ROCHI – INTERNATIONAL CONFERENCE ON HUMAN-COMPUTER INTERACTION

Proceedings of the 13th International Conference on Human-Computer Interaction RoCHI'2016, 8-9 September, Iasi, Romania

Edited by:

Adrian Iftene

A.I. Cuza University of Iasi, Romania

and

Jean Vanderdonckt

Université catholique de Louvain, Belgium

MATRIX ROM Bucharest, 2016 Publisher

MATRIX ROM

C.P. 16-162 062510 – București, România Tel.: 021 4113617, Fax: 021 4114280 E-mail: office@matrixrom.ro www.matrixrom.ro

MatrixRom publishing house is a certified publisher by the National Council for Scientific Research in Higher Education (Consiliul Național al Cercetării Științifice din Învățământul Superior)

Cover graphic design: *Sabin-Corneliu Buraga*, A.I.Cuza University, Iasi Graphic identity: *Adrian Mironescu*, Idegrafo

© Copyright 2016

All rights of the current edition are reserved by MATRIX ROM. No part of this work may be reproduced or transmitted in any form or by any means, electronic or mechanical, without permission in writing from the publisher.

RoCHI – International Conference on Human Computer Interaction (2016)

http://rochi.utcluj.ro/proceedings/en/index.php

ISSN 2501-9422 ISSN-L 2501-9422

TABLE OF CONTENTS

Organiz Forwor	zation d	iz
Invited •	paper "The design is right and the right design" – How to measure and theorize "right" in HCI and Software Engineering?	
Usabilit •	ty and accessibility evaluation Implementation of a polyglot text-to-speech synthesis in two assistive technologies Paul Fogarassy-Neszly, Aurel Patru, Dragos Daniel Iordache and Costin Pribeanu	,
•	Comments on the reliability and validity of UMUX and UMUX-LITE short scales	
•	Automated Evaluation of Menu by Guidelines Review Sara Bouzit, Gaëlle Calvary, Denis Chêne and Jean Vanderdonckt	1
•	Perceived aesthetics of user-modifiable layouts: a comparison between an unspecified design and a GUI Stefan Trausan-Matu, Brahma Dathan	22
User in •	terface development Enactment of User Interface Development Methods in Software Life Cycles <i>Iyad Khaddam, Hanaa Barakat and Jean Vanderdonckt</i>	20
•	Increasing the Accuracy of Indoor Localization Applications by Using Predefined Markers and the Phone's Camera	30
•	Using WSN and Mobile Apps for Home and Office Ambient Monitoring and Control Catalin Damian, Lenuta Alboaie and Adrian Iftene	42
•	Sphero - Multiplayer Augmented Game (SMAUG) Marian-Nicolae Pinzariu, Adrian Iftene	40
Interac •	tion Techniques Reducing Gestural Interaction Physical Stress in Virtual Environments: an Experiment Sobhi Ahmed, Laure Leroy and Ari Bouaniche	50
•	Oculus Rift 3D Interaction and Nicotine Craving: Results from a Pilot Study Ioana-Monica Ciolan, Sabin Buraga and Ion Dafinoiu	58
•	Brain Computer Interface using Machine Learning Cristian-Valeriu Soare	6
•	Methodology for Identification and Evaluation of Web Application Performance Oriented Usability Issues	
	Mihaela Ciugudean, Dorian Gorgan	69

HCI in e-learning systemsTesting the technolo

Her me rearing systems	
• Testing the technology acceptance model with Romanian university students	76
Alexanaru Balog, Costin Pribeanu	
• Measuring the negative effects of the Facebook dependence on the students' university work.	81
Gabriel Gorghiu, Valentina Iuliana Manea, Dragoș Daniel Iordache and Costin Pribeanu	01
• Evaluation of the Tesys e-Learning platform's interface Paul Stefan Popescu, Cristian Mihaescu, Mihai Mocanu and Costel Ionascu	86
MOOCBuddy: a Chatbot for personalized learning with MOOCs <i>Carmen Holotescu</i>	91
Interactive applications	
ReaderBench goes Online: A Comprehension-Centered Framework for Educational	
Purposes Marius-Gabriel Gutu, Mihai Dascalu, Stefan Trausan-Matu, Philippe Dessus	95
• Event detection in Tweets Andrei Bogdan Baran, Adrian Iftene	103
Rhythm analysis of texts using Natural Language Processing <i>Irina-Diana Niculescu and Stefan Trausan-Matu</i>	107
• Terrain Synthesis from Crude Heightmaps Alexandre Philippe Mangra, Adrian Sabou, Dorian Gorgan	113
Author Index	119

ORGANIZERS

Conference Chair

Adrian Iftene, Alexandru Ioan Cuza University, Iasi, Romania

Program Committee Chair

Jean Vanderdonckt, Université catholique de Louvain, Belgium

Organizing Committee Chair

Lenuța Alboaie, Alexandru Ioan Cuza University, Iasi, Romania

RoCHI Conference Website Administrator

Teodor Stefanut, Technical University Cluj-Napoca, Romania

ORGANIZING COMMITTEE

Andrei Arusoaie, Alexandru Ioan Cuza University, Iaşi, Romania Corina Forăscu, Alexandru Ioan Cuza University, Iaşi, Romania Daniela Gîfu, Alexandru Ioan Cuza University, Iaşi, Romania Alex Moruz, Alexandru Ioan Cuza University, Iaşi, Romania Vlad Rădulescu, Alexandru Ioan Cuza University, Iaşi, Romania Cosmin Vârlan, Alexandru Ioan Cuza University, Iaşi, Romania

PROGRAM COMMITTEE

Alexandru Balog, ICI Bucharest, Romania Sabin-Corneliu Buraga, Alexandru Ioan Cuza University, Iași, Romania Gilbert Cockton, Northumbria University, United Kingdom Grigoreta Sofia Cojocar, Babes-Bolyai University, Cluj-Napoca, Romania Karin Coninx, Hasselt University, Belgium Marian Dârdală, University of Economic Studies, Bucharest, Romania Mihai Dascălu, University Politehnică of Bucharest, Romania Philippe Dessus, Université Grenobles Alpes, France Bruno Dumas. Université de Namur, Belgium Paul Fogarassy-Neszly, BAUM Engineering, Arad, Romania Peter Forbrig, University of Rostock, Germany Alfonso Gracia Frey, Luxembourg Institute of Science and Technology, Luxembourg Vivian Genaro Motti, George Mason University, United States Juan Gonzalez Calleros, Benemérita Universidad Autónoma de Puebla, Mexico Dorian Gorgan, Technical University Cluj-Napoca, Romania Jan Gulliksen, KTH Royal Institute of Technology, Sweden Adriana-Mihaela Guran, Universitatea Babeş-Bolyai, Cluj-Napoca, Romania Adrian Iftene, Alexandru Ioan Cuza University, Iaşi, Romania Dragoş Daniel Iordache, ICI Bucharest, Romania David Lamas, Tallinn University, Estonia Alin Moldoveanu, University Politehnică of Bucharest, Romania Philippe Palanque, IRIT, Université Paul Sabatier de Toulouse, France Oscar Pastor, Polytechnic University of Valencia, Spain Lyn Pemberton, Brighton University, United Kingdom Dorin Mircea Popovici, Universitaty Ovidius, Constanta, Romania Costin Pribeanu, ICI Bucharest, Romania Traian Eugen Rebedea, University Politehnică of Bucharest, Romania Adriana Elena Reveiu, University of Economic Studies, Bucharest, Romania Cristian Rusu, Pontificia Universidad Católica de Valparaíso, Chile Carmen Santoro. ISTI-CNR, Pisa, Italy **Dominique Scapin,** INRIA, France Ahmed Seffah, Lappeeranta University of Technology, Finland Marcin Sikorski, Gdansk University, Poland Jeam-Sebastien Sotted, Luxembourg Institute of Science and Technology, Luxembourg Christian Stary, University of Linz, Austria Teodor Stefănuț, Technical University Cluj-Napoca, Romania Stefan Trăușan-Matu, University Politehnică of Bucharest, Romania Jean Vanderdonckt, Universite catholique de Louvain, Belgium Radu Daniel Vatavu, University of Suceava, Romania Panayiotis Zaphiris, Cyprus University of Technology, Cyprus

ADDITIONAL REVIEWERS

Lenuța Alboaie, Alexandru Ioan Cuza University, Iași, Romania Felix Albu, Valahia University of Târgovișt, Romania Adriana Alexandru, ICI Bucharest, Romania Dragos Catalin Barbu, ICI Bucharest, Romania Verginica Barbu-Mititelu, ICIA Bucharest, Romania Oana Balan, University Politehnică of Bucharest, Romania Elena Băutu, Universiy Ovidius, Constanta, Romania Victor Bâcu, Technical University Cluj-Napoca, Romania Tiberiu Boros, ICIA Bucharest, Romania Alexandru Butean, University Politehnică of Bucharest, Romania Dumitru Clementin Cercel, University Politehnică of Bucharest, Romania Mihaela Colhon, University of Craiova, Romania Liviu Cotfas, University of Economic Studies, Bucharest, Romania Titus Felix Furtună, University of Economic Studies, Bucharest, Romania Florin Girbacia, University Transilvania, Brasov, Romania Gabriela Grosseck, University of West, Timisoara, Romania Carmen Holotescu, Ioan Slavici University, Timisoara, Romania Vincentas Lamanauskas, Siauliai University, Lithuania Valentina Marinescu, University of Bucharest, Romania Cornelia Melenti, Technical University Cluj-Napoca, Romania Vasile Dănuț Mihon, Technical University Cluj-Napoca, Romania Cristian Mihaescu, University of Craiova, Romania Delia Mitrea, Technical University Cluj-Napoca, Romania Florica Moldoveanu, University Politehnică of Bucharest, Romania Gabriel Neagu, ICI Bucharest, Romania Mihaela Ordean, iQuest Technologies Cluj-Napoca Marian Pădure, Babeş-Bolyai University, Cluj-Napoca, Romania Camelia Mihaela Popa, IFPAR, Bucharest, Romania Elvira Popescu, University of Craiova, Romania Dan Rotar, University of Bacau, Romania Corina Sas, University of Lancaster, United Kingdom Emil Stănescu, ICI Bucharest, Romania Serban Ungureanu, Alexandru Ioan Cuza University, Iasi, Romania

Conference organized with the support of:



Alexandru Ioan Cuza University of Iasi, Romania



Université catholique de Louvain, Belgium



Louvain School of Management, Belgium



International Business Machines – IBM Belgium



RoCHI – ACM SIGCHI România

This volume was published with the support of the:



National Authority for Scientific Research and Innovation (ANCSI)

Foreword

In 2016, the International Conference on Human-Computer – RoCHI has reached its 13^{th} edition. Like all the previous editions, RoCHI 2016 provides an inter-disciplinary forum for the exchange of ideas, expertise and research results in the field of human-computer interaction.

The conference is publishing within this proceedings an invited paper and 20 regular papers organized in five thematic sessions. These papers were selected from a total of 29 submissions that were sent to RoCHI 2016. After a careful review process, each submission being assigned to a minimum of three and maximum of six reviewers in a peer-reviewed fashion (double blind), a total number of 20 short and long papers were accepted, which represents an acceptance rate of 68%.

The invited paper was entitled "The design is right and the right design" – How to measure and theorize "right" in HCI and Software Engineering? It was presented by Professor Ahmed Seffah from Lappeenranta University of Technology, Finland. It is about bridging the gap between classical design for usable user interfaces and sustainable design so that it could be measured.

The five sessions were entitled: Usability and accessibility evaluation, User interface development, Interaction techniques, HCI in e-learning systems, and Interactive applications.

The first session in the proceedings and in the conference program is *Usability and accessibility evaluation*. The session contains papers on topics that are at the core of HCI research, such as usability studies, automated evaluation, reliability and models for software design and GUI. The papers from this section present: (1) how can assistive technologies help people with disabilities (Implementation of a polyglot text-to-speech synthesis in two assistive technologies – *Paul Fogarassy-Neszly, Aurel Patru, Dragos Daniel Iordache and Costin Pribeanu*), (2) how was UMUX and UMUX-LITE scales been tested in the context of Facebook use by university students (Comments on the reliability and validity of UMUX and UMUX-LITE short scales – *Costin Pribeanu*), (3) how it is built ERGOSIM, a software that automatically evaluate the design of menu bars, pull-down menus, and sub-menus of a graphical user interface by reviewing usability guidelines related to menu design (Automated Evaluation of Menu by Guidelines Review – *Sara Bouzit, Gaëlle Calvary, Denis Chêne and Jean Vanderdonckt*) and (4) what are the results of an experiment in which students are asked to reposition rectangular shapes in a window in two cases: when they are or they are not told that the window should be considered the configuration of a graphical user interface (GUI) (Perceived aesthetics of user-modifiable layouts: a comparison between an unspecified design and a GUI – *Stefan Trausan-Matu, Brahma Dathan*).

The second section, *User interface development*, contains four papers related to user interfaces: (1) UIDLC Manager, a software that provides user interface designers and developers with methodological guidance throughout user interface development life cycle (Enactment of User Interface Development Methods in Software Life Cycles – *Iyad Khaddam*, *Hanaa Barakat, Jean Vanderdonckt*), (2) an Android application which offers the possibility to navigate through a 3D model using the movement of a mobile phone, determined by its own sensors (Increasing the Accuracy of Indoor Localization Applications by Using Predefined Markers and the Phone's Camera – *Szabolcs Orban, Teodor Stefanut*), (3) a complex system (hardware and software) that is used to monitor and control some ambient parameters (temperature, humidity, light intensity, etc.) (Using WSN and Mobile Apps for Home and Office Ambient Monitoring and Control – *Catalin Damian, Lenuta Alboaie and Adrian Iftene*) and (4) SMAUG is a game based on Sphero 2.0 robotic ball, where multiple players share the same driving control of the robot (Sphero – Multiplayer Augmented Game (SMAUG) – *Marian-Nicolae Pinzariu, Adrian Iftene*).

