (TA) or Performance tests (PT). The core is a structured discussion which aims at getting insight into the inner state of a subject during an interaction, thus revealing already slight negative emotions, like beginning frustration. This in turn provides a deeper insight into occurring issues and generates interesting solutions. The approach can be seen as stepwise thinking into the user and will be described in detail within this chapter. The Shadow Expert Technique consists of three distinct phases [5] as can be seen in fig1 below. Every phase has specific goals and actions.

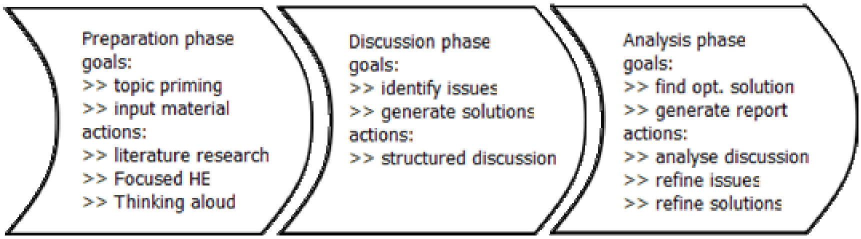


Fig. 1. Goals and actions in the three phases of SET

The first part is the preparation phase. In this phase the evaluators are primed on a specific topic (e.g. consistency) and the interface, and then split up by the investigator into two groups, which generate the input material for the discussion. Two groups are necessary here in order to later benefit from the effect of 'fresh eyes'. In the second phase, synchronized screen recordings and video material from the first phase are exchanged between the groups and reviewed in the actual discussion, in order to identify the issues and generate solutions. The discussion is recorded for the final analysis phase, which is meant to describe all issues in detail, elaborate optimum solutions and generate a report. Therefor the record of the discussion is analyzed and all issues and solutions are refined. The final report contains a list of potential problems and alternative/optimum solutions to improve the system.

3.2 Preparation Phase

The preparation phase aims at priming the evaluators on a specific topic on the one hand and on getting them to know the interface on the other hand. A literature research can provide the background on the topic and a list of specific heuristics. However it's not necessary to use a specific topic like consistency or visual complexity, as the method should also work in a more general way. In case that there is no specific topic, the Focused HE is unnecessary and we'd suggest to rather use a general HE according to Nielsen [6]. The investigator will the split the evaluators in two groups which derive tasks for a TA based on the most severe problems discovered during the FHE or HE. Thereby the tasks for are created by each group independently with the obligation not to let know the other group. Then the investigator should select the tasks for every group, such that the groups use different tasks for the TA. The TA should be run with at least seven subjects.

The outcome of the test should be a synchronized video of a screen recording, showing the interactions of the subject and a video showing the subjects facial expressions. The group will then need to evaluate the whole trial to determine a video of an average user for every task. The selection of an average user should provide more realistic data than choosing an extreme good or bad user. Further considering the rather intensive discussion phase the decision of discussing just one video per task helped to reduce the workload. However if there's enough time and resources all videos should be analyzed.

Equations (1), (2) and (3) were used to select the videos such that for each task, the user whose efficiency in a task (1) is closest to the task efficiency intersection (3) of this task was selected for the next phase. As there were five tasks, this selection resulted in five videos as input for the discussion. The advantages and disadvantages of this selection approach are discussed in chapter 5.

$$user\,efficiency\,task_x = \frac{effectiveness}{t_{task}/t_{max}} \tag{1}$$

$$total\,task\,efficiency_x = \sum_{i=1}^{\sum users} user\,efficiency\,task_{x_i} \tag{2}$$

$$task\,efficiency\,intersection_x = \frac{total\,task\,efficiency_x}{number\,of\,tasks} \tag{3}$$

3.3 Discussion Phase

In the beginning of the discussion phase both groups exchange their material, such that group 1 will discuss the output of group 2 and vice versa. Here we use the idea of ‘fresh eyes’ which can discover things that have been unrecognized due to intensive work on the material. During the discussion the group that provided the material will act as scribes, thus taking notes and eventually asking questions for clarification. The investigator will take the role of the moderator and guide the discussion. The whole session should be recorded with a proper microphone and a camcorder. It will be processed together with the notes in the last phase.

In order to proper review the tasks one or two projectors and computer for playback will be needed. An optimal setup would project the screen record synchronized right beside the subjects face. Showing the videos picture-in-picture is not optimal as the face is small and eventually overlaps areas which are important during the interaction. Composing and ordering both videos with an editor program beforehand is a proper solution to this problem. The whole discussion phase consists of two consecutive steps, which can be seen in fig2.

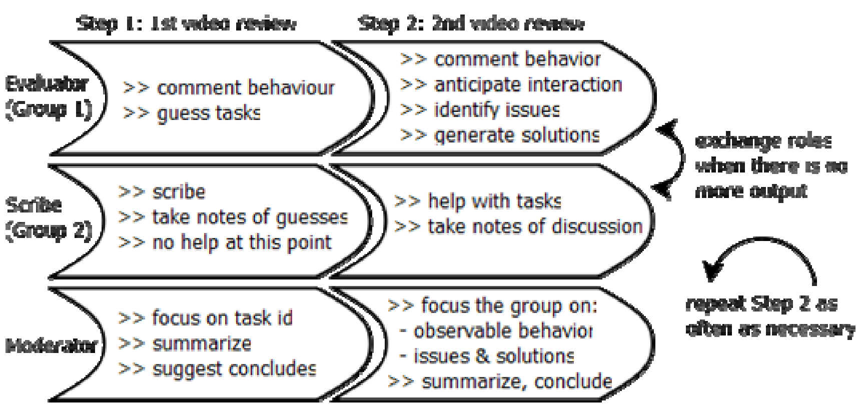


Fig. 2. Roles and actions in the discussion phase of SET

In the first step, the evaluators review the video and try to guess the tasks the user was performing. In this step, the sound of the video is muted, which avoids the evaluators to know the user's tasks in advance. Meanwhile the scribes take notes of the guesses made by the evaluators. Step 1 is finished when all tasks have been reviewed and reasonable guesses of task descriptions have been made.

In the second step the video is reviewed again, but this time with audio. Thereby the evaluators focus on understanding the user's cognitive processes involved in performing the particular tasks (by commenting on behavior and anticipation of interaction) and identifying issues as well as generating proper solutions. The audio from the thinking aloud test enables the teams to better understand the subject's expectations, intentions and emotions. Therefor the evaluators might first use problem oriented statements like The user does … because he thinks / perceives / feels… In order to generate solutions these statements can be reversed, e.g. In order to prevent that the user … thinks / perceives / feels etc. … or In order to support the user … thinking / perceiving / feeling etc. a possible solution would be… and so on. The moderator guides the discussion toward concrete solutions for identified issues. Meanwhile the scribes take notes of the discussion and may ask questions for clarification on certain statements. Step 2 can be repeated if necessary, however when there is no more output the groups should swap their roles.

3.4 Analysis Phase

In the final step of the analysis phase each group uses their notes and the record of the discussion phase to refine the results. Ideally the record is transcribed and then consolidated and clarified. Issues without solutions or multiple solutions can be investigated further, in order to generate an optimum solution. However the final report should contain the alternatives as well as the optimum solution, beside the found issues. The report should aim at suggesting the developers where and how to improve the system. Graphical mockups and flow graph of optimized interaction processes are appropriate means to communicate this. Ideally developers are part of the evaluation team or take part as silent observers in the discussion phase.

3.5 Experimental Design and Development

As already mentioned there were two trials in two consecutive years. They were used to analyze certain usability aspects. The first trial was about consistency of the interface while the second trial was concerned with visual complexity. In both cases the evaluators were primed by a three week research on the topics which led to the generation of specific heuristics for the FHE.

According to experiences in the first trial, some changes were done to optimize the technique in the second trial. In the first trial a performance test (PT) was used to generate the input material. This kind of material made observing harder, as a PT usually tries to simulate a real life setting with focus on efficiency, thus it's more difficult to find out about the users intentions and inner states. In the second trial the PT was replaced by a thinking aloud test (TA). Although replayed without audio the resulting material allowed the observers much better insight, as the subjects acted slower and expressed their emotions more clearly. Another intention of using TA was to analyze the performance of the observers. This can be done by comparing the statements of the subject with the comments of the observers, who are unaware of what the subject is talking since the video playback is muted. This way the observers are forced to more active behavior.

Another change in the material was the utilization of an average user for every task instead of determining an average user from the whole trial. This change provided the advantage to observe much more realistic issues. A disadvantage however is that the observers have to cope the constant change of users, which might somehow be disruptive to the process of getting more familiar with the users expressions.

4 Discussion and Further Works

This paper presents ongoing research on a novel usability method called Shadow Expert Technique (SET). It was developed, tested and evolved as part of an HCI master lecture held by Prof. Holzinger at Graz University of Technology. SET was applied at the university wide LMS with the focus on improving certain usability aspects of the system. There were two trials. Between the two trial slight changes were made in order to improve SET with respect to an increase in quality and quantity of the outcome. The changes as described in the previous chapter 3.3 proved to be reasonable and applicable.

In the last trial we discovered a problem that can occur during the discussion phase. At a point where the groups have changed the roles there might be an intersection of issues. Although both groups have different tasks to analyze, there might be a common base for several issues. If this base has already been identified before switching the roles it might be hard for the 'new' evaluators to come up with genuine own solutions. Our students later proposed to do the second step of the discussion each in absence of the 'scribes'. As the whole discussion is recorded anyway this should not affect the overall performance of the SET. This will be tested in the next trial.

Another slight change in the next trial might be to instruct the moderator to focus the evaluators also on more actively mimicking the subject's behavior and expressions in order to create a more intensive immersion. It might also be interesting to use psychophysiological measures and visual self-assessment in both, the TA (for subjects) and the discussion phase (for the evaluators) in order to compare the subjects emotions with the artificial elicited ones of evaluators. Questions that might be continuative for further evolving of SET concern minimum and maximum numbers of group size, more effective priming and techniques for deeper immersion.

At the current point we conclude that the Shadow Expert Technique in the current described state is an effective (not yet efficient) method for detecting big flaws and tiny issues in interaction processes. It can be used focusing on a specific topic (e.g. consistency, visual complexity) or in a general way, whereby latter is not proved yet. SET provides a deeper understanding of the end user's behavior and usually generates in the process diverse alternative applicable solutions.