The third section, *Interaction Techniques*, addresses last technologies used in new types of interaction: (1) a method with aim to improve interaction in virtual environments and make it less stressful and more effortless (Reducing Gestural Interaction Physical Stress in Virtual Environments: an Experiment – *Sobhi Ahmed, Laure Leroy and Ari Bouaniche*), (2) a study with aim to present Oculus Rift, used to assess craving of Romanian smokers (Oculus Rift 3D Interaction and Nicotine Craving: Results from a Pilot Study – *Ioana-Monica Ciolan, Sabin Buraga and Ion Dafinoiu*), (3) the design and development of a complete hardware and software solution for a brain computer interface (Brain Computer Interface using Machine Learning – *Cristian-Valeriu Soare*) and (4) a methodology for identifying and assessing a set of performance issues encountered in a particular web application, with impact on the usability level (Visual techniques for identification and evaluation of a web application usability issues – *Mihaela Ciugudean, Dorian Gorgan*).

Forth section, *HCI in e-learning system*, contains papers on novel and popular topics in HCI. The papers are presenting: (1) a case study in testing technology acceptance model to explain the adoption of Facebook by Romanian university students (Testing the technology acceptance model with Romanian university students – *Alexandru Balog, Costin Pribeanu*), (2) a case study about the relationship between the Facebook dependence and the negative consequences of the excessive use of Facebook on the students' university work (Measuring the negative effects of the Facebook dependence on the students' university work – *Gorghiu Gabriel, Manea Valentina Iuliana, Iordache Dragoş Daniel and Pribeanu Costin*), (3) the methodology and results obtained after the evaluation of the Tesys e-learning platform's interface – *Paul Stefan Popescu, Cristian Mihaescu, Mihai Mocanu and Costel Ionascu*) and

(4) MOOCBuddy, an innovative project aiming to make known the Romanian initiatives related to MOOCs (MOOCBuddy: a Chatbot for personalized learning with MOOCs – *Carmen Holotescu*).

In the last section, *Interactive applications*, are four papers covering the following themes: (1) *ReaderBench* framework, which includes multi-lingual comprehension-centered web services designed to address a wide range of individual and collaborative learning scenarios (*ReaderBench* goes Online: A Comprehension-Centered Framework for Educational Purposes – *Marius-Gabriel Gutu, Mihai Dascalu, Stefan Trausan-Matu, Philippe Dessus*), (2) how the tweets can be processed in order to obtain valuable information in real time, based on user preferences and different search criteria (Event detection in Tweets – *Andrei Bogdan Baran, Adrian Iftene*), (3) a study about rhythm in various genres of texts (Rhythm analysis of texts using Natural Language Processing – *Irina-Diana Niculescu and Stefan Trausan-Matu*) and (4) an approach to terrain synthesis from minimal-detail user-provided heightmaps (Terrain Synthesis from Crude Heightmaps – *Alexandre Philippe Mangra, Adrian Sabou, Dorian Gorgan*).

We cannot end this preface without expressing our appreciation to the members of the scientific committee and to the volunteer reviewers from the RoCHI group who helped for selecting the best papers to be presented at the conference. Moreover, we acknowledge the efforts of all the persons involved in the organization of the RoCHI 2016 Conference and thank them for their efforts!

Iasi, August 8th, 2016

Editors, Adrian Iftene Jean Vanderdonckt

"The design is right and the right design" How to measure and theorize "right" in HCI and Software Engineering?

Ahmed Seffah Human-centric Software Engineering Ahmed.Seffah@lut.fi

ABSTRACT

In this keynote, we will show how to improve the practices of interactive software and user interfaces design, by applying lessons from various fields of design to the co-creation of interactive software products, systems and services. The goal is to create software that works – really works, meaning we can measure it works, usable, profitable yet secure, resilient and sustainable – A software is being appropriate and effective for people who live in the world that the software creates. Beyond the "cook-books of the HCI gurus" and the large diversity of HCI design patterns, guidelines and principles, the talk will highlight the the importance of the qualities, properties and quantities that quantify the quality of software systems from the human perspective. How to measure the "right", "the old user friendly", the past usability, the today quality of user experience and the tomorrow sustainability. How to ground measures of HCI in the general theory of software quality and measurement. What we can learn from the successful stories and history of measurement in many fields: medical sciences, engineering, and even from natural and hearth sciences?

ABOUT

Ahmed Seffah is a professor of human-centric software engineering and HCI at Lappeenranta University of Technology, Finland. Previously, he was a professor and Concordia research chair on human-centered software engineering at Concordia University as well a visiting professor in more than 10 universities and research Centre including IBM, University of Lausanne, Daimler Chrysler and the Computer research institute of Montreal. Professor Seffah co-authored five books, the last one on the "Patterns of HCI Design Patterns and the HCI Design of Patterns". His main research is to understanding those HCI and interactive software systems design, software usability, user-centric engineering, UX design practices and all similar ones within the wider software and systems engineering processes. Most visible contributions of his work are on the gaps and bridges between design science practices and software engineering methodologies such as agile, model-driven and service-oriented while building a theories of human-centric software engineering.

Implementation of a polyglot text-to-speech synthesis in two assistive technologies

Paul Fogarassy-Neszly, Aurel Patru

Dragos Daniel Iordache, Costin Pribeanu

BAUM Engineering

National Institute for R & D in Informatics - ICI Bucharest Bd. Averescu 8-10, Bucucharest, Romania

Str. Traian Mosoiu 8, Arad, Romania pf@baum.ro, ap@baum.ro

iordache@ici.ro, pribeanu@ici.ro

ABSTRACT

The text-to-speech (TTS) synthesis is widely used in the area of assistive technologies for visually impaired people as well as for people with dyslexia or related learning disabilities. In the case of multilingual text-to-speech synthesis, both the language identification and voice switching are desirable. This paper presents the implementation of a multilingual text-to-speech in two assistive technologies: the automatic reading machine POET and the portable organizer for blind people Pronto. Both implementations have been tested with visually impaired users. The evaluation results show that the implementation of this software component makes these devices more easy and comfortable to use.

Author Keywords

Accessibility, assistive technologies, usability, technology acceptance, TTS, visually impaired users.

ACM Classification Keywords

D.2.2: Design tools and techniques. H5.2 User interfaces.

INTRODUCTION

A Text-To-Speech (TTS) system is a software component able to produce speech output, whether directly from text input or from a scanned document page previously submitted to an Optical Character Recognition (OCR) system. Many assistive technologies for visually impaired people are using text-to-speech (TTS). TTS synthesis is also used by people with reading disabilities (dyslectic, illiterate, or with learning disability) in order to make the electronic or printed documents accessible. Examples of assistive technologies using TTS synthesis are: screen readers, automatic reading machines, portable computers with voice interface, smartphones, GPS systems, as well as many other gadgets or self-voiced software.

Usually, the synthetic voice corresponding to the native (or preferred) language of the user is currently selected. If the text is written in another language, then the user has to manually select a corresponding voice for that language. In recent years, there is an increasing interest in the applications that are able to process texts written in two or more languages. There are many application areas that need polyglot text-to-speech, such as education for all and multi-cultural contexts [8, 11, 12, 13]. In this case, both a multilingual (polyglot) text-to-speech synthesis and voice switching are needed. This requires to analyze the text in order to detect the language and then to select the voice available for that language. Several approaches for the multilingual TTS exist that differ with regard to the solutions adopted for the text analysis and speech synthesis [1, 3, 9, 10, 11].

In a previous work, a software component for multilingual text-to-speech has been presented that performs both the automatic language identification and voice switching. The component has been developed during the research project iT2V that has been carried on in a consortium of three partners: BAUM Engineering, ETA Automatizari Industriale, and National Institute for Research and Development in Informatics - ICI Bucharest.

The language recognition component plays the role of an intermediate layer, voice independent, between the application and the synthesis process. Language identification is based on statistical analysis and trigrams frequency evaluation for envisaged languages [2]. The development followed four steps: alpha version (proof-ofconcept, functional version (beta), commercial version, and implementation in several applications. The development cycle of iT2V is illustrated in Figure 1.



Figure 1. iT2V development cycle.

The goal of the alpha version was to test the language detection algorithms [5]. In the second step, a preliminary functional version (beta01) was developed and tested with four candidate languages [7]. The evaluation revealed a major usability issue: switching the voice in the middle of a sentence. Therefore, is an improved functional version has been developed and tested again. Based on the testing results [6], the commercial version has been developed and then used to implement the software component. The iT2V component plays the role of a special voice (SAPIiT2V) that identifies the language and selects the corresponding synthetic voice.

This paper presents the implementation of the multilingual text-to-speech in two assistive technologies: the automatic reading machine POET and Pronto, the portable electronic organizer for blind and visually impaired users. The implementations have been evaluated for usability and technology acceptance with visually impaired users. The rest of this paper is organized as follows. Next section briefly presents the automatic reading machine POET and the portable organizer for blind people Pronto. Then the evaluation method is presented. In the next section the evaluation results of the implementations in these devices are presented and discussed. The paper ends with conclusion and future research directions.

IMPLEMENTATIONS OF IT2V

Implementation in a reading machine

The first implementation has been done on the automatic reading machine POET (trade mark of BAUM Retec AG).

This type of assistive technology integrates a computer and a scanner (including an Optical Character Recognition software - OCR - with italics adjustment) in one single device. The device is simple and is easy to use. The basic version of POET (Compact 2) has only two buttons (Start and Stop) and two control knobs to adjust the volume and the reading speed. Figure 2 presents the reading device POET Compact 2. The user can read periodicals, books, or magazines. The text could be written on one or several columns.



Figure 2. Automatic reading device POET Compact 2.

The text could be saved and then used for a new lecture or transferred to another device, such as an MP3 player. The model Compact 2+ has three buttons and a menu keypad and has been designed for users that prefer is a model for more advanced users who prefer more features and full control over the functionality. For Braille literate and deaf and blind users a Braille display can be connected to the Poet. As Braille tables are language dependent, prior language detection is necessary before producing Braille output. This is a not foreseen application of language detection.

Implementation in a portable device

The second implementation has been done on the organizer for blind people Pronto (trade mark of BAUM Retec AG).

Pronto is a portable organizer for visually impaired people that runs under the operating system Window CE. The device includes a Braille display and speech output. Pronto supports various applications, like a regular PDA (personal digital assistant). As such is more complex than other assistive technologies. Figure 3 presents the organizer Pronto. The main functions are: taking notes with a text editor, agenda & organizer, playing music, reading books in text or Braille format, Internet access.

The device is easy to handle and easy to use. It provides a keyboard for 8-dot Braille input, an 18-cell Braille display with integrated cursor routing, four function keys for fast access to applications, and a navistick for an easy and comfortable operation.



Figure 3. Organizer for blind people Pronto.

METHOD

Evaluation techniques

The evaluation has two goals: usability and acceptance. First, a usability inspection has been carried on by three experts. Then, each implementation has been tested with visually impaired users for usability and acceptance. After testing the device, participants were asked to answer a questionnaire as regards the actors that are influencing the technology acceptance. The users were also asked to mention the most important positive and negative aspects as regards each device and to weight the relative importance of three factors (ease of use, usefulness, and enjoyment).

Measures

Following measures were collected: number of usability problems by severity degree (major, moderate, and minor), description of unique usability problems, number of positive and negative aspects, the ratings of each item of the questionnaire, and the relative importance of each factor. The items in the questionnaire are presented in Table 1.

Item	Statement			
PEU1	It is easy to understand how to use iT2V with this device			
PEU2	This device is easier to use with iT2V			
PU1	iT2V is a useful facility for me			
PU2	The capabilities of iT2V correspond to my requirements			
PE1	I prefer to use this device having iT2V			
SAT1	I am satisfied with this device having iT2V			
INT1	If would like to use this device with iT2V			
INT2	I will recommend other people this device having iT2V			

Table 1. Questionnaire.

The questions are inspired from the technology acceptance model [4]. Since all the users were visually impaired, the answers have been collected via interview (ratings of each item on a 5-point Likert scale).

Participants and tasks

Seven people participated in user testing (six men and a woman). The mean age of participants is 40.3 years (SD=9.23) with a minimum of 23 and a maximum of 48 years. The user testing took place in Arad, at the Local Branch of Romanian Association of the Blind (Filiala Arad a Asociației Nevăzătorilor din România). Except for one user (university student), all are retired for medical reasons. All participants graduated a high school. The disability degree is severe (first degree - legally blind).

The familiarity with the assistive technologies is variable. However, all are familiar with the lecture on the Internet. The most used information and communication technologies are the computer, mobile phone, tablet, and scanner + OCR. Most used applications are the Internet browsers, Skype, and Facebook. The usual goals are related to information, lecture, entertainment (games), and socialization. The POET device has been tested with three tasks.

The goal of the first task was to read two documents, each of them written in two languages (Romanian and English), without iT2V. This means to stop the lecture and manually change the voices from the device option menu. In the first document the text was written in one column. In the second document, the same text was written in two columns.

The second task was identical, but with the iT2V feature selected. The third task was to read several pages from a magazine with iT2V feature selected. The Pronto device has been tested with a task performed under two conditions. The task was to open and read a document written in two languages (Romanian and English). The condition is to have the iT2V selected.

RESULTS

Reading machine POET

The usability inspection identified several usability problems for the first task. The problems are related to the navigation in the menu in order to manually change the voice. No usability problem has been detected at the second and third tasks. The accuracy of language identification was 100%. The average time saving for a voice switching was 2 minutes. The user testing confirmed the results of the usability inspection: no usability problem has been detected after selecting the iT2V feature.

Item	Factor	Mean	SD
PEU1	Ease of use	4.71	0.49
PEU2	Ease of use	4.43	0.79
PU1	Usefulness	4.29	1.50
PU2	Usefulness	4.29	1.50
PE1	Enjoyment	4.57	0.79
SAT1	Satisfaction	5.00	0.00
INT1	Intention to use	5.00	0.00
INT2	Intention to use	4.71	0.76

Table 2. Descriptives for the POET device.

Most of the users mentioned that the device is accessible, simple, easy to use, useful, and compact. They appreciated the ease of reading and the multilingual feature. As regards the negative aspects, the users mentioned the price (too high), the dimensions (too big) and the lack of the translation.

The users considered that the most important factor for the acceptance of the reading device is the usefulness (45%), followed by the ease of use (29%), and usefulness (26%). The descriptive statistics for the items in the questionnaire is given in Table 2. The mean value of items related to the ease of use is 4.57 (SD = 0.45) and of the items related to usefulness 4.29 (SD = 1.47). An analysis of correlation

based on Spearman coefficient shows a marginally significant relationship between the enjoyment and the satisfaction (rho = 0.68, p = 0.091).

The organizer for blind people Pronto

The usability inspection identified few usability problems when performing the task without iT2V. The problems are related to the navigation in the menu in order to manually change the voice. The problems are related to the device itself and somehow inherent, given the rich functionality and the lack of familiarity with the device. The user testing confirmed the results of the usability inspection: no usability problem has been detected after selecting the iT2V feature.

Overall, the usability evaluation results were similar for the two implementations. Most frequently mentioned positive aspects were the portability, the Braille feature, flexibility of use (both voice and Braille), and the multilingual feature. As regards the negative aspects, they mentioned the price and the difficulty to use.

The users considered that the most important factor for the acceptance of the reading device is the ease of use (44%), followed by the usefulness (36%), and enjoyment (20%). The descriptive statistics for the items in the questionnaire is given in Table 2.

Item	Factor	Mean	SD
PEU1	Ease of use	3.71	0.76
PEU2	Ease of use	3.86	0.90
PU1	Usefulness	4.43	1.13
PU2	Usefulness	4.43	0.53
PE1	Enjoyment	4.43	1.13
SAT1	Satisfaction	4.57	0.53
INT1	Intention to use	5.00	0.00
INT2	Intention to use	4.86	0.38

Table 3. Descriptives for the Prontop device

The mean value of items related to the ease of use is 3.79 (SD = 0.39) and of the items related to usefulness 4.43 (SD = 0.53). The analysis of correlation showed a marginally significant relationship between the perceived enjoyment and the satisfaction (rho = 0.72, p = 0.068) and a significant relationship between the perceived enjoyment and the perceived usefulness (rho = 0.84, p = 0.018).

Discussion

The evaluation has been focused on the implementation of the iT2V component and not on the device itself. In this respect, no usability problem has been identified. In both cases, the accuracy of voice switching was excellent.

The component saves time and make the use of these assistive technologies much simpler. The answers to the questionnaires show a difference between the ease of use of the two devices. Pronto has a more rich functionality which makes it more difficult to use. Nevertheless, in both cases the users have been satisfied and expressed the intention to use these technologies.

There are several limitations of this work. First of all, the number of users is small, so the results could be seen as

only exploratory. Second, the users were not familiar with these devices. A reason is the price: six out of seven users said that they can't afford these technologies. Third, the time for evaluation was limited, so only few tasks were assigned. The reasons were the availability of users and the time needed to explain how to use a new device.

CONCLUSION AND FUTURE WORK

In this paper two implementations of the iT2V software component for multilingual text-to-speech and automatic voice switching have been presented.

The evaluation results showed that iT2V is usable, useful and enjoyable. A device with iT2V is easier to use. Since the users are visually impaired, manually changing the voice could be done via an audio menu, which is both difficult to use and time-consuming.

In the next future the language recognition will be implemented in COBRA, the screen reader software. This will be very useful, especially for internet browsing, when beside pages in local language, pages in English should be accessed. Also, in countries with two or more official languages this facility will be useful, as well as for any polyglot user who often are switching between documents in different languages.

Also, beside the language dependent speech synthesis switch, the Braille table will be selected accordingly, taking into account that not only synthetic voices, but also Braille tables are language dependent. This will allow the user to access documents in different languages, both through speech synthesis and Braille display, without any option menu exploration.

ACKNOWLEDGMENTS

This work is partly supported by the IT2V research project (29DPST/2013), financed by UEFISCDI under the PNCDI II Innovation Program.

REFERENCES

- Bourlard, H., Dines, J., Magimai-Doss, M., Garner, P. N., Imseng, D., Motlicek, P. and Valente, F. Current trends in multilingual speech processing, *Sadhana* 36(5), (2011), 885-915.
- 2. Cavnar, W. and Trenkle, J. N-gram-based text categorization. *Proceedings Document Analysis and Information Retrieval* (SDAIR-94), (1994).
- 3. Chen, C. P., Huang, Y. C., Wu, C. H. and Lee K. D. Polyglot speech synthesis based on cross-lingual frame

selection using auditory and articulatory features. *IEEE/ACM TASLP* 22 (10), (2014), 1558-1570.

- 4. Davis, F.D., Bagozzi, R.P. and Warshaw, P.R. Extrinsic and intrinsic motivation to use computers in the workplace. *Journal of Applied Social Psychology*, 22 (14), (1992), 1111-1132.
- Fogarassy-Neszly, P. and Gherhes, V. Applications for dynamic language identification. *Proceedings of RoCHI 2014* Conference, Popovici D., M. & Iordache D.D. (Eds.), (2014), 51-54.
- Fogarassy-Neszly, P., Zinveliu, Z. and Pribeanu, C. A software component for polyglot text-to-speech synthesis: user interface and beta testing results. Proceedings of RoCHI 2015 Conference, Dardala, M., Rebedea, T.E. (Eds.), (2015), 145-148.
- 7. Pribeanu, C. and Fogarassy-Neszly, P. Beta testing of a dynamic language identification software component preliminary results. *Revista Romana de Interactiune Om-Calculator* 7(3), (2014), 259-272.
- Ramani, B., Actlin Jeeva, M.P., Vijayalaksmi, P., and Nagarajan, T. Cross-lingual voice conversion-based polyglot speech synthesizer for Indian languages. *Proceedings INTERSPEECH*, (2014), 775-779.
- 9. Romsdorfer, H. and Pfister, B. Text analysis and language identification for polyglot text-to-speech synthesis. *Speech Communication* 49, (2007), 697-724.
- 10. Shiga, Y. and Kawai, H. Multilingual speech synthesis system. *Journal of the National Institute of information and Communication Technology* 59(3/4), (2012), 21-28.
- 11. Traber, C., Huber, K., Nedir, K., Pfister, B., Keller, E. and Zellner, B. From multilingual to polyglot speech synthesis. Proceedings of EUROSPEECH, Budapest, Hungary, (1999), 835–838.
- 12. Tripathi, M., and Shukla, A. Use of assistive technologies in academic libraries: A survey. Assistive Technology, 26(2), (2014), 105-118.
- 13. Turunen, M. and Hakulinen, J. Mailman-a multilingual speech-only e-mail client based on an adaptive speech application framework. In *Proceedings of Workshop on Multilingual Speech Communication* MSC 2000, (2000), 7-12.
- 14. Udvari-Solner, A. and Thousand, J. S. Creating a responsive curriculum for inclusive schools. Remedial and special education, 17(3), (1996), 182-191.

Comments on the reliability and validity of UMUX and UMUX-LITE short scales

Costin Pribeanu

National Institute for Research and Development in Informatics - ICI Bucharest Blvd. Maresal Averescu, nr.8-10, Bucharest, Romania

pribeanu@ici.ro

ABSTRACT

Recent work on usability and user experience shows several concerns on the validity of evaluation instruments. There is a debate on the use of standardized scales versus short scales, such as UMUX and UMUX-LITE or even a single-item measure of usability. Nevertheless, there are relatively few papers reporting the testing of these scales together for reliability and validity. This paper aims at discussing the UMUX and UMUX-LITE scales that have been tested in the context of Facebook use by university students. From a theoretical point of view, both scales are questionable. From an empirical point of view, the testing results confirmed a lack of unidimensionality as well as a poor reliability and convergent validity of these scales.

Author Keywords

Usability scales, UMUX, UMUX-LITE, factor analysis, validity, Facebook use.

INTRODUCTION

Recent work on usability and user experience shows several concerns on the reliability and validity of the scales used to measure the perceived usability. There is currently a debate on the use of standardized scales versus short scales, such as UMUX [6] and UMUX-LITE [14] or even a single-item measure of usability [13]. There are many pros and cons as regards the reliability and validity as well as the practical benefits. Nevertheless, there are relatively few papers reporting the testing of these scales for reliability and validity.

This paper aims at discussing the UMUX and UMUX-LITE scales from both a theoretical and an empirical point of view. Additionally, the perceived ease of use (PEU) is analyzed that is a widely used concept in the context of technology acceptance [5]. Since PEU is a short scale taping on several key usability aspects, it could be a better alternative than UMUX and UMUX-LITE.

Theoretically, the analysis is following the scale development recommendations. Empirically, the analysis is focused on the scale testing that has been carried on by using two samples collected during a larger study on Facebook use by university students [9]. The first sample is from the pilot study and is used to assess the UMUX and UMUX-LITE scales. The second sample is from a subsequent study using a revised evaluation instrument and is used to assess the UMUX-LITE and PEU scales.

The rest of this paper is organized as follows. The following section briefly presents recent approaches in the area of scale development with an emphasis on the scale development process and the existing usability scales. In section 3, the analysis of UMUX and UMUX-LITE is presented based on two empirical studies. The same assessment criteria are used to analyze the PEOU scale. The paper ends with conclusion and future research directions.

RELATED WORK

Scale development

The interest in developing rating scales increased after the adoption of the ISO 9241-11 standard that included satisfaction as a key usability aspect. As Lindgaard & Kirakowski [15] pointed out, the landscape of scale development in HCI shows many usability scales, many approaches, as well as many opinions as regards the scale reliability and validity.

When analyzing the reliability and validity of the usability scales, two aspects are usually neglected: the theoretical meaning and the multidimensional nature of the usability concept. Psychometrics is not a favorite area of expertise in HCI [15]. As such, the scale development process is not well understood as regards both the ordering of steps to be carried on and the suitable techniques that should be used in each step.

Long time ago, Gerbing & Anderson [7] outlined an updated paradigm for scale development that includes a confirmatory factor analysis (CFA) to assess the scale unidimensionality. They underlined that, only after achieving an acceptable unidimensionality level, the reliability could be assessed. This precondition is usually ignored in the existing papers reporting the development and testing of usability scales. The authors rely on the traditional approach that only includes the Cronbach's alpha coefficient, the item-to-total correlations, and the exploratory factor analysis (EFA).

More recently, MacKenzie et al. [16] emphasized the importance of the conceptualization as a first step in the scale development process. They noticed that an adequate conceptualization is difficult and requires a review of the literature on the meaning of related constructs, aspects these constructs refer to, dimensionality and preliminary research with domain experts or practitioners. Another important issue is to specify if the construct is measured reflectively or formatively.

Usability in the ISO standards

ISO 9241-11 standard defined usability as the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use [11]. The ISO standard 9126-1 defined usability as the capability of a software product to be understood, learned, used, and attractive to the user, when used under specified conditions [10]. Later on, both definitions were integrated in the ISO standard 25010 [12], under two key terms:

- Quality in use: the degree to which a product used by specific users meets their needs to achieve specific goals with effectiveness, efficiency, safety, and satisfaction in specific contexts of use.
- Usability: the degree to which a software product is able to satisfy the following needs when used under specified conditions: appropriateness recognizability, learnability, operability, error protection, user interface aesthetics, and accessibility.

As pointed out by Bevan et al. [1], the quality in use defines usability as a high level concept, focusing on the outcomes of the interaction rather than on the characteristics that make a product usable. Unfortunately, this distinction is rarely made in the mainstream of HCI literature.

Measurement scales for the perceived usability

A well-known usability scale is System Usability Scale (SUS) that has been developed by Brooke as a simple, "quick and dirty" scale [3]. SUS has been widely used and is considered an industrial standard [2, 14].

Several authors noticed that SUS lacks unidimensionality. For example, Borsci et al. [2] found that a learnability dimension of SUS might emerge under certain conditions (when administrated to experienced users). SUS has been also criticized for using both positive and negative wording, since this may lead to mistakes (made by respondents) and mis-coding (made by researchers) [17].

More recently, Finstad proposed the UMUX (Usability Metrics for User Experience) as a shorter alternative to SUS. UMUX have been criticized for dimensionality [4, 14] and for using negative wording [14, 17]. Lewis et al. [14] found that UMUX has a bi-factorial structure with positive tones aligning with one factor and negative tones with the second factor.

An even shorter scale that is based on UMUX has been proposed by Lewis et al. [14]. UMUX-LITE was intended as a very quick, two-item scale, that uses the first and the third item from UMUX. The authors found that UMUX-LITE is unidimensional and has acceptable reliability. However, they recommended using this scale with caution until it will be validated across a wider variety of systems.

A well-known scale measuring the perceived ease of use (PEU) has been developed and tested in the context of technology acceptance studies. The technology acceptance model (TAM) has been developed by Davis et al. [5], in order to explain and predict the technology acceptance on a large variety of technologies. Although it is a short scale with a widely recognized psychometric quality, PEU has been rarely used in the HCI studies.

ANALYSIS OF UMUX AND UMUX-LITE

Method

The analysis follows the recommendations in the literature for scale development and assessment of dimensionality, reliability, and validity [7, 8, 16]. The first step is to analyze the conceptualization based on the definition of concepts in the literature. Then, the dimensionality is assessed via exploratory and confirmatory factor analysis. After demonstrating that the construct is unidimensional, the reliability could be analyzed checking the magnitude of Cronbach's alpha and the item-to-total correlations. The convergent validity is assessed by the examination of the composite reliability (CR), and average variance extracted (AVE).