Acknowledgements. Headings We cordially thank the following students of the lecture 706.046 “AK Human-Computer Interaction: Applying User Centered Design” for their enthusiasm in carrying out the evaluations: Markus Fassold, Claus Bürbaumer, Marco Garcia, Daniela Mellacher, Thomas Gebhard, Christian Partl, Patrick Plaschzug, Dieter Ladenhauf, Muhammad Salman Khan, Marco Lautischer, Georg Michael Lexer, Pulkit Chouhan and Alois Bauer.

References

1. Bevan, N.: Measuring Usability as Quality of Use. Software Quality Journal 4(2), 115–130 (1995)
2. Desmet, P.: Measuring emotion: development and application of an instrument to measure emotional responses to products. In: Blythe, M.A., Overbeeke, K., Monk, A.F., Wright, P.C. (eds.) Funology, pp. 111–123. Kluwer Academic Publishers, Norwell (2005)
3. Ekman, P., Friesen, W.V.: Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press, Palo Alto (1978)
4. Frijda, N.H.: The emotions. Cambridge University Press, Cambridge (1986)
5. Holzinger, A., Stickel, C., Fassold, M., Ebner, M.: Seeing the System through the End Users Eyes: Shadow Expert Technique for Evaluating the Consistency of a Learning Management System. In: Holzinger, A., Miesenberger, K. (eds.) USAB 2009. LNCS, vol. 5889, pp. 178–192. Springer, Heidelberg (2009)
6. Holzinger, A.: Usability engineering methods for software developers. Comm. ACM 48, 71–74 (2005)
7. Holzinger, A., Searle, G., Kleinberger, T., Seffah, A., Javahery, H.: Investigating Usability Metrics for the Design and Development of Applications for the Elderly. In: Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I. (eds.) ICCHP 2008. LNCS, vol. 5105, pp. 98–105. Springer, Heidelberg (2008)
8. Javahery, H., Seffah, A.: Refining the usability engineering toolbox: lessons learned from a user study on a visualization tool. In: Holzinger, A. (ed.) USAB 2007. LNCS, vol. 4799, pp. 185–198. Springer, Heidelberg (2007)
9. Kamper, R.J.: Extending the usability of heuristics for design and evaluation: Lead, follow get out of the way. International Journal of Human-Computer Interaction 4(3-4), 447–462 (2002)
10. Lang, P.J.: Behavioral treatment and biobehavioral assessment: Computer applications. In: Sidowski, J., Johnson, J., Williams, T. (eds.) Technology in Mental Health Care Delivery Systems, pp. 119–137. Ablex, Norwood (1980)
11. Leonard, D., Swap, L.: When Sparks Fly: Igniting Creativity in Groups. Harvard Business School Press, Boston (1999)
12. Lewis, C.H.: Using the "Thinking Aloud" Method. In: Cognitive Interface Design. Tech. rep., IBM RC-9265 (1982)
13. Nielsen, J., Molich, R.: Heuristic evaluation of user interfaces. In: CHI 1990, pp. 249–256. ACM, New York (1990)
14. Nielsen, J.: Finding usability problems through heuristic evaluation. In: CHI 1992, pp. 373–380 (1992)
15. Ortony, A., Turner, T.J.: What's basic about basic emotions? Psychological Review 97, 315–331 (1990)
16. Rhee, C., et al.: Web interface consistency in e-learning. Online Information Review 30(1), 53–69 (2006)
17. Rizzolatti, G., Fogassi, L., Gallese, V.: Mirrors in the Mind. Scientific American 295(5), 30–37 (2006)

18. Rizzolatti, F.G.L., Gallese, V., Fogassi, L.: Premotor cortex and the recognition of motor actions. Cognitive Brain Research 3, 131–141 (1996)
19. Stickel, C., Ebner, M., Steinbach-Nordmann, S., Searle, G., Holzinger, A.: Emotion Detection: Application of the Valence Arousal Space for Rapid Biological Usability Testing to enhance Universal Access. In: Stephanidis, C. (ed.) UAHCI 2009. LNCS, vol. 5614, pp. 615–624. Springer, Heidelberg (2009)
20. Virzi, R.A.: Refining the test phase of usability evaluation: how many subjects is enough?'. Human Factors 34(4), 457–468 (1992)

On Interactive Interfaces for Learning Chinese Character Writing Online

Manlai You, Yu-Jie Xu, and Cherng Chin

National Yunlin University of Science and Technology, Graduate School of Design
Douliu, Yunlin 64002, Taiwan
{youm,g9930807,cchin}@ yuntech.edu.tw

Abstract. Learning to write Chinese character is a crucial process for non-native speakers in Chinese learning and a big stumbling block also. Luckily there are numbers of websites providing online learning of Chinese character writing which may benefit learners in overcoming the barriers. In this study we focus on these websites and emphasizes on the instruction of Chinese character writing. There were 24 of the websites screened from the Internet for further investigation. After intensive analysis and comparison, eight main factors in interface design of Chinese character writing were identified. They are: Media of presentation, Grid pattern, Tracing outline, Color coding, Stroke labeling, Play control, Stroke speed, and Writing exercise. Based on the analysis of the advantage and disadvantage of different implementations of these 8 main factors, an improved interface design proposal for Chinese character writing instruction is then propose to the follow-up research.

Keywords: Chinese characters, Chinese learning, Chinese character writing, Interactive interface, Interface design.

1 Introduction

Currently, Chinese learning has become very popular in the world. However, not every learner can have formal guidance from qualified instructors. Thanks to the rise of network learning technology, this deficiency can now partly resolved by well designed online interactive interfaces. Websites for Chinese learning are not only growing in number but also increasing in quality on a daily basis. Chinese characters are basically an ideographic writing system with its own specific characteristics [1]. Many researchers have explored different aspects of learning Chinese character writing, such as the stroke order [2][3][4], the radical [5], the component structure [6][7]. Still others emphasize on other relative information, such as the font factor of Chinese character or even the digital auxiliary learning platform. Stroke by stroke, repeat and repeat to practice is the traditional method to learn how to write the thousand-some most frequently used Chinese characters, not only for students who treat Chinese as a foreign language, but also for the Chinese native speakers too. Through the practice of hand-writing, one can increase his/her apprehension of the construct structure of the Chinese character, leading to a full understanding of how to handle Chinese character writing [6][7]. Luo, after investigating with western students

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 652–661, 2011.

learning Chinese character writing, points out that if the learners are adult and who can fully handle the principle of Chinese character writing then the tutorial learning could be successful with appropriate auxiliary tools [8]. This again verifies that character writing plays an essential role in the learning of the Chinese langue.

This study investigates the websites that provide Chinese character writing lessons and then proposes an improved interface design for learning character writing. Firstly, based on qualitative observation and analysis, the interactive interfaces adopted in a sample of websites that teach character writing are discussed with emphasis on the instruction of stroke order. Secondly, after comparison and evaluation, crucial factors of learning stroke order of the Chinese characters and the best practices in the existing interfaces are explored and identified. Finally, an improved version of Chinese character stroke order teaching interface is proposed, which takes advantages of existing best practices along with contemporary online teaching techniques. The new version supposedly would be better accepted by beginners of the Chinese language in learning character writing. The rest of this paper further describes the content of these three main stages of this study in more detail.

2 Selection of Websites and Essential Factors for Character Writing Instruction

Searching with the key words "Chinese character" in the social bookmark sharing site Delicious, more than 2,000 websites about Chinese characters learning were assembled. But with the limitation of the budget, only the free websites or trial versions were selected. All candidates should provide complete functions or unique interactive features for Chinese character writing interface design. Totally 24 websites for stroke order learning are selected based on the rule for the further investigation (Table 1).

There are 24 websites in public domain being analyzed and synthesized through the discussion of four students in PhD program of a design school. This group watched out the basic functions and presentation styles of each website, and compared them to table out both advantages and disadvantages of the interactive interface design with the identity or similarity of detail functions. Based on the importance of learning effects, 8 main factors were then categorized. Those were sorted out by the user focus order when they started the study. Each of the factors is then briefly defined as follows:

1. Media of presentation: The media used in writing demonstration can either be in the format of animation, image, video, or any of their combinations.
2. Grid pattern: The auxiliary pattern underlying the writing field for references in locating the beginning, movement, and ending of a stroke and its relationship with other strokes and character components. The most frequently used patterns are Blank Square, Cross, Diagonal, and Quadruple Cross.
3. Tracing outline: The tracing outline is an outline of the character to be written and mapped on the Grid in the writing field. The outline can either be a solid paled shadow or with hollowed strokes. These outlines are with colors in pink or red generally.

Table 1. The 24 selected websites with instruction of Chinese character writing

1	bestvoc.com	11	Learning Program for Stroke Order of Frequently Used Chinese Characters	16	Orchild Pavilion: Chinese as a second Language
2	Chinese Learner	12	漢字讀音筆順	17	Arch Chinese[1]
3	Chinese Tools.en	13	Chinesisch-Deutsche Gesellschaft e.V. Hamburg	18	Chineseetime[1]
4	Chinese-Tools.com	14	Mandarin Teaching Course	19	Hello-Han1
5	Learn Chinese Online	15	United Nations Chinese Programme	20	Skritter1
6	LearnChineseABC.com			21	Yellowbridge[1]
7	Mobilese			22	Chinese Characters Software[1,2]
8	Simple-Chinese.com			23	eStroke[1,2]
9	UCS College			24	Hanzi Master[1,2]
10	Centre for Teaching Chinese as a Foreign Language University of Oxford				

[1] The websit allows trial use. [2] With software available for download for offline use.

4. Color coding: Colors are used for coding different strokes or components of the Chinese character during the writing demonstration. Generally, three types of strokes/components are color coded: the stroke currently being written, the stroke/s which have been written, and the radical of the character.
5. Stroke labeling: Whether or not a stroke being labeled with its order number and direction.
6. Play control: Availability of user control over the animation, such as play, stepping play, pause, repeat, and stop. The control can either be on the whole or on the components/elements of the animation/video.
7. Stroke speed: Drawing speed of stroke and the pause between stokes in the demonstration in video or animation.
8. Writing exercise: Availability of Chinese character writing exercises, including online practice and download lessons for offline practice.