Empirical studies

UMUX and UMUX-LITE have been tested in a larger study on the use of Facebook (FB) by university students. Two samples collected during these studies are used for the analysis of the psychometric quality of UMUX and UMUX-LITE. The respondents were asked to answer questions related to demographics, enrollment, FB usage (the size of their FB network, frequency of use, minutes per day), and to evaluate items on a 7-point Likert scale.

The first sample has been collected in 2014 and consists of 152 students (110 female, 42 male) from two universities in Lithuania. The negatively worded items in Table 1 were recoded. The first and third items in Table 1 represent the UMUX-LITE scale.

Item	Statement	Μ	SD
U1	FB's capabilities meet my requirements	4.02	1.38
U2	Using FB is a frustrating experience	3.44	1.64
U3	FB is easy to use	4.91	1.15
U4	I have to spend too much time correcting things with FB	4.12	1.63

Table 1. Descriptives for UMUX scale (N=152).

The second sample has been collected in 2015 and consists of 414 students (258 female, 156 male) from a Romanian university. Since the testing results from the first study revealed poor psychometric properties of UMUX, the scale has been removed from the evaluation instrument and replaced with PEU. However, the first item has been preserved in order to test again UMUX-LITE.

Item	Statement	М	SD
U1	FB's capabilities meet my requirements	4.36	1.51
PEU1	It is easy to learn how to use FB	6.10	1.27
PEU2 / U3	FB is easy to use	6.21	1.17
PEU3	My interaction with FB is clear and understandable	5.69	1.38

Table 2. Descriptives for UMUX-LITE and PEU (N=414).

The PEU scale has been developed by adapting items from the existing scales. The three items tap on three usability attributes: understandability, learnability, and operability. As it could be observed, the item U3 in Table 1 is identical with the item PEU2 in Table 2.

Conceptualization

In most studies discussing UMUX and UMUX-LITE the reliability and validity is limited at dimensionality which is assessed with exploratory factor analysis techniques, reliability, and correlation with other usability scales. The main shortcoming of UMUX (and, consequently, of UMUX-LITE) is the poor conceptualization which is due, on one side, to the underlying usability definition and, on another side, to the misunderstanding of the nature of the measurement model.

Both scales take the roots from the quality in use concept and try to measure the user's subjective satisfaction. Nevertheless, the measured variable is not the satisfaction (there is no item saying "I am satisfied with...". Rather, the operationalization of the construct is done with a mix of items measuring various aspects of user experience and usability. As such, the conceptualization is ambiguous and is not clear what UMUX is actually measuring.

The first item is the most ambiguous and undermines the conceptualization of both UMUX and UMUX-LITE. The fit between the user's needs and requirements could refer to anything: ease of use, aesthetics, flexibility, robustness, safety, usefulness, enjoyment, etc.

The second shortcoming is the lack of a clear definition of the nature of the measurement model. The measurement model describes the relationship between a construct and its measures [7, 16]. According to the direction of the causal relationship, the constructs could be reflective (from the construct to its measures) or formative (from measures to construct). It is also possible to define multidimensional constructs where the dimensions are specified as first order constructs.

Failure to adequately specify the measurement model leads to a poor operationalization and a lack of validity. The point is that the conceptualization and validation recommendations are different for reflective and formative measurement models. Unidimensionality and inter-item correlation are required for reflective constructs, since all items are supposed to measure the same thing. For a more detailed discussion on the scale development of reflective and formative constructs, see MacKenzie et al. [16].

Therefore, neither UMUX nor UMUX-LITE could be adequately assessed as measurement scales, since both of them suffer from a lack of clear definition of the construct domain. As regards the measurement model, although is not specified, it is assumed to be reflective according to the assessment techniques used by the authors.

Dimensionality

The dimensionality of UMUX and UMUX-LITE has been analyzed within the two empirical studies. The first study enabled testing of UMUX and UMUX-LITE. The principal component analysis with Varimax rotation for UMUX resulted in two factors explaining 34.56%, respectively 26.69% of the variance. The same analysis for UMUX-LITE resulted in one factor explaining 60.06 % of the variance.

A confirmatory factor analysis has been then carried on. The results revealed the lack of dimensionality for both constructs. The loadings of the underlying construct on its measures (α regression coefficients) is below the cutoff value of 0.60 [8]. The results are presented in Table 3.

.	UMUX			UMUX-LITE		
Item	EFA			EFA		
	1	2	CFA (α)	1	CFA (α)	
U1	.707	364	0.66	.775	0.65	
U2	.700	.332	-0.38			
U3	.609		0.31	.775	0.34	
U4		.919	0.32			

Table 3. Dimensionality of UMUX and UMUX-LITE (N=152).

The second study enabled the analysis of dimensionality of UMUX-LITE and PEU. The principal component analysis with Varimax rotation for UMUX-LITE resulted in one factor explaining 68% of the variance. The same analysis for PEU resulted in one factor explaining 80.61% of the variance.

The confirmatory factor analysis for UMUX-LITE revealed a low item loading of U1 (α =0.49). The same analysis for PEU confirmed its unidimensionality (item loadings: 0.86, 0.92, and 0.73).

The results of the two empirical studies demonstrate the lack of dimensionality for the UMUX and UMUX-LITE short scales as well as the limitations of the exploratory factor analysis for testing the dimensionality.

Reliability

The Cronbach's alpha was unacceptable low in the first study: 0.213 for UMUX and 0.331 for UMUX-LITE. The item-to-total correlations were in the range of 0.05-0.23for UMUX, respectively 0.20 for UMUX-LITE.

In the second study, the Cronbach's alpha for UMUX-LITE was low (0.517) and the item-to-total correlation also low (0.36). Cronbach's alpha for PEU was 0.874 and the item-to-total correlation in the range of 0.68-0.82.

Convergent validity

Convergent validity refers to the degree to which the measures of a construct that are supposed to be related, are in fact related.

In the first study, the low item loadings make no sense to test the convergent validity of UMUX. For UMUX-LITE, the composite reliability of 0.412 and the average variance extracted of 0.279 demonstrate the lack of convergent validity. The second study confirmed the poor convergent validity of UMUX-LITE (CR=0.552, AVE=0.390). The

convergent validity for PEU was very good (CR=0.885, AVE=0.722).

Interpretation of scores

The final step in scale development is to provide the prospective researchers with some recommendations for the interpretation of scores. Since this step is beyond the purpose of this study, it will not be discussed. However, is important to note that the correlation of a scale under consideration with other existing scales does not ensure the scale validity. It is expected that two scales pointing to similar usability aspects correlate. The problem is that if the scale under consideration and the reference scale are not unidimensional, then a comparison leads to ambiguous if not erroneous conclusions.

DISCUSSION AND CONCLUSION

In the area of HCI, several misconceptions exist as regards the scale development and the validity criteria. The two short scales analyzed in this paper suffer from an ambiguous definition of the construct. An adequate conceptualization should include the specification of the dimensionality and the nature of the construct (reflectively vs. formatively measured).

It seems that the relationship between the definition of the target construct and the criteria for its assessment are not well understood: if a usability scale is not unidimensional, then more than one thing (e.g. usability) is measured. In other words, what is actually measured is not what has been supposed to be measured.

A problem with the conceptualization of existing short scales in HCI, such as SUS, UMUX, and UMUX-LITE is the confusion between usability and quality in use. The quality in use is a multidimensional construct since it taps on different concepts. Another problem is the overlapping between two HCI concepts: usability and user experience.

It is advisable to keep apart the scales measuring the pragmatic and hedonic aspects (each scale should undergo a separate validation procedure). For example, the perceived ease of use refers to pragmatic aspects, while the perceived enjoyment refers to hedonic aspects. Both scales have been widely used and validated in technology acceptance studies.

The empirical studies confirmed the recommendation of Gerbing and Anderson to use confirmatory factor analysis to assess the dimensionality [7]. The exploratory factor analysis is clearly not enough.

The two empirical studies show that both UMUX and its shorter version, UMUX-LITE, suffer from poor reliability, lack of unidimensionality, and poor convergent validity. The correlation with other usability scales, which is frequently mentioned as an argument for reliability, is a poor surrogate when the candidate scale and the reference scale does not measure the same thing.

This paper does not deny the practical value of the short questionnaires which are less expensive and could provide a useful feedback for the developers. However, these should not be referred as usability or UX scales. As it was shown, the PEU scale is unidimensional and reliable. PEU provides with a useful feedback on some key aspects of usability and could be combined with other short scales pointing to other usability / user experience aspects. This approach enables a step-by-step development of valid and reliable evaluation instruments and a flexible choice of scales, according to the objectives of the evaluation.

ACKNOWLEDGMENTS

This work was supported by the Romanian grant financed by ANCS under COGNOTIC 1609 0101 / 2016.

REFERENCES

- 1. Bevan, N., Carter, J. and Harker, S. ISO 2041-11 revised: What have we learned about usability since 1998? *Human-Computer Interaction Design and Evaluation, Proc. HCI International Conference, LNCS 9169*, Springer, (2015), 143-151.
- Borsci, S., Federici, S., Bacci, S., Gnaldi, M. and Bartolucci, F. Assessing User Satisfaction in the Era of User Experience: Comparison of the SUS, UMUX, and UMUX-LITE as a Function of Product Experience. *International Journal of Human-Computer Interaction*, 31(8), (2015), 484-495.
- 3. Brooke, J. SUS-A quick and dirty usability scale. *Usability evaluation in industry*, 189 (194), (1996), 4-7.
- Cairns, P. A Commentary on short questionnaires for assessing usability. *Interacting with Computers* 25(4), (2013), 312-316.
- Davis, F.D., Bagozzi, R.P. and Warshaw, P.R. User acceptance of computer technology: A comparison of two theoretical models, *Management Science*, 35 (8), (1989), 982-1003.
- 6. Finstad, K. The usability metric for user experience. *Interacting with Computers* 22(5), (2010), 323-327.
- 7. Gerbing, D.W. and Anderson, J.C. An updated paradigm for scale development incorporating unidimensionality and its assessment. *Journal of Marketing Research* 25(2), (1988), 86-192.
- 8. Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E. and Tatham, R.L. *Multivariate Data Analysis*. 6th ed., Prentice Hall, 2006.
- 9. Iordache, D.D., Pribeanu, C., Lamanauskas, V. and Raguliene, L. Usage of Facebook by university students in Romania and Lithuania: a comparative study. *Informatica Economica* 19(1), (2015), 46-54.
- 10. ISO/IEC 9126-1, Software Engineering –Product quality. Part I: Quality Model, 2001.
- 11.ISO 9241-11, Ergonomic requirements for office work with visual display terminals (VDTs) - Part 11 *Guidance on usability*, 1998.
- 12. ISO/IEC 25010, Systems and software engineering Systems and software product Quality Requirements and Evaluation (SQuaRE) – System and software quality models, 2011.

- 13. Konradt, U., Wandke, H., Balazs, B. and Christophersen, T. Usability in online shops: scale construction, validation and the influence on the buyers' intention and decision. *Behaviour and Information Technology* 22, (2003), 165–174.
- 14. Lewis, J. R., Utesch, B. S. and Maher, D. E. UMUX-LITE: when there's no time for the SUS. Proceedings of *CHI 2013*, ACM, (2013), 2099-2102.
- 15. Lindgaard, G. and Kirakowski, J. Introduction to the Special Issue: The tricky landscape of developing

rating scales in HCI. *Interacting with Computers* 25 (4), (2013), 271-277.

- 16. MacKenzie, S. B., Podsakoff, P. M., and Podsakoff, N. P. Construct measurement and validation procedures in MIS and behavioral research: Integrating new and existing techniques. *MIS quarterly*, 35(2), (2011), 293-334.
- 17. Sauro, J. and Lewis, J. When designing usability questionnaires, does it hurt to be positive? Proceedings of the *CHI* 2011, ACM, (2011), 2215-2223.

Automated Evaluation of Menu by Guidelines Review

Sara Bouzit^{1,2}, Gaëlle Calvary², Denis Chêne¹, Jean Vanderdonckt³

¹Orange Labs, 28 chemin du Vieux Chêne, F-38240 Meylan (France)

{sarah.bouzit, denis.chene}@orange.com

²Univ. Grenoble Alpes, LIG,

CNRS, LIG, F-38000 Grenoble (France) - {sara.bouzit, gaelle.calvary}@imag.fr

³Université catholique de Louvain, Louvain School of Management, Louvain Interaction Lab.

B-1348 Louvain-la-Neuve (Belgium) - jean.vanderdonckt@uclouvain.be

ABSTRACT

This paper presents ERGOSIM, a software that automatically evaluate the design of menu bars, pull-down menus, and sub-menus of a graphical user interface by reviewing usability guidelines related to menu design. In this method, a menu design is parsed against the definition of usability guidelines in order to detect potential usability problems manifested by any occurrence where a guidelines is not respected. Four evaluation strategies are enabled depending on the end user's preferences: an active strategy initiated by the system, a passive strategy initiated by the designer, a mixed strategy collaboratively initiated by both the designer and the system, and a strategy by conceptual units based on the domain. From an initial corpus of 312 usability guidelines compiled from different sources on menu design, a final knowledge base of 58 implemented usability guidelines has been obtained for automatic evaluation. By examining how each usability guideline for menu design is expressed, we discuss to what extent such guidelines could be automated in an automated process by guidelines review.

Author Keywords

Automatic evaluation; computer-aided design; evaluation strategy; heuristic evaluation; menu bar; pull-down menus; sub-menus; usability guidelines.

ACM Classification Keywords

Human-centered computing, Graphical user interfaces. *Human-centered computing - Heuristic evaluations*. Human-centered computing - User interface management systems

INTRODUCTION

In order to assess the usability of a User Interface (UI) and therefore to improve it, the temptation has been followed since years to replace a human (manual) evaluation of this usability by a system (automatic or semi-automatic) evaluation for several reasons [8,13,19,31]: to reduce human resources (e.g., by being released from involving usability experts), to reduce budget resources (e.g., by reducing the time and the resources needed to conduct such an evaluation), to guarantee the quality of the results (e.g., to ensure consistency across several evaluations, even if automated evaluation cannot cover all aspects, to establish a systematic evaluation by reducing missing spots, to minimize false positive and false negative), or to give the label of a usability standard (e.g., by certifying that a particular UI is compliant with a style guide, a corporate design guide or an established standard).

These reasons are considered even more important when the UI has some special status: a UI for a safety-critical system [32] for which it is crucial not to miss any potential defect, a very large UI for which there are so many screens that evaluating them becomes too tedious and repetitive [10], an adaptive UI for which adaptation could give rise to many different configurations to evaluate [11]. Many variables need to be decided when automatically evaluating a UI [8, 14, 19]:

- What type of method: several evaluation methods that are good candidates for conducting an automated evaluation, but they do not give all the same type of results. For instance, heuristic inspection, standard compliance, guideline review, cognitive walkthrough, user testing.
- *What type of software*: on-line vs off-line software or mixed solutions exist in order to capture UI data at runtime as well as users or contextual data or not.
- What type of usability knowledge: usability guidelines as well as accessibility guidelines are two representative examples of usability knowledge used for guideline review.
- What type of UI: as opposed to stand-alone UIs which are more difficult to grasp regarding their code access, web UIs are in principle easier for accessing the HTML code, parsing it, and conducting evaluation. Stand-alone UIs have other sources, like log files, resource files, screen analysis, static analysis, dynamic analysis.
- *What type of scope*: the scope of the evaluation could be also very different, ranging from local evaluation of typed UI elements to global evaluation of all UI elements, including their presentation, navigation, and their contents.

What type of purpose: detecting usability problems or certifying that there should not be such usability problems are two inverse approaches. Between exists the wish to have a simple diagnosis to assess the current usability of a UI in order to locate its quality with respect to competitors.



Figure 1. ERGOSIM main screen.

This paper presents ERGOSIM (Figure 1), a software for automated usability evaluation of the menu of a Graphical User Interface (GUI) with the following original aspects:

- *Type of user interface*: a GUI is the focus of the application with standard menus, no adaptable or adaptive menus are considered since they stem for other evaluation methods. Although ERGOSIM is developed on the MS Window platform, it is expected that any IBM Common User Access (CUA)-compliant menu is targeted by ERGOSIM, therefore not assuming that a particular operating system is required to design the menu.
- Type of method: guideline review has been decided in order to confront the GUI against a set of usability guidelines that belong to the literature. It is expected that the menu is built at design time (not at run-time) by a designer or a developer. Guideline review [23] consists of selecting a set of relevant guidelines and to examine a UI against this set of guidelines [32]. Two models are prevalent [4]: a binary model where a guideline is considered violated when there is at least instance on the GUI where the guideline is not respect or a linear model where all occurrences of guideline violations count per screen, along with a weight expressing the level of importance of the guideline. The binary model is mainly used here, but with different strategies that will be detailed.
- *Type of software*: a stand-alone application has been decided to enable the designer to build the menu bar, the pull-down menus and the submenus during the development phase of detailed design. In order to preserve continuity with the rest of the development life cycle, ERGOSIM can

export a menu design as a resource file to be included in a Windows application project. Other export formats, like UsiXML, could be imagined as well, but are not covered. Similarly, it could be imagined that a resource file could be imported for further evaluation.

- *Type of scope*: ERGOSIM focus on only one part of the GUI: the menu bar with its pull-down menus, cascading menus and sub-menus. Several reasons motivate this choice: the menu has never been covered per se by automated evaluation, the menu is probably one of the most frequently used interaction technique in many interactive applications and systems [2], many different types of menu exist [2] although this paper is not aimed at evaluating them all, there is a significant body of knowledge on menu [22], many usability guidelines are widespread in the literature, menu design is a familiar design activity and the menu is an object that could be easily controlled.
- *Type of purpose*: the goal of ERGOSIM is to support the designer while designing the menu, not to conduct an evaluation afterwards when the entire GUI is developed and to help novice designers learning usability knowledge regarding menu design in context [18, 26].

In order to introduce ERGOSIM and to explain how menu usability guidelines are automatically evaluated, the remainder of this paper is structured as follows: Section 2 will review some selected contributions in the area of UI automatic evaluation without conducting a systematic literature review, Section 3 will elaborate on the design and the implementation of ERGOSIM, Section 4 will discuss to what extent usability guidelines have been implemented in ERGOSIM, and Section 5 will conclude the paper by discussing future avenues to this work.

RELATED WORK

There are many pieces of work related to automatic UI evaluation in general, like Ivory's state-of-the-art [19], although it is no longer up-to-date. A good review is provided in [8, 14]. In this section, we only review some selected work with a focus on menus.

METROWEB [9] enables the designer to access to one or many usability knowledge bases that are presented as hypermedia with faceted search. A typical knowledge base consists of guidelines of any type, along with its ergonomic criteria [5], its linguistic level [30], its impact factor, and positive/negative examples illustrating good and bad practice related to the guideline. Multiple knowledge bases could be accessed and a faceted search could query these bases like "Give me all guidelines related to menu design" by selecting appropriate values for search criteria [30]. Selected guidelines could then be exported in a special section, e.g., for producing an evaluation report. Although METROWEB provides adequate access to usability knowledge, it is the designer's responsibility to correctly apply them or evaluate them. It has been demonstrated that designers relying on METROWEB manipulate more usability guidelines than without and that the UI resulting from this exercise satisfy more guidelines than without [9].

ERGOVAL [17] is a pioneering attempt to automatically evaluate GUIs against usability guidelines (by guideline review) for stand-alone applications. The authors report in their feasibility analysis that a ratio of **40%** has been reached between the guidelines candidates to automated evaluation and their feasible final implementation. They ask the question: what is the limit of automated evaluation?

BOBBY [10] automatically evaluate accessibility guidelines of web sites by guideline review of W3C accessibility guidelines. A pilot study revealed that the ratio could reach up to **50%** for accessibility of web sites since the HTML code of web pages is in principle easily accessible. Nowadays, this ratio is no longer that high with dynamic web pages and CSS3 style sheets [24]: it is around **30%** according to a qualitative estimation.

KWARESMI [4] also automatically evaluate usability and accessibility guidelines of web pages, either on-line or offline, by guideline review. This process is structured as follows: any candidate guideline is first encoded in GDL (Guideline Definition Language), a XML-compliant language for specifying a guideline based on first-order predicate logic on HTML tags, then incorporated into an evaluation base that is then parsed on-demand for a set of web pages. The advantage is that the evaluation engine is independent of the guidelines encoded in one or many knowledge bases [29]. Although no empirical study has been conducted yet on this software, it is also estimated that a ratio of **30%** of automatable guidelines could be reached, but with different types of restrictions depending on the tags involved in the GDL rule. Several other software follow the same principle, such as ErgoManager [1], ErgoCoIn [21], MAUVE [29], with a higher degree of flexibility when it supports dynamic web sites instead of a current version of a web page at runtime.

RITA [6] provides a more comprehensive framework for automatic GUI evaluation by considering not only a large set of guidelines, but also by relating them to quantitative data, such as task execution time, task completion rate, error rate, and interaction traces. In this way, RITA establishes a bridge between a qualitative evaluation based on guidelines review and a quantitative evaluation based on metrics.

EISEVAL [15] automatically evaluate usability guidelines on the GUI of an interactive application implemented according to the paradigm of a multi-agent software architecture. In this way, the evaluation consists of a set of autonomous agents which can query different parts of the interactive application so as to gather date and establish a diagnosis based on these data. Multiple agents could be incorporated that conduct different types of evaluation independently of each other, perhaps also with the same guidelines or different ones. In [7], a system is presented that hold the evaluation logic in the very right widgets used by the end user, instead of other modules of the interactive applications.

MENUSELECTOR [25] is a software for rapid prototyping of menu designs by considering different physical parameters like location, orientation, selection mechanism, and group clustering. This approach is purely syntactical since there is no automatic evaluation of the menu being designed, but the automatically generated HTML code could be subject to a further analysis conducted in another software.

MENUDESIGNER [28] is aimed at automatically generating a menu bar, associated cascading menus and menus items based on an activity chaining graph representing possible hierarchical navigation based on a task model. This approach remains static (the menu structure is generated once for all), without any adaptation and could lead to inconsistent menus when items are arranged.

MENUOPTIMIZER [3] is aimed at helping designers and developers to optimize the menu structure by maximizing consistency vs performance based on ant colony algorithm. While MENUOPTIMIZER reveals the popularity of menu items by a color line under each menu item, thus leaving the menu structure untouched, it does not provide end users with an adaptive menu. Matsui & Yamada [20] relied on a genetic algorithm to generate a menu structure that is optimized for its usage.

Adaptivity Animated transitions [12] have also been successfully used to explain to the end user how a UI has been adapted, including for menus [12]: each adaptation operation performed on a GUI is captured, scripted and could be played or replayed at the end user's pace, thus providing some visual explanation of the adaptation. The major drawback was the lack of animation control: not all steps should be animated equally to understand.

Usability Category	Guidelines
Optimizing the user experience	29
Hardware and software	4
The homepage	12
Page layout	9
Navigation	27
Scrolling and paging	3
Headlines, titles and labels	18
Links	21
Text appearance	18
Lists	13
Screen based controls (widgets)	27
Graphics, images and multimedia	17
Writing web content	18
Content organization	8
Search	16
Total Guidelines	240

Table 1. Guidelines implemented in USEFul.

The closest work to ErgoSIM is probably USEFul [13, 14], a complete framework for automatic evaluation of web sites against usability guidelines, also by guideline review. Table 1 summarizes the various categories of guidelines implemented in USEFul. Each guideline cannot be implemented with the same level of support. For this purpose, USEful distinguishes three levels of implementation [14]:

- 1. *Green*, when the guideline can be fully implemented: the framework is able to automatically determine whether this guideline applies to the web site being evaluated and the results related to this guideline are conclusive since these types of guidelines are typically measurable, with clearly defined parameters.
- 2. Amber, when the guideline is harder to fully implement in the USEFul framework: certain patterns have been used in order to determine whether this guideline may apply to the web site being evaluated and then transformed into a corresponding code. This guideline could be upgraded by augmenting the guideline evaluation by other mechanisms than guideline review, such as with machine learning processes or artificial intelligence algorithms. The results provided by USEFul for this guideline consist of data that can assist the designer in checking whether it applies to the web site being evaluated or not and support the designer in conducting the evaluation of this guideline which required human interpretation.
- 3. *Red*, when the guideline is too abstract to warrant any implementation and requires user intervention or too advanced algorithms to make it possible for it to be implemented in the framework. Through the use of such sophisticated algorithms, a guideline could be upgraded to "amber" or "green" levels. In its current definition, USEFul lists this guideline so that the

designer can be manually checked if it applies to the web site being evaluated.

In conclusion, one can observe that today there is no software like ERGOSIM to perform automatic evaluation of usability guidelines related to menus of GUIs of interactive applications at design time. Some software could be however tailored for this purpose, although they mainly work over web applications for which the HTML code is downloadable as opposed to a stand-alone application.

DEVELOPMENT OF ERGOSIM

Design options

This section presents the major design options decided for developing ERGOSIM and discusses the rationale behind.

Multi-view visualization. We hereby define a *UI view* as any representation of a final UI involved in a development life cycle. A UI view may be textual, graphical or both, based on a data structure or not [12]. By observing existing UI development methods and development life cycles, UI views can be roughly classified into three categories (Figure 2):

- 1. *Conceptual View* (CV): describes a conceptual representation of a UI of interest based on semantics, syntax, and stylistics. Typical examples include: UI models for domain, functional core, resources, and dynamic aspects. A conceptual view is the designer's view at early stage.
- 2. *Internal View* (IV): consists of the UI code in any programming or markup language. An internal view is the typical developer view for developing a particular UI.
- 3. *External view* (EV): refers to the final UI that is visible and executable by the end user.



Figure 2. Possible paths between UI views.

During the development life cycle, at design-time as well as at run-time, various UI stakeholders can create, retrieve, modify, delete, or simply execute any UI view or view element: for instance, while a designer is responsible for the conceptual view, the developer is responsible for the internal view, and the end user accesses the external view for comments, testing, and validation. A development path may be initiated from any view and could proceed with any other view, including itself, which are respectively represented by arrows and loops in Figure 2. ERGOSIM structures its environment similarly into three views (Figure 1) [12]:

1. *External view* (EV): refers to the final representation of the menu as it is visible and executable by the end

user with the same Look & Feel as it should be in the end.

- 2. *Internal View* (IV): consists of the menu structure decomposed into levels of the menu hierarchy, as it is stored for instance in a resource file.
- 3. *Conceptual View* (CV): describes a conceptual representation of a menu in terms of design options and parameters for each menu item or menu group, which includes:
 - a. The *menu label*, which contains the textual label of the menu item, along with the "&" character representing the mnemonic of the menu item. The character after this delimiter is underlined.
 - b. The *mnemonic* of the menu item, which is the character to be pressed by combining it with the "Alt" key instead of selecting it by pointing, e.g. "Alt + 3 for "Save". The range of possible mnemonics for a label is automatically generated from the menu label (Figure 3) and the designer can choose among them by moving a cursor on it. This is important since a usability guideline states that a mnemonic should always be chosen among the real letters of the label, preferably those that are pronounced.
 - c. The *menu activation status*, qui specifies whether the item is by default activated or deactivated (greyed).
 - d. The *menu shortcut*, qui defines the sequence of keys to be pressed for directly accessing the menu item, which consists of normal keys, i.e., A, B, C, ..., X, Y, Z, 1, 2, ...9, 0, F1, F2, ..., F11, F12, Del, Ins, ... and control keys, i.e., « Crtl », « Alt », « Shift ».
 - e. The *menu attachment type*, which specifies whether a menu item is related to displaying a sub-menu (for instance, a pull-down menu or a cascading menu), to opening a dialog box or a secondary window, or to triggering directly a function of the application.
 - f. The *contextual help message*, which specifies the message to be displayed in the status bar when the menu item is highlighted, but not yet selected.