Each of the factors is constituted of several construct elements with different possible implementations. These are to be discussed in detail, one by one, with examples and illustrations in the following sections.

Fig. 1. Stroke order presented with images (left) and video (middle). Example of animation verisimilitude problem (right).

2.1 Media of Presentation

The Chinese character writing demonstration includes Interactive animation, image (Fig.1 left), and video (Fig.1 middle). Interactive animation is the main stream for the presentation in the Chinese character writing demonstration, because the Flash can provide abundant control ability for learners. However, attention should be paid to the fluency and stroke processing, and to avoid the stroke shape in the circle does not conform to the real situation in calligraphy (Fig.1 right). The image media is a series of still pictures which do not take the advantage of digital media for instruction, but just transferring the earlier paper textbook contents to website only. Video is a media used to record handwriting by a real person, both camera angle and resolution can affect the quality of the demonstration. According to the interactive ability, video is better than image. However, these two are not as superior as Interactive animation. From the point of view of the importance of media interaction for learners, this study considers website design should choose Interactive animation, and treats the video as an auxiliary, and abandons the image media, except for the presentation of font styles, etymology, or something important to be saved.

2.2 Grid Pattern

The auxiliary pattern underlying the writing field is to assist users to locate the movement of the stroke from the beginning to the end, and its relationship with other strokes and character components. The surveyed websites use four kinds of Grid, including Blank Square, Cross, Diagonal and Quadruple Cross (Fig. 2). The main difference of these websites is the contorlbility of these grids, No.17 and No.23 provide all four kinds of Grid, among them the learner can choose his/her own preference. The other websites just provide one kind of Grid without choice, except No.2 and No.13 do not provide any Grid for writing demonstration at all. As an assessment result, those websites that provide Grid is better than none, more choice is also better than no choice too.

2.3 Tracing Outline

The tracing outline is an outline of the character to be written and mapped on the Grid in the writing field. The outline can either be a solid paled shadow or with hollowed strokes. These outlines are with colors in pink or red generally (Fig.3 a,b). An unabridged outline provides more information and guidance than an empty Grid, i.e., the outline increases apprehension for the learner. However, it seems that all kinds of outline and color work equally fine as long as a tracing outline can match observation requirement for writing demonstration. But for the advanced learners, the tracing outline can be omitted when their familiarity with Chinese characters have been upgraded to a higher level. This study considers that a website that provides the tracing outline is better than not, and that provides an option to switch among different tracing outlines is better than not.

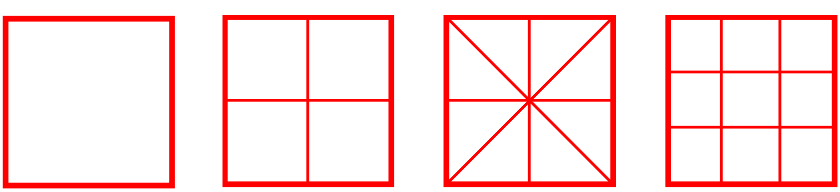


Fig. 2. Grid patterns include Blank Square, Cross, Diagonal and Quadruple Cross

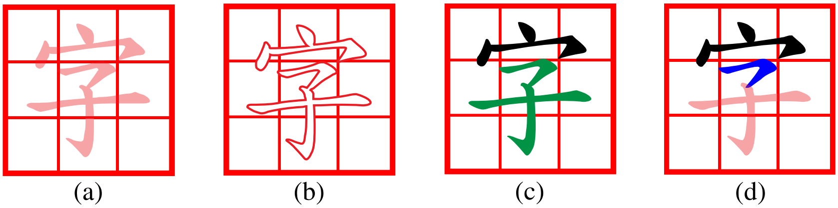


Fig. 3. Examples of solid paled (a) and hollow (b) tracing outline, color coding: black is written, green is the radical of the character (c), and blue is the ongoing written stroke (d)

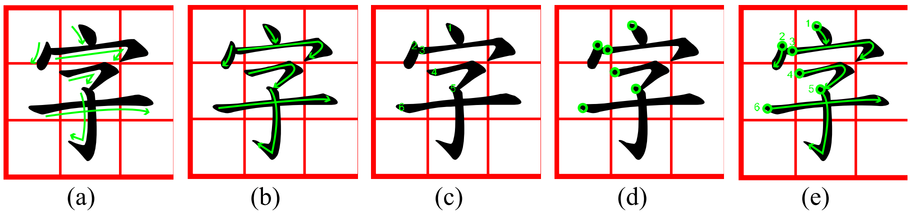


Fig. 4. Examples of external (a) and internal (b) labeling of writing directions, order numbers (c), start points (d) and all strokes labeled (e)

2.4 Color Coding

Colors are used for coding different strokes or components of the Chinese character during the writing demonstration. Generally, three types of strokes/components are color coded: the stroke currently being written, the stroke/s which have been written, and the radical of the character (Fig.3 c,d). The radical is the most important taxonomy of the Chinese character construction and indexing. The usage of color can point out the radical to enhance the learner's impression. Currently, websites designer use bright colors for the ongoing writing stroke to mark it out and prompt the stroke order, when the strokes are going continuously the finish stroke will be replaced by default color and the next stroke will be brightened. The favorite color of each learner is quite different, both No.17 and No.22 websites provides palette function for learners to select preferred color by themselves. This study suggests that a website that provides color coding is better than not, provides color coding for both radical and ongoing stroke are better than for only one of them, provides a palette function for choosing preferred colors is better than not, and provides a switch option is also better than not.

2.5 Stroke Labeling

The stroke labeling labels the strokes with order numbers and direction when the Chinese character are written, including both starting and ending points, and direction vectors with an arrow sign (Fig.4 b-e). The Stoke labeling is an auxiliary function of stroke order in Chinese character writing. After a learner has crossed the basic level, some or all the stroke labeling functions can be switched off. Currently, two kinds of labeling styles exist. They are internal and external labels (Fig.4 a,b). The internal label must be fitting within the stroke, hence is limited with the stroke width, leading to low visibility. Even more, the stroke color may overwrite the internal labels and totally erases the given information. The external label, though resolves this problem, can also mess up the clearness of the character strokes when put labels beside/along the strokes. This study suggests that stroke labeling function should be provided with switch options for each operation, and the location of the stroke labels should be external. Even better if the stroke label appears just before the stroke animation starts, and disappears after stroke finishes, it can avoid the label shrinking and mess up conditions to happen.

2.6 Play Control

The Play control lets the learner handle the Interactive animation demonstration control. The controllability effect the interaction between the learner and the Interactive animation. Examples of play control are Continuous writing play and Single stroke writing play. The Continuous stroke writing play is a complete processing of the whole Chinese character, writing from the first stroke to the last one, including Play, Pause, Stop, and Repeat. The Single stroke writing play is a stepping processing of the Chinese character writing, one stoke followed by another, including Play and Repeat. The Continuous stroke writing play is currently the dominant function in the analyzed websites, except for No.12 and No.18, which provide a Single stroke writing play function. As to the control button design, both text and symbolic icons are used. The media player icons have been unified and widely used and accepted for a long time, hence it is not necessary to create a brand new set of icons for the Play control function, lest the learners to be confused. This study suggests that the Interactive animation designer should provide both Continuous and Single writing play functions, and the icon designs should follow the general standard to decrease possibility of misusage by the Chinese character writing learners.

2.7 The Stroke Speed

The Stroke speed refers to the control of the Interactive animation writing demonstration speed by the learner, including stroke speed control, interval time control between strokes, delay time control between characters. Each learner has his/her own feel about speed, also the perception of Stroke speed differs from beginners to advanced learners. Flexible speed control is required to satisfy different levels of learners. There are three types of speed control: stroke speed control which controls the average speed of each single stroke, interval time control which controls

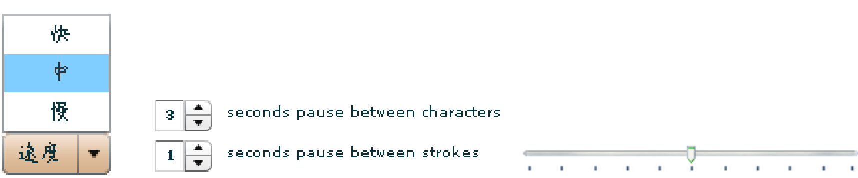


Fig. 5. Examples of Stroke speed control method

the delay time between the single strokes, and the delay time control which controls the delay time between the characters' writing play.

Currently, only No.17 provides a complete stroke speed control function, where the learner has to setup the speed in another webpage, and return to the original webpage to test it. That means modification results cannot be seen immediately, which is very inconvenient for the learner. Some other websites provide partial stoke speed control function. For example, No.11 provides a three step speed choice: fast, medium, and slow, No.22 provides a slider bar to modify the speed, and No.17 provides a Second input box for the interval time control between stokes and delay time control between characters (Fig. 5). This study considers that Speed control is important for animation play control, and Stroke speed control is more importance than interval time control and delay time control. As to the implementation of controls, the slider bar is far better than the speed switch and Second input box, because of its immediate response, intuitiveness and ease of use for the learner.

2.8 The Writing Exercise

The writing exercise, the key objective of the Chinese character writing website, is more important than the writing demonstration part. Both online and offline exercises are provided with vocabulary lessons. The online exercises can provide real time interactive response, but the offline exercises cannot.

Online exercises include practice by tracing, practice by drawing, and practice by memory (Fig. 6). Practice by tracing outline provides a character outline for the learner to trace. Practice by drawing provides a sample character in a Grid and an empty Grid for the learner to simulate the sample and draw the character in the empty Grid. Practice by memory provides a Grid and the Pinyin of a character along with its meaning, and asks the learner to write the Chinese character independently.

Currently, No.11 and No.20 provide practice by tracing, with interactive hinting of the next stroke during the exercise. No.20 also provides practice by memory with a hinting button, which, when pressed, shows the complete tracing outline. As soon as the learner starts to draw, the outline disappears immediately. But tracing in No.11 would disappear when learner draws outside the tracing outline. No.20 lets the learner write the strokes by memory, if learner's line matches the standard position and direction of the stroke, the correct stroke will show up and pasted in the practice space, else after several errors the practice system will show out the hinting stroke.