The menu bar is consequently equipped with traditional facilities for menu management, such as creating, updating, deleting a menu item, a group of items, an entire menu or a menu bar. Note that this conceptual view could be expanded in the future with other parameters, such as the associated earcon or the gesture to trigger the same item, but these options are not covered yet by usability guidelines.

Multi-strategy evaluation. The target users of ERGOSIM, theoretically any stakeholder involved in the UI development life cycle but practically the most often, designers and developers, could exhibit very different profiles in terms of background and level of experience [26].

Adrian Iftene, Jean Vanderdonckt (Eds.)

Dénomination de l	item						_
&Enregistrer							
L							
Mnémonique							_
							-
Enr		n i	«	÷	÷	è	;
	Ŭ	9 '	·	`	<u> </u>	Ŭ	· ·
Etat de l'item	Baco	courci cl	avier				
Activé	S	-	Ctr	l+S			
C Désactivé		<u>M</u> aj					
						- 6 - 1	
	I ₩ ĭ	ru ∣	Ą	τ.		alidei	

Figure 3. Conceptual view of a menu item.

In order to support this variation, the automatic evaluation of the menu being designed could be achieved flexibly according to parameters specified in the *Evaluation parametrization* window (right part of Figure 1), which offers four evaluation strategies:

- 1. An *active evaluation strategy*, where the end user decides when and how the automatic evaluation will take place. This strategy is qualified as "active" because the stakeholder actively participates in the evaluation.
- 2. A *passive evaluation strategy*, where the system automatically evaluates the menu being designed without any intervention of the end user. This strategy is qualified as passive since the stakeholder has no control on the evaluation process.
- 3. A *mixed initiative strategy*, which is located mid-way between the active and the passive strategies, where the end user can parametrize the evaluation based on several parameters:
 - a. The *amount of user actions*: which captures the amount of all elementary actions performed by the user, such as menu item editing, pull-down menu editing, etc. In this way, it is possible to trigger the evaluation every 5 actions.
 - b. The *type of user actions*: which categorizes the level of actions performed by the end user: elementary, intermediate, or complete. In this way, it is possible to trigger the evaluation when a complete pull-down menu is finished, therefore not interrupting the user in the design process. Task switching between a design activity and an evaluation activity should be minimized.
 - c. The *amount of usability problems*, which specifies the amount of usability problems detected after which the evaluation could be triggered. In this way, it is possible to trigger an evaluation after a certain amount of problems has been detected, which is particularly useful when problems are generated in cascade: one usability problem may immediately induce some other related problems.
 - d. The *type of usability problem*, which specifies the level of importance of a detected violation of a usability problem, ranging from 1 (cosmetic) to 5 (critical). The level of importance of a guideline is stored in its definition or automatically suggested from the linguistic level: the higher the linguistic level is, the higher becomes the level of

importance. In this way, it is possible to trigger an evaluation when a problem with a given severity is detected, and not just after any occurrence of a detected violation.

4. A *strategy based on conceptual units*, where the end user decides when the system should evaluate significant parts of the menu being designed, based on conceptual units. A *conceptual unit* is defined as a non-elementary menu group unit, such as an entire group of menu items delineated by separators in a pull-down menu, an entire pull-down menu, a cascading menu or the whole menu bar. In this way, the evaluations could be triggered as soon as a significant part of the menu has been completed, not before. This strategy is qualified as "based on conceptual units" since the evaluation scope is on a menu part that has some semantic meaning, not an elementary item.

Note that for the moment in the mixed-initiative strategy, the evaluation is triggered only based on simple conditions with a threshold, such as when the amount of problems \geq 5 or when 2 important problems have been detected. A cursor between these strategies (Figure 1) enables the end user to gracefully evolve between strategies.

Parametrizable feedback. In addition to the evaluation strategy, the end user may want to specify the *level of feedback detail* that governs the way feedback messages are presented to the end user after an evaluation has been performed. This level of feedback could be stated to [16]:

- *Elementary*, when only the short title of the usability guideline violated is presented for each occurrence of a usability problem, along with its location (see the message window at the bottom of Figure 1).
- *Intermediate*, when the complete title of the guideline violated is presented for each occurrence of a usability problem, along with its location and the level of importance.
- *Detailed*, when the message contains the full set of information on any detected usability problem: the complete title, the ergonomic criteria from Bastien & Scapin [], the linguistic level, positive and/or negative examples, references where the guideline is documented, along with information on the location and a possible help on how to fix it.

Any occurrence of a detected usability problem is displayed in the message window with or without a timestamp. The message window could be purged at any time via an appropriate push button. The amount of information displayed in the message window can be tailored (Figure 4).

Management of user profiles. A user profile could be created and updated at any time that captures the parameters:

The *level of experience*, which specifies the level of usability experience in general and more specifically for menu design: low experience, medium experience, or high experience. Based on this value, ERGOSIM can automatically assign predefined values to other

parameters, like the evaluation strategy so that the end user should not necessarily fill in all the parameters before starting. In this way, if the end user selected "low experience", ERGOSIM will pick the passive evaluation strategy and the elementary feedback. It the end user estimates herself as "moderately experienced", ERGOSIM will pick the mixed-initiative evaluation strategy with a feedback every 5 significant actions. If the end user estimates herself "highly experienced", ERGOSIM will pick the active strategy. The user can change the values of these parameters at any time.

- The *evaluation strategy* that is preferred by the end user, according to the aforementioned definition.
- The *level of feedback detail* that is preferred by the end user, according to the aforementioned definition.
- The evaluation parametrization options (Figure 5).

Options	\times
Affichage Evaluation	
✓ Afficher la barre d' <u>o</u> utils	Ok
✓ Afficher la <u>b</u> arre d'état	
Fenêtre de messages	Annuler
Hauteur : 5 🗲 lignes Nettoyer	Aide
Nouvel item de menu	
Afficher la dénomination par défaut	
Fenêtre du menu Largeur de l'indentation : 5 🚖 espaces	
Confirmer la <u>f</u> ermeture de la fenêtre de travail	

Figure 4. Evaluation display options.



Figure 5. Evaluation parametrization options.



Figure 6. Activation of selected guidelines.

Evaluation parametrization options (Figure 5) enable the end user to tailor various options that drive the evaluation of guidelines themselves:

- The *list of possible guidelines*: any guideline can be activated or de-activated momentarily and this configuration can be saved in a configuration file (Figure 6).
- The *list of inappropriate terms*: since no natural language understanding is incorporated, the end user may want to specify a series of terms that hold a negative connotation, whose usage is therefore prohibited. For instance, "Abort" in English is inappropriately translated into "avorter" in French, which is irrelevant (Figure 7).
- The *list of interdependent terms*: for the sake of the evaluation based on conceptual units, a series of constraints could be imposed to establish and maintain semantic relationships between terms that have some interdependency. For instance, "Save" and "Save as" should be located one after another, "Open" and "Close" or synonyms should be grouped in a same group of menu items to convey disclosure.
- The *list of computer-based terms*: this includes terms that are considered as jargon terms belonging to the area of computer science, such as "bandwidth", "baud rate", "bitmap", "memory dump". These terms should be avoided.
- The *list of abstract terms*: in order to evaluate the general guideline stating that a menu item should be ideally structured in a simple sentence composed of an action verb followed by an object on which the action is executed (action-object paradigm) or vice-versa (object-action paradigm), this list contains verbs that are considered too abstract or generic to be used in appropriate menu design.
- The *definition of standard menu items*: standardized menus as found in standards like IBM Common User Access (CUA), in software vendors or operating systems style guides (e.g., MacOs, Ubuntu, MS Windows) can be defined once for all in a profile so as to be compliant with these sources (Figure 7).

iste des dénominations à connotation	×
avorter	<u>M</u> odifier
Liste des dénominations à connotation	
abandonner	<u>N</u> ouveau
accéder disponible booter exécuter toucher implémenter invalide	\$ \$
tuer	<u>S</u> upprimer
<u>F</u> ermer Aid	e

Figure 7. Definition of inappropriate labels.

liste des items standards		>
Item standard		
Dénomination de l'îtem : &Fichier		Modifier
✓ Appartient à la barre de menu		
Dénomination de l'îtem garent :		
Sous-menu ou menu <u>d</u> éroulant		
<u>R</u> accourci clavier :		
Liste		
'&Fichier', 1, 1, ,	^	Nouveau
"&Nouveau", U, U, 'Utrl+N', '&Fichier' '&Duvrir', 1, 0		
'&Ouvrir', 0, 0, 'Ctrl+O', '&Fichier'		
"&Enregistrer', 0, 0, 'Ctrl+E', '&Fichier'		ŵ
'&Quitter', 0, 0, 'Alt+F4', '&Fichier'		
'&Edition', 1, 1, .		-0-
"&Annuler", U, U, "LtrI+2", "&Edition" "&Couper", D, D, "CtrI+2", "&Edition"		
'Co&pier', 0, 0, 'Ctrl+C', '&Edition'		Supprimer
'C&oller', 0, 0, 'Ctrl+V', '&Edition'	~	Tobbung
<u>F</u> ermer	Aide	

Figure 8. Definition of standard items.

Development of ErgoSim

ERGOSIM has been developed in Borland Pascal for Windows 7.0 because of the object-oriented facilities offered by the procedural language and its corresponding environment, but also for the object-oriented database in which each usability guideline will be stored as a record. The software architecture of ERGOSIM is composed of three modules:

1. The evaluation triggerer: this module receives as input any action executed by the end user performed on the internal or the conceptual view of the menu and the values of options contained in the user profile, the most important being the evaluation strategy with its parameters. This module then triggers an evaluation of the menu being designed based on the evaluation strategy and other parameters on the end users actions performed since the last evaluation. These actions include, but are not limited to: modifying the label of a menu item, inserting a pull-down menu, inserting a new menu item, modifying the menu bar, defining the shortcut of a menu item, moving a group of items from one sub-menu to another menu, moving a group of items to a sub-menu, using standard menu items in their standard format.

- 2. The *evaluation engine*: this module receives as input a knowledge base of usability guidelines and the internal representation of the menu from the triggerer and performs the evaluation according to parameters set by the triggered to return the results of the evaluation. After an evaluation has been performed, all parameters regarding the amount of actions, problems, etc. is reinitialized. The system does not keep trace of usability problems that are not solved: it simply rechecks them at any evaluation. The evaluation engine is independent of the knowledge base containing the evaluation logic.
- 3. The *evaluation presenter*: this module receives from the evaluation engine the results of a performed evaluation and produces the output according to the evaluation display options (Figure 4) and user profile. The results are displayed in the message window (bottom right of Figure 1).

AUTOMATIC EVALUATION OF USABILITY GUIDELINES

Although ERGOSIM could accommodate one or several different knowledge bases, it was decided to compose one comprehensive knowledge base containing all the possible guidelines on menu design. For this purpose, we compiled usability guidelines from two major sources: Scapin's *guide ergonomique* [23] and Vanderdonckt's *ergonomic guide* [27], which is itself a compilation of usability guidelines. This compilation resulted into a base of 362 unique usability guidelines (without double entries), which is considered as the set of initial guidelines subject to automatic evaluation.

From this initial set, only 58 usability guidelines out of 362, have been finally implemented, which represents a ratio of **16%**. If we count usability guidelines that are intrinsically respected by the operating system, the software environment of ERGOSIM or ERGOSIM itself due to its implementation (e.g., some guidelines are intrinsically respected when displayed in the external view), this ratio reaches to **36.5%**.



Figure 9. Distribution of implemented usability guidelines.

The full distribution of usability guidelines is graphically depicted in Figure 9. Usability guidelines fall into 4 categories depending the level with which they could be implemented, similarly to the USEFul's level of implementation [14]:

- 1. *Irrelevant or not applicable guidelines*: this category contains guidelines whose application is probably relevant to menu design in general, but not for menu bar, pull-down menus, and cascading menus or guidelines that cannot be applied practically. Table 2 reveals some significant examples of such guidelines along with a comment explaining why they cannot be applied.
- 2. Non-implementable guidelines: this category contains guidelines whose interpretation and/or application is impossible to replicate by a software for different reasons (Figure 9): guidelines are expressed at a too high level of abstraction that prevent them to be interpreted by an automaton, guidelines that require additional information related to the user, the platform, the environment or the whole context of use that is unknown at design time, guidelines that require additional information that cannot be obtained by any means, guidelines that require understanding of the natural language in which the guideline is expressed, and guidelines whose development would be so complicated that they would require a significant amount of time for a small benefit. Note that these reasons are independent of the environment in which ERGOSIM is implemented. Table 3 reveals some significant examples of such guidelines.
- 3. Intrinsically respected guidelines: this category contains guidelines that are intrinsically respected either by the operating system (here, MS Windows), the development platform (here, Borland Pascal) or the environment of ERGOSIM itself. For instance, the guideline "every menu item should be either activated or deactivated" is automatically ensured by the conceptual view of ERGOSIM. Similarly, the guideline "Shortcuts should always be visible" is straightforwardly ensured by MS Windows. The guideline "A main menu should always exist" is also intrinsically established by MS Windows since a menu bar is always created, even if minimal. The guideline "Shortcuts should always made visible" is ensured by ERGOSIM itself since the conceptual view automatically propagates this design choice on the external view, thus making them visible automatically. The guideline "Menu items should be perceptually distinct from each other" is ensure by both MS Windows and Borland Delphi since menu items in the external view always presented with the same space between and separators defined by the end user in the group.
- 4. *Implemented guidelines*: this category contains guidelines implemented in some way in ERGOSIM, which are further refined into three sub-categories depending on their complexity and the level with which the scope of the guidelines could have been addressed (Figure 9):

- a. *Easy guidelines*: guidelines that are straightforwardly implemented. For instance, "The menu breadth should not exceed 8 items" required 17 Lines of Code (LOC) in Borland Pascal, "The menu items should have unique labels" required 28 LOC.
- b. *Difficult guidelines*: guidelines that are implemented but with some restrictions in their interpretation [30]. For instance, "The numbering of menu items should continuous" required 60 LOC, "Menu mnemonics should be phonetically distinct" required 43 LOC, based on an existing SOUNDEX algorithm testing whether two strings are phonetically close or not.
- c. Very difficult guidelines: guidelines that are implemented with advanced techniques or significant restrictions. For instance, "Mutually exclusive items or interdependent items should be grouped together" required 76 LOC, "Menu items should avoid abstract terms and prefer action verbs" is only 26 LOC because it merely tests that all items do not belong to a list of predefined terms considered abstract or not.

Guideline statement	Reason	
Menu items should be consistent from one application to another	ERGOSIM evaluates one menu design at a time and cannot compare with other menu design for other case studies.	
Full screen menus should be displayed at once, with one item per line	ERGOSIM does not cover full screen menus	
Linear menus should match user's expectations	There are no linear menus in ERGOSIM	
Network menus should follow a natural flow	There are no network menus in ERGOSIM	
Contextual menus should be displayed at their right location (top, bottom, left, right) depending on the task	ERGOSIM is focusing on menu bars, pull-down menus and cannot relocate such menus at different locations	
Items of pull-down menus attached to a label of the menu bar could be colored in the same way	This guideline is mostly applicable to web sites and item coloring is an unsupported feature	

Table 2. Examples of irrelevant guidelines.

Guideline statement	Reason	
The menu design should be based on a metaphor of a mini-world, based on real task options	Impossible to interpret unless a model of the mini-world is available along with a task model	
Use menu selection technique that is precise enough	This implies to rely on a menu performance model, which exists,	

	but is another input.
The complexity of menus should reflect the end user's experience	This requires to access a user model
Only action verbs of natural language should be used	This requires a thesaurus of action verbs for the natural language used
Menu items should have unique meanings	This requires a module for natural language processing based on a semantic network
Menu items should avoid any humor	This requires an interpreter of natural language

Table 3. Examples of non-implementable guidelines.

CONCLUSION

This paper presents ERGOSIM, a software that automatically evaluate the design of menu bars, pull-down menus, and sub-menus of a graphical user interface by reviewing usability guidelines related to menu design in a design-time environment, thus preserving the continuity between design activities and evaluation activities. ERGOSIM automatically evaluate 16% of usability guidelines for menu, or 36% if we count intrinsically respected guidelines. This is rather different from ERGOVAL [17], whose authors argue that the ratio should be between 44% and 78% or from BOBBY [10], whose authors argue that a ratio of 50% was reached. In the last case, web sites were automatically evaluated against usability and accessibility guidelines, which is considered as an easier case since the HTML code is accessible, perhaps also with the Document Object Model (DOM) containing the structure of the web page and the CSS. A closer observation of guidelines that are finally supported by the whole environment could be classified as follows by linguistic level:

- Guidelines belonging to the physical and alphabetical levels are almost always established by construction of the menu bar, the pull-down menus and the cascading menus. Changing the alphabet is also possible, but in another environment.
- Guidelines belonging to the lexical level are almost all supported since they are all easy to implement.
- Guidelines belonging to the syntactical level are often supported, sometime with a more advanced technique.
- Guidelines belonging to the semantic level could be sometimes implemented provided that some restriction, e.g. by replacing the full scope by a list of admissible values, is adopted.
- Guidelines belonging to the pragmatic and the goal levels are almost never possible to implement, unless additional models are made accessible, thus required artificial intelligence techniques, such as intelligent model-checking techniques, machine learning techniques, relevance feedback or reinforcement learning.

ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers for their constructive feedback on an earlier version of this paper and Grenoble INP for supporting this collaboration.

REFERENCES

- de Abreu Cybis, W., Scapin, D.L., Morandini, M. ErgoManager: A UIMS for monitoring and revising user interfaces for Web sites. In *Proc. of the 1st Int. Conf. on Web Information Systems and Technologies* (WEBIST'2005, Miami, 26-28 May 2005). SciTe Press, (2005), 281–286.
- Bailly, G., Lecolinet, E., Nigay, L. MenUA: A Design Space of Menu Techniques (2009). Retrieved from http://www.gillesbailly.fr/menua/ on May 27th, 2016.
- Bailly, G., Oulasvirta, A., Kötzing, T., and Hoppe, S. MenuOptimizer: interactive optimization of menu systems. In Proc. of the 26th ACM Symposium on User interface software and technology (UIST'2013). ACM Press, New York, (2013), 331–342.
- Beirekdar, A., Vanderdonckt, J., Noirhomme-Fraiture, M. A Framework and a Language for Usability Automatic Evaluation of Web Sites by Static Analysis of HTML Source Code. In *Proc. of 4th Int. Conf. on Computer-Aided Design of User Interfaces* (CADUI'2002, Valenciennes, 15-17 May 2002). Kluwer Academics Pub., Dordrecht, (2002), 337–348.
- 5. Bastien, Ch. And Scapin, D.L. Evaluating a User Interface with Ergonomic Criteria, *International Journal of Human-Computer Interaction* 7, (1995), 105–121.
- 6. Charfi, S., Ezzedine, H., and Kolski, Ch. RITA: a useR Interface evaluaTion frAmework. *Journal of Universal Computer Science* 21, 4, (2015), 526–560.
- Charfi, S., Trabelsi, A., Ezzedine, H., and Kolski, Ch. Widgets Dedicated to User Interface Evaluation. *Int. J. Hum. Comput. Interaction* 30, 5, (2014), 408–421.
- 8. Charfi, S. and Ezzedine, H. Evaluation tools through user participation techniques: Features, limitations, and new perspectives. In *Proc. of Int. Conf. on Advanced Logistics and Transport* (ICALT'2014, Hammamet, 1-3 May 2014). IEEE Press, (2014), 13–18.
- 9. Chevalier, A., Fouquereau, N., and Vanderdonckt, J. The influence of a knowledge-based system on designers' cognitive activities: a study involving professional web designers. *Behaviour and Information Technology* 28, 1, (2009), 45–62.
- 10. Cooper, M., Limbourg, Q., Mariage, C., and Vanderdonckt, J. Integrating Universal Design into a Global Approach for Managing Very Large Web Sites. In Proc. of the 5th ERCIM Workshop on User Interfaces for All (UI4All'99, Dagstuhl, 28 November-1 December 1999). A. Kobsa, C. Stephanidis (Eds.). GMD Report 74, GMD - Forschungszentrum Informationstechnik GmbH, Sankt Augustin, (1999), 131–150.
- 11. Dessart, C.-E., Genaro Motti, V., and Vanderdonckt, J. Showing user interface adaptivity by animated transitions. In *Proc. of ACM Conf. on Engineering*

Interactive Computing Systems (EICS'2011). ACM Press, New York, (2011), 95–104.

- Dessart, C.-E., Genaro Motti, V., and Vanderdonckt, J. Animated transitions between user interface views. In Proc. of ACM Working Conf. on Advanced Visual Interfaces (AVI'2012). ACM Press, (2012), 341–348.
- 13. Dingli, A. and Cassar, S. An Intelligent Framework for Website Usability, 2014. Advances in Human-Computer Interaction 2014. DOI: http://dx.doi.org/10.1155/ 2014/479286
- 14. Dingli, A. and Mifsud, J. USEFul: A Framework to Mainstream Web Site Usability through Automated Evaluation. *International Journal of Human Computer Interaction* 2, 1, (2011), 10–30.
- 15. Dung Tran, Ch., Ezzedine, H., and Kolski, Ch. EISEval, a generic reconfigurable environment for evaluating agent-based interactive systems. *Int. Journal Human-Computer Studies* 71, 6, (2013), 725–761.
- 16. Ericsson, M., Baurén, M., Löwgren, J., and Y. Wærn. A study of commenting agents as design support. In *Proc. of ACM SIGCHI Conference on Human Factors in Computing Systems* (CHI'1998). ACM Press, New York, (1998), 225–226.
- 17. Farenc, Ch., Liberati, V., Barthet, M.-F. Automatic Ergonomic Evaluation: What are the limits? In Proceedings of 2nd International Workshop on Computer-Aided Design of User Interfaces (CADUI'96). Presses Universitaires de Namur, Namur, (1996), 159–170.
- 18. Fischer, G., Lemke, A.C., Mastaglio, T.W., and Mørch, A.I. The Role of Critiquing in Cooperative Problem Solving. ACM Trans. Inf. Syst. 9, 2, (1991), 123–151.
- 19. Ivory, M.Y. and Hearst, M.A. The state of the art in automating usability evaluation of user interface. *ACM Computing Surveys*, (2001), 470-516.
- 20. Matsui, S. and Yamada, S. Genetic algorithm can optimize hierarchical menus. In Proc. of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'2008). ACM Press, NY, (2008), 1385– 1388.
- 21. Morandini, M., de Abreu Cybis, W., and Scapin, D.L. A Prototype to Validate ErgoCoIn: A Web Site Ergonomic Inspection Technique. In *Proc. of HCI International*'2009, (2009), 339–348.
- 22. Norman, K.L. *The Psychology of Menu Selection*, Designing Cognitive Control at the Human/Computer Interface. Ablex Publishing Corporation, Norwood, New Jersey, 1991.
- 23. Scapin, D.L. Guide ergonomique de conception des interfaces homme-ordinateur. Research report INRIA n°77. Institut National de Recherche en Informatique et en Automatique, Le Chesnay, 1986.
- 24. Schiavone, A.G. and Paternò, F. An extensible environment for guideline-based accessibility evaluation of dynamic Web applications. *Universal*

Access in the Information Society 14, 1, (2015), 111–132.

- 25. Spoidenne, J., Vanderdonckt, J. MenuSelector: Automated Generation of Dynamic Menus with Guidelines Support. In Proc. of 10th Int. Conf. on Human-Computer Interaction (HCI International'2003). Vol. 1, J. Jacko, C. Stephanidis (Eds.). Lawrence Erlbaum Associates, Mahwah, (2003), 233–237.
- 26. Sumner, T., Bonnardel, N., and Kallak, B.H. The Cognitive Ergonomics of Knowledge-Based Design Support Systems. In Proc. of ACM SIGCHI Conf. on Human Factors in Computing Systems (CHI'1997). ACM Press, New York, (1987), 83–90.
- 27. Vanderdonckt, J. *Guide ergonomique des interfaces homme-machine*. Presses Universitaires de Namur, Namur, 1994.
- 28. Vanderdonckt, J. Computer-Aided Design of Menu Bar and Pull-Down Menus for Business Oriented

Applications. In *Proc. of* 6th *Int. Workshop on Design, Specification, Verification of Interactive Systems* (DSV-IS'99). Springer, Vienna, (1999), 73–88.

- 29. Vanderdonckt, J. and Beirekdar, A. Automated Web Evaluation by Guideline Review. *Journal of Web Engineering* 4, 2, (2005), 102–117.
- 30. Vanderdonckt, J. Development Milestones towards a Tool for Working with Guidelines. *Interacting with Computers 12*, 2, (1999), 81–118.
- 31.Xiong, J., Diouf, M., Farenc, Ch., and Winckler, M. Automating Guidelines Inspection: From Web site Specification to Deployment. In *Proc. of CADUI'2006*, 273–286.
- 32. Xu, J., Ding, X., Huang, K., and Chen, G. A Pilot Study of an Inspection Framework for Automated Usability Guideline Reviews of Mobile Health Applications. In *Proceedings of Wireless Health 2014* on National Institutes of Health (WH'2014). ACM Press, New York, (2014), 1–8.

Perceived aesthetics of user-modifiable layouts: a comparison between an unspecified design and a GUI

Stefan Trausan-Matu University Politehnica of Bucharest 313 Splaiul Independentei, Bucharest, Romania

and

Research Institute for Artificial Intelligence and

Academy of Romanian Scientists stefan.trausan@cs.pub.ro

ABSTRACT

This paper presents the results of an experiment in which students are asked to reposition rectangular shapes in a window in two cases: when they are or they are not told that the window should be considered the configuration of a graphical user interface (GUI). The main aim of the research is to investigate if users have different aesthetical preferences in the two distinct cases. Other goals are to see if, in the case of a GUI, the aesthetical criteria relate to cognitive ergonomics and if there are some preferred shape patterns.

Author Keywords

Beauty measures; aesthetics; human-computer interfaces design; cognitive ergonomics

ACM Classification Keywords

H.5.2 User Interfaces: Graphical user interfaces (GUI).

INTRODUCTION

The omnipresence of computer-based devices such as smart phones, tablets, laptops and the large amount of time spent by people interacting with them obviously show the importance of paying a major attention to the design of their human-computer interfaces. In this direction, user models and cognitive ergonomics rules are considered in the Computer-Human Interaction (CHI).

However, one important human factor gets considerable less attention in the design of human-computer interfaces: aesthetics. People buy various fancy cases for their phones, upload or even buy nice backgrounds for their laptops or ringtones for the phones. Therefore, it is obvious that users seek aesthetical features of their devices. Even if "beauty is in the eye of the beholder", several general aesthetic principles may be identified, such as symmetry or harmony of colors.

There are also mathematical formulas in aesthetics, the most well-known being the golden ration or the ratios of nice sounding musical chords remarked by Pythagoras. In addition to these, we can mention George David Birkhoff's "measures of beauty" [1], Matila Ghyka [2-4], and Pius Servien [5], which tried to identify mathematical relations that characterize beauty.