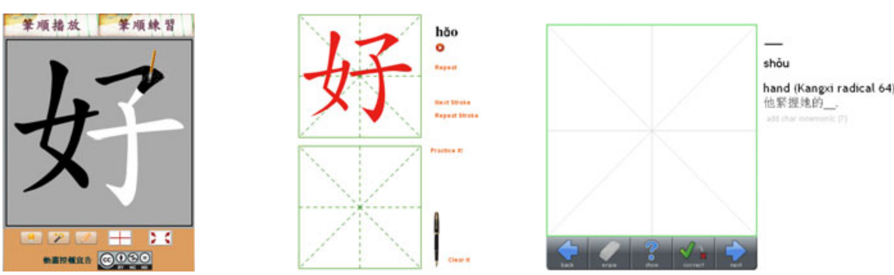


Fig. 6. Examples of online writing exercise

Practice by drawing can be found in No.6, No.18 and No.19, where No.6 and No.19 have the grid pattern inconsistency problem between the sample character and the learner's writing grid, and the learner's drawing stroke in No.19 is too thin to distinguish. Both No.18 and No.19 arrange the sample grid and the learner grid vertically, and No.6 arranges them horizontally. All websites provide clean and rewrite functions, and show the cursor as a pen icon, but the Grid pattern, pen icon, and stroke width cannot be modified by the learner, and no Evaluation function is provided. Controls for learners are: Clean, Rewrite, and Skip to next exercise.

This study purposes the following recommendations according to the analysis of the current website design for the writing exercise:

1. Online writing exercise is a better interactive method for Chinese character writing as compared to offline exercise. Practice by tracing, by drawing, and by memory should all be provided for choice by learners of different levels of proficiency.
2. The learner's strokes should be faithfully presented in the grid in practice by tracing. Both vertical and horizontal arranges should be provided for the learner to choose with.
3. Online writing exercise should provide different pen types, stroke widths, and grid patterns for the learners to choose. Practice by drawing should keep the grid patterns consistent between the sample and the exercise.
4. Evaluation mechanism should be provided for the exercises and a portfolio should be kept for each learner.
5. The exercise grid for practice by drawing should provide a translucent tracing outline as the base map for hinting, which becomes transparent with the increase of the exercise times.

3 Design Proposals

According to the survey of the current Chinese character writing websites, this study proposes a design improvement in the 8 design factors:

1. Interactive animation is the main stream for media presentation, but the stroke rendering algorithm used in animation should conform to the real situation stroke processing.
2. Four kinds of Grid patterns should be provided, namely: Blank Square, Cross, Diagonal, and Quadruple Cross. Availability a Grid pattern switch option would be even better.
3. The tracing outline as a base map is useful, and it would be even better if this function provides a switch option for the learner.
4. Color coding for both radical and ongoing strokes are a necessary function, still better if a pallet is provided for color selection with a switch option.
5. Stroke labeling for a stroke's starting and ending positions and direction vectors are very helpful for the learner, where an external labeling with a switch option is suggested. In addition, an intelligent stroke labeling which appears when stroke starts and disappears when it finishes is the state-of-the-art for match requirement.

6. Continuous stroke play function includes four operations: play, pause, stop, and repeat. Single stroke play function includes: play and repeat. The operation symbols for icons should follow the popular standard.
7. Stoke speed controls includes stroke play speed control, interval time control, and delay time control between characters. The slider bar is easier to use than other control styles. Furthermore, it not only feels clean and direct, but also responds immediately.
8. Writing exercise is a critical function of Chinese character writing website.
9. Online exercise is the best choice which provides better interactivity. There are three methods of practice: practice by tracing, practice by drawing, and practice by memory. All of the three should be provided with choice flexibility.
10. The strokes written by the learner should be presented faithfully, grid pattern and stroke width should be modifiable by the learners, and the arrangement of the sample and the practice grids should be flexible to change by left-handed writers too.
11. The evaluation function requires artificial intelligence technology, with which the process and result of a learner can be assessed and kept in a portfolio for each learner.

In addition, based on the analysis of the disadvantages of current Chinese character writing websites, and the suggestions for design improvements in the 8 factors, a few points for further improvement in webpage interface design are proposed as follows:

1. Easiness of identification and recognition should be considered for icon or symbol designs, they should be familiar to and easy to understand by the learner.
2. Because the learning materials distribute in different web pages, the process of finding out the required information from the database becomes an uncomfortable experience, that is to say, a large number of webpage entries, searches, and returns happen sequentially. This study considers that a good Chinese character writing website design should integrate the distributed information into a well constructed database, and simplify the functions used for learner interface design.
3. Any learner operation on the website should get response immediately, and get the desired feedback to learner as soon as possible.
4. Provide a clear location index of the website structure, and decrease depth of the website tree structure to avoid a labyrinth situation.
5. Digital handheld device will become more and more popular, the Chinese character writing website design should consider the display panel size and resolution, and provide the flexibility setup function to match learner's different contexts and environments.
6. Chinese being a foreigner language to the learner, the website learner's interface design should be presented in the learner's native languages, including the help messages and user guide.

4 Conclusion

The destination of this study is an overview survey for learning Chinese character writing, Tabling and cross comparing over thousand detail items, there are most important 8 factors filtered out which will be used for follow up study to purpose an improve version of interactive learning interface for Chinese character writing. Follow up study require participate with none-native learners and experts of Chinese character writing instruction. New observations, interviews, and experiments will respond the suggestion of participants. The step refinement will figure out a interactive interface template for Chinese character writing instruction websites design.

References

[1] Tan, L.H., Spinks, J.A., Eden, G.F., Perfetti, C.A., Siok, W.T.: Reading depends on writing, in Chinese. J. Proceedings of the National Academy of Sciences 102, 8781–8785 (2005)

[2] Chen, G.S., Jheng, Y.D., Lin, L.F.: Computer-based assessment for the stroke order of Chinese characters writing. In: The Second International Conference on Innovative Computing, Information and Control, Kumamoto, Japan (2007)

[3] Chen, G.S., Yao, H.C., Jheng, Y.D.: On-line assessment for the stroke order of Chinese characters writing. In: 16th International Conference on Computers in Education, Workshop Proceedings: Supplementary Proceedings of ICCE 2008, pp. 132–138 (2008)

[4] Law, N., Ki, W.W., Chung, A.L.S., Ko, P.Y., Lam, H.C.: Children's Stroke Sequence Errors in Writing Chinese Characters. Reading and Writing 10, 267–292 (1998)

[5] Taft, M., Chung, K.K.H.: Using radicals in teaching Chinese characters to second language learners. J. Psychologia 42, 243–251 (1999)

[6] Lam, H.C., Ki, W.W., Law, N., Chung, A.L.S., Ko, P.Y., Ho, A.H.S., Pun, S.W.: Designing CALL for learning Chinese characters. J. Computer Assisted Learning 17, 115–128 (2001)

[7] Lam, H.C., Pun, K.H., Leung, S.T., Tse, S.K., Ki, W.W.: Computer-assisted-learning for learning Chinese characters. J. Communications of COLIPS: An International Journal of Chinese and Oriental Languages Information Processing Society 3(1), 31–44 (1993)

[8] Luo, W.: Teaching of Chinese characters at Beijing Language and Culture University. In: Guder, A., Jiang, X., Wan, Y. (eds.) The Cognition, Learning and Teaching of Chinese Character, pp. 411–426. Beijing Language and Culture University Press (2007) [in Chinese]

Personalization in Mobile Learning for People with Special Needs

Saeed Zare

Digital Media in Education, University of Bremen
Bibliothek Str. 1, 28359 Bremen, Germany
zare@tzi.de

Abstract. The domain of learning context for people with special needs is a big challenge for digital media in education. Usage of mobile technology is growing, and it affects other technologies by bringing in new innovation and methods. The reason for this growth is not only ease of use and mobility, but also improvements in interaction and functionality in different contexts. Meanwhile, the difference between cell phones and handheld computers is becoming less and less evident. Such convergence offers the opportunity of ubiquitous learning "anytime, anywhere", so that the learners do not have to wait for a fixed time and place for learning to take place. Mobile learning can be seen as a bridge between higher level of abstracted knowledge and practical experiences, which supports the personalization in a learning process.

Keywords: Personalization, Mobile Learning, Inclusive Design, Learning Disabilities.

1 Introduction

This study has highlighted various research issues and analyses a framework that enables the personalization and autonomy in a learning process. By confronting the two needs -personalization and learning activities across real context- to the abilities of new learning technologies, it became clear that mobile learning would support personalized learning across contexts. But a major problem remained: how would it be possible for this target group to have personalized learning process and autonomously use via technologies?

In order to build extendible models, mobile learning concepts are stated and models of personalized learning processes used by teachers of students with learning and cognitive disabilities are analyzed and conceptualized. Our approach takes into consideration classrooms and workplace learning situations like small sequences of learning units for training of a certain skills, guided training for a particular machine, or workflows training for a specific task. This system should support or enable the personalization accomplishment of given tasks in order to foster the experience that cognitive complex tasks can be completed autonomously.

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 662–669, 2011.

1.1 Mobile Learning

With the expansion of mobile technology, new qualities for media contextual use cases and ubiquitous computing arise. Mobile technology is becoming a focal point of new technologies, which provide new designs, new interfaces, and new interactions. Mobile learning is a widely accepted term for describing a learning process with mobile technologies. Many researchers believe that mobile learning is situated in the future of learning [1][2], ubiquitous learning [3] and seamless learning [4]. Kevin Walker (Institute of education) [2] says:

"Mobile learning is not something that people do; learning is what people do. With technology getting smaller, more personal, ubiquitous, and powerful, it better supports a mobile society. (…) Mobile learning is not just about learning using portable devices, but learning across contexts."

This citation poses an interesting view on mobile technology and learning across context. Mobile learning is neither an extended version of e-learning nor a portable Computer-Based Training (CBT). It has its own characteristics and didactical methods as well as direct interaction between learners in context. This mobile activity is embedded in a didactical framework. One leading aspect of this framework is that mobile learning is adapting to a specific context. In this context the social interaction becomes meaningful to cognition. Not only social contexts but also relationships to objects become an important part of the context. Mobile learning can influence both the individual and community at large. Therefore, these tools impact two-way learning between groups and individuals. Learners are no longer limited to one place. Moreover, mobile devices support personalized and collaborative learning. They also let the learner interact with others face to face, instead of sitting at a personal computer.

1.2 People with Learning Disabilities

Based on statistics and census from Federal Statistical Office of Germany [6], 10.5 percent of the total population in Germany has disabilities. Also 4.2 percent of people in Germany are affected by severe mental disabilities [6] and need to be involved in personalized and individualized process of learning. It is scientifically proven that a mental disability is an intellectual impairment [7] and commonly called a cognitive disability [8] that points to the mental process of knowing.

Mental disabilities are considered as developmental disabilities that appear in childhood under the age of 18. Levels of mental disability are mostly based on the classification of the American Association on Mental Deficiency (AAMD uses IQ testing which shows the intelligence quotient, function performance of a person and the score from which their IQ test). People with the average score between 70 and 85 are considered slow learners, but not learning disabled, and people under 70 are called learning disabled [9].

2 Personalization

Personalization is one of the principles in the design of our study. Personalized learning usually occurs in traditional learning in informal ways. Traditionally, successful trainers using this method by differentiate between a learner's attitude and behaviors and through receiving learner feedback.

Hawkridge and Vincent's discussions about the use of digital media and computers by people with learning disabilities further determine the limitations and lack of this kind of personalization for learners [10]:

"Computers can ease learning difficulties. They can help learners to overcome their difficulties. They cannot work magic. They are not necessarily the best solution. Because each learner's needs are slightly different, there are few standard rules."

In 1992, Hawkridge and Vincent's citation was revolutionary. They looked toward the possibilities that digital media could provide for people with learning difficulties. At that time graphical user interfaces (GUI) were not intuitive and interactive functionalities had not been developed. However, they saw certain use cases for the disabled where the computer might be helpful.

John Traxler in the book "Mobile Learning" points to diversities, differences and individualities, which can be recognized by personalized learning and adapted to the user. Ally claims that mobile technologies support productive and meaningful processes for the learners to enhance their abilities according to their own autonomy [11].

2.1 Personalization, Individualization, and Customization

According research in software design, the analysis finds that models tend to associate personalization with individualization with respect to customization. It lies in the end-user's ability to control the device and its related data. Individualization lets teachers and learning software designers tailor materials to match scaled assessments of learner's interest whereas personalization lets the learner interact with the material on the device. In other words, individualization is a one-way process from teacher to learners while personalization is two-way.

Personalization means fitting specific content or presenting information according to an individual learner's needs. It is the capacity to tailor learning content and interactions to match learner abilities and needs that make the use of mobile technologies unique. In customization, the control of process is from the learner side and learners select material and leaning processes according to their own interests. Figure 1 depicts the differences between personalization, individualization, and customization.

The report of the teaching and learning in 2020 review group [13] argues personalization serves a moral purpose and social justice and stating:

"Put simply, personalizing learning and teaching means taking a highly structured and responsive approach to each child's and young person's learning, in order that all are able to progress, achieve and participate. It means strengthening the link between learning and teaching by engaging pupils – and their parents – as partners in learning."

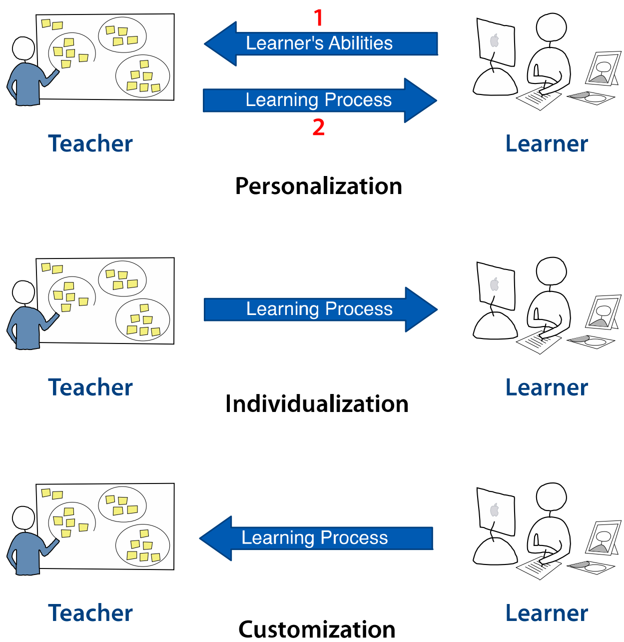


Fig. 1. Personalization, Individualization, and Customization [12]

2.2 Learner Position in a Personalized Learning

At that time, interface limitations and hardware interactions made the early tools a less appropriate choice for teachers looking for helpful learning aids. Since this time, there have been tremendous improvements in computer technologies including: hardware, user interface, interaction patterns, database, web technologies and audio-visual capabilities. Today intelligent software implemented on client-server solutions, based on fully developed interactive patterns enable us to focus on personalization and individualization. Nevertheless, developers still face challenges in building fully personalized functionalities in the core of a learning system.

Figure 2 visualizes the main concepts of personalization mentioned by [13][11][5] studies. It depicts the learner position in a personalized learning system where the learner is supported by personalized learning content, activities, assessment, evaluations and personalized planning of the learning process. Even though in this case, everything is personalized for the learner, the learner's role can be improved with the following possibilities within a learning process [12]:

- Open questioning (anytime)
- Sharing learning objectives/activities
- Success criteria
- Focused marking
- Flexible/enough time for a specific learning activity

In non-personalized systems, learners may receive high level, redundant or irrelevant blocks of information, which causes disinterest and boredom. A

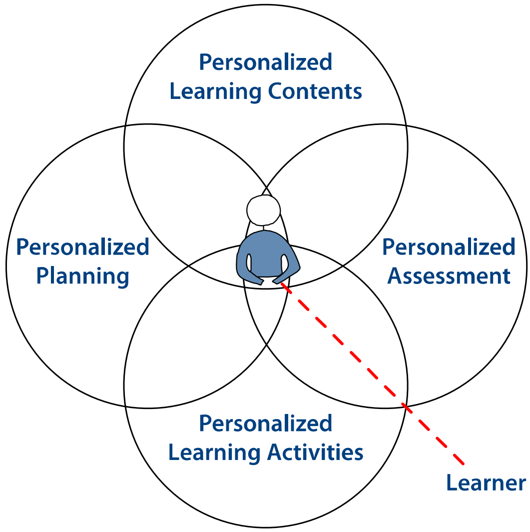


Fig. 2. Learner central to a personalized learning system [12]

personalized mobile learning system identifies an individual's profile and history as it designed to provide appropriate learning patterns, attributes and interactions based on the learners' profile. The tool is set to meet the individual needs at a time and place when and where the learner needs it.

Personalized learning contains key components that have direct affects on the learner's psychological, cognitive and social abilities. An optimal personalized learning system should focus on the importance of user autonomy, self-motivation and self-management. Learner autonomy can be greatly increased as learners improve their feelings about what they learn in their own personalized learning environment.

3 Approach for Personalization in Mobile Learning

In last decades, interface limitations and hardware interactions made the early tools a less appropriate choice for teachers looking for helpful learning aids. Today intelligent software implemented on client-server solutions, based on fully developed interactive patterns enable us to focus on personalization and partly individualization. Nevertheless, developers still face challenges in building fully personalized functionalities in the core of a learning system. Even though in this case, everything is personalized for the learner and his role that can be improved with respect to different aspects.

During our field studies in inclusive classroom, we participated different types of didactical lesson formats. With the camera, the situation focusing the interaction between teacher and student with learning impairments have been documented. Afterwards these materials were analyzed with the teacher and a psychologist according the quality and specification of the teacher student interaction. This study revealed four main results:

1. Usually the teacher develops a teaching style that allows having time for the special needs of each student with impairments. The diversity of needs brought them to establish a quality of interaction that becomes a direct interaction adapted to the individual specification.
2. Part of the learning impairment is based on a lack of self-perception and self-control.
3. The quality of feedback is related to the ability to develop self-efficiency in learning and to foster motivation.
4. Encouraging these students to imitate successful solutions and principles to cope with problems extends their learning abilities.

In learning technologies, models of personalization and adaptation according to the learning process on mobile devices are discussed. We analyzed these concepts for personalization according our four results. Doing these, we decided to reformulate a model of personalization that covers the multiple strategies of learner-centered interaction of inclusive classrooms. The following five aspects implement this model of personalization:

1. A profile that is defined at the beginning (including the initial analytical criteria).
2. An update-functionality coupled with a monitoring of user behavior and feedbacks on tests.
3. A specific algorithm for analyzing the users outcome according specific developmental models.
4. Given developmental tasks by teachers according the specific individual need of a specific student.
5. Specific learning material composed of units according an individual profile and adapted to the specification of an individual development of a specific student.

The personalization aspects render further possibilities for developing mobile learning systems for diverse groups of users and contexts. We considered the following parameters [12] in our concept to build a system that consists of technology, learner and teaching efforts:

- Adaptation of needs for the individual user and learning style
- The importance of personal well-being
- Motivating and encouraging design
- Orienting the user through the learning process
- Fostering competencies and self-confidence
- Supporting self-control or self-awareness of learner behavior during the performance, by focusing attention and recording behavior
- Different self-assessments which encourage the learners to self-regulate
- Modeling workflow according to the description of educational experts and teachers with respect to fortifying perceived competence and self-direction of the target group

In recursive implementations, experts and students have evaluated the mentioned aspects. The experts were mostly interested in the aspect of orientation and the self-confidence aspect had impressed the students. We believe that a deeper study on psychological aspect such a perception and conceptual learning can improve this model of personalization.

4 Conclusion

The suggested methodology defines a comprehensive learning process for the mentally disabled to support them in formal and informal learning. This study is part of a broader collaboration with learners and teachers for a development process with specific requirements of an engineering process. The aim of this process is to generalize all the specific requirements of the targeted field and the actions of the mentally disabled. The goal is to enable existing practices with digital media and along with establishing media that provides opportunities for a new user-appropriate practice.

The monitored learner behaviors can be analyzed according to different learning patterns. The system can use different patterns and will recognize which one is suitable for a specific learner to promote better performance. The successful learning patterns can be reused and enhanced for subsequent learning sessions. In addition, via a teacher portal, different possibilities can be added to the criteria of a learner's profile, such as ethnic, cultural background or gender.

The interaction feedback and motivating audio files can be selected from a list of predefined templates or can be manually added by the teacher. This choice may be considered similar to the well known exchange of mobile phone ring tones. In addition, a teacher portal supports visualization preview from the content which allows teachers to directly see the output-learning page according to content preparation. This can help them better orient students and prepare learning processes.

The mentioned personalization has the potential that has similarities to a teacher student interaction of classroom. These results have been successfully implemented on different mobile devices to support personalization for promoting the learning abilities. Several workshops and experiments conducted by our researchers indicate that this system provides an efficient and effective personalized learning method for the mentally disabled. These real context interactions with users and their teachers also brought forth the challenges facing this learning system.

Acknowledgments. I would like to thank Heidi Schelhowe and Michael Lund for their valuable comments and advises. Also special thanks to Digital Media in Education group at University of Bremen for their support.

References

1. Keegan, D.: Mobile Learning: The Next Generation of Learning. Distance Educational International (2005)
2. Sharples, M., Walker, K., Winters, N., et al.: Big Issues in Mobile Learning: Report of a workshop by the Kaleidoscope Network of Excellence Mobile Learning Initiative. The Learning Sciences (2007)

3. Rogers, Y., Price, S., Randell, C., Stanton-Fraser, D., Weal, M., Fitzpatrick, G.: Ubi-learning: integrating outdoor and indoor learning experiences. Communications of the ACM (2005)
4. Chan, T., Roschelle, J., Hsi, S., Kinshuk, S., Sharples, M., Brown, T., et al.: One-to-one technology-enhanced learning: an opportunity for global research collaboration. Research and Practice in Technology Enhanced Learning (2006)
5. Clarke, J.: Changing Systems to Personalize Learning. Introduction to the Personalization Workshops. The Education Alliance at Brown University (2003)
6. Schwerbehinderte Menschen 2007. Wirtschaft und Statistik 2/2010. Statistisches Bundesamt, Wiesbaden (2010), `http://www.destatis.de`
7. Barbotte, E., Guillemin, F., Chau, N.: The Lorhandicap Group: Bulletin of the World Health Organization (WHO): Prevalence of impairments, disabilities, handicaps and quality of life in the general population: a review of recent literature, Genebra, vol. 79(11) (2001) ISSN 0042-9686
8. Clark, J.: How Do Disabled People Use Computers. In: Building Accessible Websites (2002) ISBN 0-7357-1150-X
9. Pitsch, H.J.: How does a trainer working with the mentally disabled differ from anyother teacher or trainer?: Agora XII. Training for mentally disabled people and their trainers. Permitting the mentally disabled a genuine and appropriate exercise of their rights. Cedefop Hrsg.: Thessaloniki, July 5-6 2001. Office for Official Publications of the European Communities, 127-140 Luxembourg (2003)
10. Hawkridge, D., Vincent, T.: Learning Difficulties and Computers: Access to the Curriculum. Jessica Kingsley Publishers (1992) ISBN-13: 978-1853021329
11. Ally, M.: Mobile Learning. Transforming the Delivery of Education and Training. Published by AU Press, Athabasca University (2009) ISBN 978-1-897425-43-5
12. Zare, S.: Intelligent Mobile Learning Interaction System (IMLIS): A Personalized Learning System for People with Mental Disabilities. University of Bremen (2010)
13. 2020 Vision, Report of the Teaching and Learning in 2020 Review Group. DfES Publications (2006) ISBN: 978-1-84478-862-0

Intelligent Provision of Semantic Web Services in Educational Systems

Marija Zelić[1], Marko Rosić[1], and Vlado Glavinić[2]

[1] Faculty of Science, University of Split, Teslina 12, 21000 Split, Croatia
[2] Faculty of Electrical Engineering and Computing, University of Zagreb, Unska 3, 10000 Zagreb, Croatia
{marija.zelic,marko.rosic}@pmfst.hr, vlado.glavinic@fer.hr

Abstract. The developed adaptation techniques in adaptive educational systems can be quite efficient; however, the systems' learning resources are usually stored in closed environments. This seems insufficient considering that the present-day learning environment encompasses large volumes of data and services on the Web. Our research focuses on the development of open, dynamic and adaptive systems that provide universal access to knowledge. This paper presents a novel unifying architecture for e-learning systems that enables intelligent provision of semantic Web services pertinent to the field of e-learning. The SWEA architecture enables the complete automation of the learning process by using a set of different agent types that perform semantic matchmaking of the required services.

Keywords: Semantic Web, semantic Web services, agent technology, e-learning systems.

1 Introduction

The research on adaptive educational systems and intelligent user interfaces puts the user at the center of all developments, by taking into account the users' heterogeneity - their different background, interests, learning styles, expertise and goals. Even though some of the developed adaptation techniques are quite efficient, the systems' learning resources are usually stored in closed environments and accessible only to closed communities [1]. This, however, seems insufficient considering that the present-day learning environment encompasses large volumes of data and services on the Web, and offers an abundance of unused learning resources. Even more, the ubiquitous computing environment provides users with instant access to knowledge and information infrastructure services, in any place and at any time, by using different access devices - spanning from low-performance mobile phones to high-performance laptops. Consequently, the term adaptivity gets a much wider context - besides the requirements on an intuitive, dynamic and adaptive human-computer interaction, educational systems should provide universal access to knowledge and support the user in finding the relevant learning resources on the Web.

Consequently, our research focuses on the latest research and development technologies that enable adaptivity in the context of a ubiquitous learning

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 670–679, 2011.

environment - the Semantic Web and the agent technology. This paper presents a novel unifying architecture for e-learning systems that enables intelligent provision of semantic Web services pertinent to the field of e-learning. SWEA (Semantic Web-based E-learning Architecture) enables the complete automation of the learning process by using a set of different agent types that perform semantic matchmaking of the required services, based on the learner profile and the learning situation. The agent technology has already proven effective in the development of distributed, dynamic and intelligent systems; however, the agents' ability to automatically reason about data and services on the Web greatly enhances system adaptivity. Our research is based on initial studies that have investigated different approaches to the interoperation of agents and semantic Web services, reporting on the great benefits of their combination. As there is no comprehensive study on an overall e-learning architecture based on the integration of the aforementioned technologies, our objective is to model SWEA and enable the development of adaptive e-learning systems in a ubiquitous learning environment. The rest of the paper is structured as follows. Section 2 elaborates on the technology infrastructure, while Section 3 presents the SWEA architecture along with a lightweight prototype implementation to illustrate the overall system behavior. Section 4 provides the concluding remarks and outlines guidelines for future work.

2 Technology Infrastructure

Technology enhanced learning takes many different forms, including formal and informal learning, individual and collaborative, desktop, mobile and ubiquitous learning, etc. Hence the educational systems should support effective behavior in real-world, dynamic and distributed environments. In such a context, the agent technology is considered one of the most innovative approaches to designing and implementing open, dynamic and intelligent systems [2]. Accordingly, the systems comprise autonomous, reactive and proactive components, i.e. agents, which usually represent animate objects - typically human beings. One of the inherent agent properties is the ability to form social structures (e.g. teams, communities, coalitions), called multi-agent systems (MASs). Agents intrinsically model the phenomena of the real world into a distributed software system. Standardization efforts in the field are led by non-profit associations of companies and organizations, such as FIPA (Foundation for Intelligent Physical Agents) and OMG Agent PSIG (Agent Platform Special Interest Group, subgroup of the Object Management Group). A de-facto standard middleware for the development of agent-based applications is JADE (Java Agent DEvelopment Framework). The agent technology has been the subject of discussion and investigation within the scientific community for several years already, but it is only recently that it has received a great deal of attention due to its relation to the Semantic Web.

The emergence of World Wide Web has provided users with access to vast volumes of distributed data and information services. However, search engines are not intelligent and provide results numbering in hundreds of thousands, many of them containing irrelevant pages. Berners-Lee has envisioned the Web where computers are capable of analyzing all the data on the Web [3]. This vision has become known as the Semantic Web (SW) and is now an initiative led by the W3C with participation of a large number of researchers and industrial partners. The Semantic Web extends the present-day Web by defining the semantics of data and services on the Web in a precise, machine-understandable form. This enables agents to find and reason about specific information on the Web, thus solving the aforementioned problems of information overload and searching the information space. Ontologies are used as shared vocabularies that enable effective and accurate communication of meaning, hence enabling knowledge sharing and reuse. In addition, ontologies can be used as reusable components in modeling of intelligent and knowledge-based applications. The OWL Web Ontology Language is a W3C standard for publishing and sharing ontologies on the Web.

In the last decade, more and more organizations are adopting the service-oriented approach to infrastructure development. The service-oriented architecture enables the systems to be assembled from a loosely coupled collection of Web services. Web services provide a standard means of interoperating between different applications, running on different platforms, and are therefore considered ideal for the development of distributed and interoperable systems. The concept of semantic Web services (SWS) can be traced in an article published in 2001 [4]. The main idea behind semantic Web services is to ensure semantic annotation of the services, processes, message-based conversations, transactions, etc, thus enabling automation of service discovery, execution, composition as well as interoperation. Semantic Web services are described by using ontology-based expressions, hence enabling agents to perform semantic matching of service capabilities, e.g. if an agent cannot find the appropriate service, it can search for a service that provides a partial match. Semantic Web services turn the Web into a collection of reusable resources that implement various useful functions and enable the development of intelligent distributed applications. The technology is based on a set of W3C member submissions; however, OWL-S (Semantic) and WSMO (Web Service Modeling Ontology) are considered the main ones.

Even though both the agent technology and semantic Web services enable the development of open, dynamic and service-based systems, they are conceptually different - agents provide services based upon the principles of autonomy, social ability, proactive, reactive and goal-directed behavior, while Web services are passive and provide distributed functionality [5]. Initial studies have been conducted to explore different approaches to their interoperation and/or integration [4], [5], [6], [7], [8], and delivered promising results that show enormous potential in the development of more powerful and intelligent applications. Examination of these studies reveals that a distinction should remain between high-level agent societies and Web services - agents are to use Web services as external functionality. In the last few years a number of research projects have been carried out in the field of e-learning, which deal with the use of either the agent technology or Semantic Web technologies for realizing specific aspects of e-learning systems. However, only a very few research projects actually exploit the combination of the agent technology and Semantic Web technologies, e.g. [9], [10]. Various tools that facilitate semantic Web service authoring, discovery, monitoring, etc., are underway, either as part of an execution environment for a specific W3C initiative (e.g. WSMX [11]) or as separate research initiatives (e.g. at Carnegie Mellon University and Stanford University).

3 SWEA E-Learning Architecture

As there is no comprehensive study on an overall e-learning architecture based on the integration of the aforementioned technologies, our objective is to model the SWEA architecture and enable the development of open, dynamic and adaptive e-learning systems that provide universal access to knowledge. Based on the idea of providing a service-oriented technical infrastructure, SWEA encompasses a set of distinct services that encapsulate the application logic. However, the notion of a service is expanded and modeled into different layers - semantic Web services provide passive functionality, while agents are active components that complement the Web service technology. As the Semantic Web is the underlying infrastructure, the agent-based services enable various levels of intelligent and adaptive system behavior, hence improving the efficiency and effectiveness of human-computer interaction. Agents and semantic Web services are loosely coupled components with standards-based interfaces, and can be deployed by different universities and organizations. Consequently, various customized e-learning applications can be developed by consuming different services. This approach offers great possibilities in terms of modularity, reusability and interoperability. SWEA logical structure has three layers that are discussed as follows (Fig. 1).

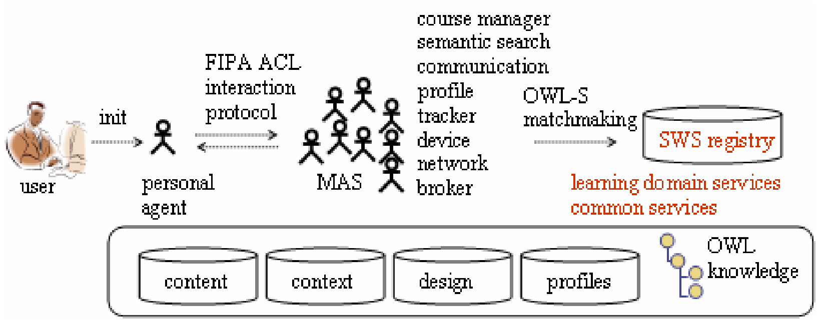


Fig. 1. SWEA e-learning architecture

3.1 Functionality and Layers

The data layer provides a systematic view of ongoing research in Semantic Web-based education and identifies the general categories of ontologies related to various aspects of the learning process (Table 1). Ontological engineering in education is an active research field and mostly relies on the conventional metadata standards. The most referred initiatives are [12], [13], and [14]. In addition to enabling knowledge sharing and reuse, SWEA ontologies also enable adaptivity of the learning process. For instance, if an agent has access to ontology-based knowledge on the learner and her/his learning situation, it can find the matching services and retrieve customized learning content and activities. Due to their expressiveness and support for reasoning, ontologies have been used in a number of adaptive systems.

Table 1. SWEA ontologies

Ontology	Description
Content	A family of ontologies for different learning domains, e.g. physics, biology, music, etc. Learning content should be annotated with domain ontology concepts so as to enable semantic search and reasoning about data.
Learning context	Learning or presentation context of the learning content. A simple learning object can be in the form of an introduction, a discussion, an exercise, an example, a figure, a presentation, etc. The learning context description enables context-aware retrieval of learning resources.
Mobile/ubiquitous context	Spatial and temporal aspects of the learning situation, e.g. location, time, delivery media, surrounding objects, etc. Context modeling enables adaptation of the system to the environment without explicit user intervention.
Learning design	Formal representation of the building blocks of an instructional design. The course structure is dynamic because it depends on the learner profile and comprises learning objects related in different ways, e.g. a learning object can precede, contain or reference another learning object.
Learner profiles	Representation of the learner's characteristics: her/his personal information, preferences, performance, learning history, etc. Managing a learner profile is an intensive research field, with techniques ranging from stating of preferences to machine learning techniques.

Table 2. e-Framework services

Services	Description
Common services	Basic services - e.g. logging, alert, forum, chat, authentication and authorization, federated search, archiving, scheduling, etc.
Learning domain services	Domain-specific services, e.g. (i) course creation: curriculum, course management, (ii) course delivery: activity management, learning flow, (iii) assessment: marking, grading, (iv) record keeping: reporting, e-portfolio, quality assurance, etc.

The semantic Web services layer in SWEA is the data layer's front-end, comprising services that retrieve and update content from the shared repositories. Our idea stems from the e-Framework initiative [15], an international effort to establish a service-oriented technical framework for designing e-learning infrastructures. The underlying goal is to provide a common set of service definitions, so that the systems can expose their application logic as services and share them with other applications. Applications can be developed or acquired as needed, by utilizing the services to create customized solutions. This enables the development of modular and flexible systems, and also supports cross-organizational activities. SWEA puts the services defined by e-Framework in the Semantic Web context. The service descriptions, i.e. their interface, functionality, behavior, prerequisites and consequences, are based on SWEA ontologies, hence enabling agents to perform their automated discovery, selection, composition and invocation. The e-Framework provides a set of reference models and a set of services to be populated by the e-learning community (Table 2).

The agent-based layer serves as an application layer that efficiently and effectively utilizes resources from the underlying layers. The agent technology is increasingly being considered as an appropriate support to the service-oriented architecture, since it is communication-centric, relies on distributed systems and involves autonomous and proactive components [16]. SWEA identifies a set of different agent types that perform assigned tasks by communicating among each other, searching and reasoning about semantic Web services that are suitable for the task at hand. Their knowledge and communication are based on SWEA ontologies, so as to enable semantic interoperability. The agent-based layer provides a high-level functionality of the architecture and enables the development of e-learning systems by instantiating suitable agents. Hence, the agents can be reused in different contexts to fit a variety of pedagogic and organizational models. SWEA is to provide a thorough specification of the agents and their functionality, behavior, interaction templates, etc. Table 3 identifies some of the basic agents.

Table 3. SWEA basic agents

Agent	Description
Broker	Mediator between service requestors and service providers. Performs automated discovery of semantic Web services, negotiates with service providers and decides on the optimal solution.
Device	Performs dynamic content adaptation encompassing layout changes, reconfigurations of content format, etc., basing on the access device profile.
Network	Provides information on the Quality-of-Service (QoS) values of the wireless network currently in use.
Course manager	This agent guides the learner through the learning session based on her/his profile (it decides whether the learner should see additional explanations on a topic, take a test, etc.), thus performing adaptation of the learning process.
Semantic search	Performs semantically enhanced searches across the shared repositories, using semantic matching and reasoning capabilities.
Communication	Enables communication among learners, creation of different chat rooms, forming of buddy lists, transferring files, etc.
Profile	Monitors the learners' behavior and uses data mining techniques on collected data to extract behavior patterns and rules, thus enabling better adaptation to learners.
System	System agents manage the operation of the agent platform, e.g. creation of agents, monitoring of agents, discovery of agents, their mobility, etc. System agents are usually provided with the agent platform itself.
Personal	Provides the user with a GUI. This agent is engaged in communication and information sharing with the other agents so as to support the user in accomplishing her/his tasks. The agent also performs routine support tasks and altogether provides the user with an enjoyable learning session.
Tracker, Monitor	Local tracking and monitoring of the user's behavior, updating her/his local profile, suggesting subsequent actions, recommending events, reacting to system notifications, etc.

3.2 Adaptivity and Context-Awareness Issues

Over the last few years, a number of adaptive systems have been exploiting ontologies for the purpose of user modeling. Ontology features, i.e. available formats for sharable knowledge representation and support for reasoning, have proven to be useful for adequate representation of knowledge about a user and for performing adaptation of the system to the user.

Even though most information modeled about the user is typical, the actual profile representation may vary from system to system. For instance, both IEEE PAPI and IMS LIP specifications deal with user profiles from very different perspectives: LIP uses rich constructs for representing different learner aspects, while PAPI is more focused on the performance proper. Therefore, the existing projects use one of the two main approaches to the problem of interoperability of user profiles:

- Development of a central user profile ontology that acts as a common ground for multiple applications, e.g. GUMO (General User Model Ontology) [17].
- Manual mapping of existing user profile ontologies for the purpose of either identifying the conversion rules or developing an integrated model, e.g. in [12] the authors have developed a user profile that inherits categories from both PAPI and LIP, and extends them with new concepts to support current learning trends.

As this area of research is rather young, some of the activity may seem unbalanced and solutions to some problems have not yet been proposed.

In addition to providing personalized contents based on the user's characteristics, it is important for the system to adapt to the learning situation itself. The underlying idea of present-day educational systems is to provide users with "anytime, anywhere" learning experience. The advent of mobile technologies and wireless infrastructure can be leveraged to improve the learning outcomes, e.g. students might use their mobile devices for spontaneous and short study phases. Any information about the situation in which the learning process occurs can be referred to as *context*. In particular, a ubiquitous learning environment encompasses two underlying contexts:

- *The learning context* is mostly characterized by learning objects and learning paths, pedagogical approaches, tools, materials, people (in terms of social networking) and learning activities.
- *The mobile/ubiquitous context* refers to the spatial and temporal aspects of the situation, such as the location, delivery media, surrounding objects, etc.

Adequate representations of contextual information are essential for providing context-aware learning contents. Ontologies are particularly suitable in this aspect, due to their flexibility, expressiveness, extendibility and support for reasoning. Various approaches in modeling the context utilize ontologies as the underlying context model, e.g. CoOL (Context Ontology Language) [13], CONON (CONtext ONtology) [18], etc. However, there is still a need for standardized formats and protocols in the development of context-aware systems.

The data layer specification in SWEA encompasses ontologies for modeling of users as well as contextual information, so as to enable intelligent agents to perform adequate adaptations of the learning process and hence provide users with an enjoyable and successful learning experience. The following section describes a small learning scenario along with a lightweight prototype implementation to illustrate the overall system behavior.

3.3 Prototype

A lightweight prototype has been developed so as to test the technical feasibility of the approach. The learner is able to access her/his university portal and select an introductory course on computer science. The learner's personal agent informs the university's course manager agent on the learner's request, her/his profile and the terminal device profile. Based on the collected data, the course manager agent decides on the learning path and invokes the suitable services - if the learner has accessed the course for the first time, she/he is presented with an initial quiz; otherwise the suitable learning content is presented. The available services are a course delivery service and a quiz service. The services obtain the learning resources from the shared repositories. After obtaining the learning resources, the course manager agent builds the GUI for resource delivery and sends it to the personal agent for display. The following illustrations show a fragment of the domain ontology (Fig. 2), a process description of a semantic Web service that returns a question item for a specified concept in the domain ontology, and the result of service execution on the GUI of the learner's personal agent deployed on a mobile device simulator (Fig. 3).

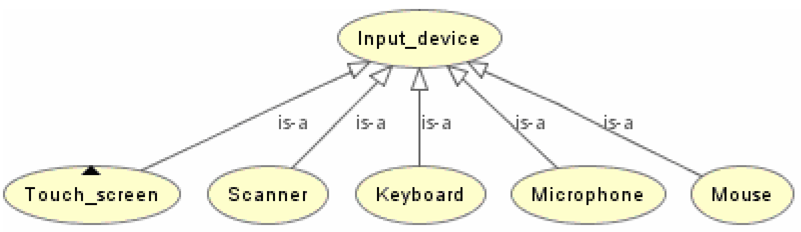


Fig. 2. A fragment of the domain ontology on computer system components

Process description of the Semantic Web Service ComsysQuizService

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF
 xml:base="http://localhost/swea/owl/ComsysQuizService.owl#">
  <process:AtomicProcess rdf:ID="ComsysQuizProcess">
    <service:describes rdf:resource="#ComsysQuizService"/>
  </process:AtomicProcess>
  <process:Input rdf:ID="Concept">
    <process:parameterType
     rdf:datatype="http://www.w3.org/2001/XMLSchema#anyURI">
        http://localhost/swea/owl/Comsys.owl#Concept
    </process:parameterType>
  </process:Input>
  <process:Output rdf:ID="QuestionItem">
    <process:parameterType
     rdf:datatype="http://www.w3.org/2001/XMLSchema#anyURI">
        http://localhost/swea/owl/Comsys.owl#QuestionItem
    </process:parameterType>
  </process:Output>
</rdf:RDF>
```

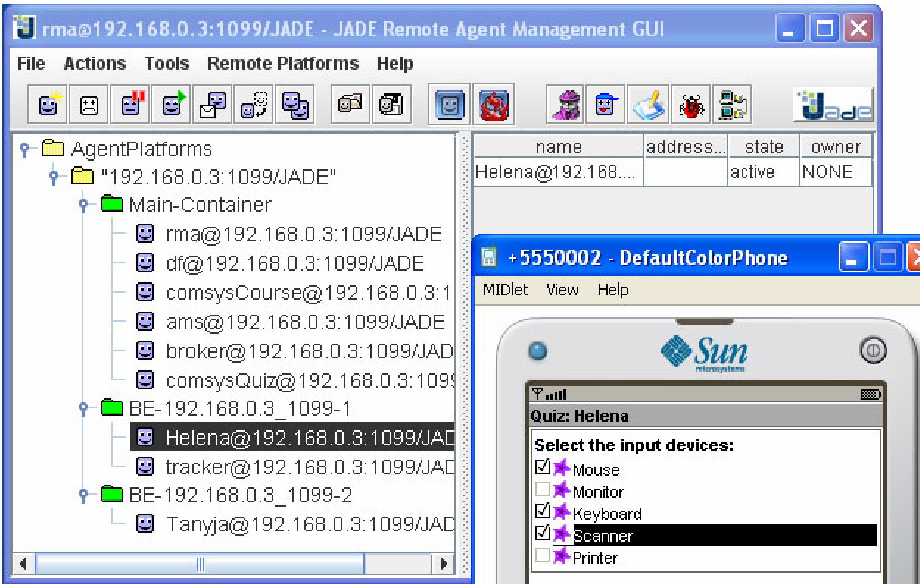


Fig. 3. The service result displayed on the GUI of the learner's personal agent

4 Conclusion and Future Work

Today's knowledge society places immense importance on learning. However, the present-day e-learning systems do not meet the needs for adaptive learning experience in open, dynamic and distributed environments. The e-learning systems should be active companions that enable users to seamlessly acquire relevant learning content and activities, communicate with their peers, in any place and at any time, by using any device. The learning technology developments have identified the need for a stable technical infrastructure that would provide a common basis for the development of e-learning systems. This paper presents a novel unifying architecture named SWEA that encompasses a set of loosely coupled components - agents that provide active functionality by communicating among each other and searching for the required services, and semantic Web services that provide distributed functionality. The architecture is based on the Semantic Web, so as to enable intelligent and adaptive system behavior. Our future work involves the definition of thorough specifications for each layer and the development of a complete prototype system that is to be evaluated with specialized evaluation methods.

Acknowledgements. This paper describes the results of research being carried out within the project 036-0361994-1995 *Universal Middleware Platform for e-Learning Systems*, as well as within the program *Intelligent Support for Ubiquity of e-Learning Systems*, both funded by the Ministry of Science, Education and Sports of the Republic of Croatia.

References

1. Devedžić, V.: Semantic Web and Education. Springer, Berlin (2006)
2. Luck, M., Ashri, R., d'Inverno, M.: Agent-Based Software Development. Artech House Publishers, Norwood (2004)
3. Berners-Lee, T., Fischetti, M.: Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web. HarperCollins Publishers, New York (2000)
4. McIlraith, S.A., Son, T.C., Zeng, H.: Semantic Web Services. IEEE Intelligent Systems 16, 46–53 (2001)
5. Blacoe, I., Portabella, D.: KWeb/IST-2004-507482 Deliverable D2.4.4 – Guidelines for the integration of agent-based services and web-based services (2005)
6. Ermolayev, V., Keberle, N., Plaksin, S., Kononenko, O., Terziyan, V.Y.: Towards a Framework for Agent-Enabled Semantic Web Service Composition. International Journal of Web Services Research 1, 63–87 (2004)
7. Gibbins, N., Harris, S., Shadbolt, N.: Agent-based Semantic Web Services. Web Semantics: Science, Services and Agents on the World Wide Web 1, 141–154 (2004)
8. Stollberg, M., Lausen, H., Lara, R., Haller, A., Fensel, D.: D3.2 v0.1 WSMO Use Case Modeling and Testing (2004)
9. Kerschberg, L., Jeong, H., Kim, W.: Emergent Semantics in Knowledge Sifter: An Evolutionary Search Agent Based on Semantic Web Services. In: Spaccapietra, S., Aberer, K., Cudré-Mauroux, P. (eds.) Journal on Data Semantics VI. LNCS, vol. 4090, pp. 187–209. Springer, Heidelberg (2006)
10. Lopes, A.L., Botelho, L.M.: Executing Semantic Web Services with a Context-aware Service Execution Agent. In: Huang, J., Kowalczyk, R., Maamar, Z., Martin, D., Müller, I., Stoutenburg, S., Sycara, K. (eds.) SOCASE 2007. LNCS, vol. 4504, pp. 1–15. Springer, Heidelberg (2007)
11. Bussler, C., Cimpian, E., Fensel, D., Gomez, J.M., Haller, A., Haselwanter, T., Kerrigan, M., Mocan, A., Moran, M., Oren, E., Sapkota, B., Toma, I., Viskova, J., Vitvar, T., Zaremba, M., Zaremba, M.: Web Service Execution Environment (WSMX) (2005)
12. Dolog, P., Nejdl, W.: Challenges and Benefits of the Semantic Web for User Modelling. In: Proc. Workshop on Adaptive Hypermedia and Adaptive Web-Based Systems at the 12th International World Wide Web Conference, Budapest, pp. 1–12 (2003)
13. Strang, T., Linnhoff-Popien, C., Frank, K.: CoOL: A Context Ontology Language to enable Contextual Interoperability. In: Stefani, J.-B., Demeure, I.M., Hagimont, D. (eds.) DAIS 2003. LNCS, vol. 2893, pp. 236–247. Springer, Heidelberg (2003)
14. Jovanović, J., Gašević, D., Knight, C., Richards, G.: Ontologies for Effective Use of Context in e-Learning Settings. Educational Technology & Society 10, 47–59 (2007)
15. The e-Framework for Education and Research (2009), `http://www.e-framework.org`
16. Poggi, A., Tomaiuolo, M., Turci, P.: An Agent-Based Service Oriented Architecture. In: Proc. 8th AI*IA/TABOO Joint Workshop From Objects to Agents: Agents and Industry: Technological Applications of Software Agents, Genova, pp. 157–165 (2007)
17. Heckmann, D.: Ubiquitous User Modeling. Akademische Verlagsgesellschaft Aka GmbH, Berlin (2006)
18. Wang, X., Zhang, D., Gu, T., Pung, H.K.: Ontology Based Context Modeling and Reasoning using OWL. In: Proc. 2nd IEEE Conference on Pervasive Computing and Communications Workshops, Orlando, pp. 18–22 (2004)

Correction to: Intercultural Dynamics of First Acquaintance: Comparative Study of Swedish, Chinese and Swedish-Chinese First Time Encounters

Jens Allwood, Nataliya Berbyuk Lindström, and Jia Lu

Correction to:
Chapter "Intercultural Dynamics of First Acquaintance: Comparative Study of Swedish, Chinese and Swedish-Chinese First Time Encounters" in: C. Stephanidis (Ed.):
***Universal Access in Human-Computer Interaction*, LNCS 6768,**
https://doi.org/10.1007/978-3-642-21657-2_2

In an older version of this paper, there was an orthographical error in the title. The letter "r" was missing from the word "First". This has been corrected.

The updated online version of this chapter can be found at
https://doi.org/10.1007/978-3-642-21657-2_2

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, p. C1, 2021.
https://doi.org/10.1007/978-3-642-21657-2_73

Author Index