Brahma Dathan Metropolitan State University Saint Paul, Minnesota, USA Brahma.Dathan@metrostate.edu

Cognitive ergonomics principles focus on eliminating cognitive load in using interfaces to computers, which means easing the understanding of their functionality and the finding of needed controls. Some of these principles, even if they are not explicitly linked to aesthetics, they are in consonance with them. For example, the grouping principle, which states that similar controls should be grouped, for example, that a series of check-boxes should be put on vertical or horizontal line, at equal distances, is in consonance to aesthetic principles of order, harmony or symmetry.

In CHI there were researches to identify measures or rules for assuring aesthetics of the Graphical Users Interfaces (GUIs) [6-13]. Some of them included experiments in which users were asked to grade the aesthetics features of interfaces [8, 11, 13]. Our approach introduces a new idea: to analyze how users would rearrange a set of GUI components in order to be both esthetical and usable. The considered research questions were: 1) do users have different aesthetical preferences in the distribution of shapes in the case of an unspecified image and a GUI? 2) Might some patterns of arranging shapes be identified? Consequently, an experiment was performed in which several groups of computer science students were put to rearrange a set of interface elements shapes and the results were analyzed according to several aesthetical criteria.

This paper presents in the next section an overview of the proposed mathematical relations for "measuring" beauty, and some researches in the analysis of the relations between the design of GUIs and aesthetics. The third section presents our experiment and associated discussions. The paper is completed by conclusions.

RELATED WORK

The mathematics of beauty

There were several approaches during history to identify mathematical formulae of nice looking shapes or pleasant to listen groups of sounds. Probably the most well-known mathematical measure of beauty is the golden section (or ratio), which was used for many artists. This ratio is related directly with Fibonacci numbers, which, in fact, both represent a natural law of growth, they may be found in many places in nature, for example in flower petals, spiral galaxies, snail shells' logarithmic spiral, etc. [14]. An extensive analyze of the presence of the golden ration throughout history was made by Matila Ghyka [2]. Maybe the fact that we consider beautiful shapes following the golden ratio might be explained exactly by its omnipresence in nature. Other approaches to analyze mathematical or physical properties of beautiful things were proposed by the same Matila Ghyka [3, 4] and by Pius Servien [5].

George David Birkhoff had important contributions in mathematics but he also investigated aesthetics, introducing a "measure of beauty". Birkhoff considered that beauty is directly proportional with order or organization (O) and inverse proportional with complexity (C) [1]:

$$\mathbf{M} = \mathbf{O} / \mathbf{C}$$

In other words, we may say that Birkhoff considered beautiful something simple and with an ordered/organized structure. He also gave metrics to various polygonal shapes, according to the above formula.

Measurements of aesthetics in Human Computer Interaction

The idea to analyze graphical user interfaces aesthetic features using metrics and other criteria was considered by a series of researchers [6-13]. Some approaches classify visual techniques, visual complexity elements or features that may be used for placing interface components [6, 10, 13]. For example, Vanderdonkt and Gillo identify five groups of visual techniques [6], from which we mention: *Physical* (balance, symmetry, regularity, alignment, proportion, and horizontality), *Composition* (simplicity, economy, etc.), *Association and dissociation* (unity, repartition, grouping, and sparing), *Ordering* (consistency, predictability, sequentiality, and continuity), and *Photographic* (sharpness, roundness, etc.).

Buanga identified a series of factors to be taken into account when analyzing graphical user interfaces, some of them being [10]: Visual balance (in vertical or horizontal axis), Proportion, Repetition, Rhythm, and Unity. Zain et al. consider Balance, equilibrium, symmetry, sequence, rhythm, order and complexity [11].

Many approaches use questionnaires for getting an image on the users' perception of the aesthetics of the interfaces [8, 11, 13]. These questionnaires are usually analyzed with statistical tools.

Mathematical formulae were proposed for various metrics for the analysis of the aesthetics of interfaces [7,9-12]. Some approaches developed software applications for evaluating the quality of existing interfaces (for example, the Web-based evaluator tool QUESTIM (Quality Estimator using Metrics) [9]) or Aesthetic Measurement Application (AMA) [11].

As a result of the performed research, some approaches propose design rules for arranging GUI components [6, 10].

THE EXPERIMENT

The research started from an inquiry about how people prefer to organize a set of shapes in a computer desktop window. One hypothesis from which we started was that people have different aesthetical preferences in the distribution of shapes in the case of an unspecified image and a graphical user interfaces. Another hypothesis was that people prefer some particular arranging, like rhythm, phenomenon more remarked in music and poetry [15].

The experiment consisted in assigning to several groups of students a homework in which they had to run a Java application which displayed a set of rectangles in a window (see Figure 1 for the initial configuration, common for all the groups) and they should rearrange them in a configuration that they consider aesthetical. A part of them were told from the beginning that the shapes in the window are elements of a GUI. A second part had to repeat the assignment and only in the second step they were told that they should think to a GUI configuration.



Figure 1. The initial configuration of the shapes to be moved.

Four groups of computer science students were involved in the experiment:

1) 13 master students in Artificial Intelligence,

2) 17 master students following an advance database systems course in master,

3) 11 sophomore students,

4) 34 students of a semantic web master course.

The first three groups were from the Automatic Systems and Computer Science Faculty of the University Politehnica of Bucharest and the latter from the Faculty of Engineering in Foreign Languages from the same university. The first three groups were of Romanian nationality and the fourth had Romanian, Arabic, and other European nationalities (Austrian, Serbian, etc.), therefore different cultural traditions were covered.

We will present first the results of the analysis of the 4th group. The experiment with this group had two steps. Students received a java program that displays an initial arrangement of rectangles (see Figure 1) and squares and allows them to rearrange them. The program generates an output file that students had to send by email to the professor.

In the first step, students received the following indications:

The program displays a window with several rectangles. Your task is to move the rectangles in such a way that the whole arrangement looks most pleasing to you, aesthetically. You must arrange the rectangles so that all the rectangles are displayed and no rectangle overlaps with another. If needed, you may increase the size of the window.

After the deadline of a week, students got the indications for the second step, in which they were announced to reconsider the arrangement they made in the first step, this time taking into account that the shapes are of a graphical user interface:

Re-run the RECTANGLES task, but now considering that the rectangles are: 6 windows, one text area (the long, slim rectangle), and 4 check boxes (the four small rectangles) in a Graphical User Interface (GUI). You should arrange them to be as nice looking as possible, but also easy to use as a GUI.

We have manually evaluated comparatively the arrangements of the shapes made by students in the two steps according to the following features, giving grades from 0 to 10 (the standard grades in schools and universities in Romania; 10 is the best):

- vertical symmetry;
- horizontal symmetry;
- repetitions (sequences) of shapes;
- figurativity (some students arranged the shapes in a figurative way see, for example, Figure 2);
- rhythm breaking.

The results were that vertical symmetry average grade changed from 6.647 to 6.76, horizontal symmetry from 3.441 to 4.2, sequencing from 4.529 to 9.821, figurativity from 7.5 to 5.5.

We see that the changes of averages due to considering that the shapes are parts of a graphical user interface were mainly on the considerable increase of sequencing, a moderate increase of horizontal symmetry and decrease of figurativity.

These results are not surprising. The decrease of figurativity shows probably that users prefer an interface more directed to usability, without being a figurative image. The increase of sequencing is also probably related to a need of a more ordered configuration of shapes, fact also recommended by the cognitive ergonomics grouping principle.



Figure 2. First step - a figurative image.

A comparison of these results with those of the first group of students is interesting because for the latter there was a single step. They were announced from the beginning that they should arrange the rectangles as for a graphical user interface:

The program displays a window with several rectangles. Your task is to move the rectangles in such a way that the whole arrangement **looks most pleasing to you, aesthetically**. You must arrange the rectangles so that all the rectangles are displayed and **no rectangle overlaps with another, and there is even a small space between them**. If needed, you may increase the size of the window.

You should also consider that **the rectangles are shapes in a Graphical User Interface (GUI)**: 6 windows, one text area (the long, slim rectangle), and 4 check boxes (the four small rectangles). You should arrange them to be as nice looking as possible, but also easy to use as a GUI.

The results were very similar to the second step of the fourth group (vertical symmetry average grade 6.75; sequencing 9.167), excepting horizontal symmetry, which had an average of 1.5.

CONCLUSIONS AND FUTURE WORK

The experiment presented in the paper is a first step in a series in which we want to analyze the relation between users' preferences on aesthetics and usability in GUIs. A clear difference was remarked between the cases when students were or were not told that the configuration they consider as aesthetical should be of a GUI. In the second case, for example, many of them rearranged the rectangles in a figurative configuration, which was no more the case in the first case, when they preferred a sequential organization to a figurative configuration. This result is consonance to cognitive ergonomics principle of grouping. As a consequence of the above considerations we may conclude that both our research questions have positive answers.

The results will be further analyzed, including more elaborated statistical tools and other evaluators. More experiments will be performed, including also other domains, for example texts and also automated tools [9, 11].



Figure 3. Second step of the same student from Figure 2 - a non figurative, interface-like shape.

ACKNOWLEDGMENTS

We would like to thank to one of the anonymous reviewers for the encouragement and the extensive very useful recommendations.

REFERENCES

- 1. Birkhoff, G. D. *Aesthetic Measure*. Harvard University Press, Cambridge, MA, USA, 1933.
- 2. Ghyka, M. Le nombre d'or. Rites et rythmes pythagoriciens dans le développement de la civilisation occidentale, Paris: Gallimard, 1931.
- 3. Ghyka, M. *The Geometry of Art and Life*, New York: Sheed and Ward, 1946.

- 4. Ghyka, M. *Esthétique des proportions dans la nature et dans les arts*, Paris: Gallimard, 1927.
- Servien, P. Les rythmes comme introduction physique à l'esthétique: nouvelles méthode d'analyse et leur application notamment à la musique, aux rythmes du français et aux mètres doriens – Boivin, Paris, 1930s.
- Vanderdonckt, J. and Gillo, X. Visual Techniques for Traditional and Multimedia Layouts. *Proc. Of AVI'94*, ACM Press, (1994), 95-104.
- Altaboli, A. and Lin, Y. Investigating effects of screen layout elements on interface and screen design aesthetics. *Adv. in Hum. Comp. Int.* (2011), 5:1–5:10.
- 8. Hamborg, K.C., Hülsmann, J. and Kaspar, K. The interplay between usability and aesthetics: More evidence for the "what is usable is beautiful" notion. *Adv. in Hum. Comp. Int.* (2014), 15:15–15:15.
- Zen, M. and Vanderdonckt, J. Automated evaluation of HCI metrics, *Proeedings. of IEEE RCIS*'2014, IEEE Press, (2014), 1-12.
- Buanga, P.M. Automated evaluation of graphical user interface metrics, Département D'ingénierie Informatique, Universite Catholique De Louvain, Ecole Polytechnique de Louvain, 2011.
- 11.Zain, J. M., Tey M. and Soon G. Y. Using Aesthetic Measurement Application (AMA) to Measure Aesthetics of Web Page Interfaces, *Fourth International Conference on Natural Computation*, Jinan, (2008), 96-100.
- Chek, D., Ngo, L., Samsudin, A. and Abdullah, R. Aesthetic Measures for Assessing Graphic Screens, *Journal of Information Science and Engineering* 16, (2000), 97-116
- 13. Miniukovich, A. and De Angeli, A. Visual impressions of mobile app interfaces. In *Proceedings* of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational (NordiCHI '14). ACM, New York, NY, USA, (2014), 31-40.
- 14. Dvorsky, G. 15 Uncanny Examples of the Golden Ratio in Nature, http://io9.gizmodo.com/5985588/15uncanny -examples-of-the-golden-ratio-in-nature, last accessed on 31 May 2016.
- 15. Balint, M. and Trausan-Matu, S. A critical comparison of rhythm in music and natural language, *Annals of the Academy of Romanian Scientists, Series on Science and Technology of Information*, 9(1), (2016), 43-60.

Enactment of User Interface Development Methods in Software Life Cycles

Iyad Khaddam, Hanaa Barakat, Jean Vanderdonckt

Louvain School of Management, Université catholique de Louvain

Place des Doyens, 1 - B-1348 Louvain-le-Neuve, Belgium {iyad.khaddam, hanaa.barakat, jean.vanderdonckt}@uclouvain.be

ABSTRACT

This paper presents UIDLC Manager, a software that provides user interface designers and developers with methodological guidance throughout user interface development life cycle. A methodologist firstly creates a dashboard model of the life cycle according to a corresponding meta-model in order to define a development path, decomposed into development tasks which structure the path into actions, and dependencies which serve as methodological milestones. A user interface designer or developer then enacts a previously defined development path by instantiating and interpreting a dashboard model while being provided with methodological guidance to conduct this development path. This guidance consists of steps, sub-steps, cheat sheets, and methodological actions. This approach is validated by applying it on nine classical user interface development life cycles, on two approaches for forward model-driven engineering of user interfaces based on a user interface description language, and on a linguistic approach for user interface software evolution.

Author Keywords

Dashboard model, dependency, development path, evolution, method enactment, method engineering, methodological guidance, model driven engineering, user interface description language, user interface development life cycle.

ACM Classification Keywords

• Software and its engineering ~ Software implementation planning • Software and its engineering ~ Software development methods • Software and its engineering ~ Software configuration management and version control systems • Human-centred computing ~ HCI theory, concepts and models.

INTRODUCTION

When User Interface (UI) stakeholders such as designers, modellers, analysts, graphical designers, and developers are involved in a UI development life cycle (UIDLC), they often ask what they need to do when and how, therefore complaining about the lack of methodological guidance [6, 9, 25]. This is particular applicable to Model-Driven Engineering (MDE), where the sequence of development steps should be rigorously followed in order to guarantee the quality of the results, as opposed to flexible [14], open UIDLCs where the process can be initiated from different starting points [19].

MDE of UIs [19] explicitly relies on a structured transformation process, namely involving Model-to-26

Model transformation (M2M), Model-to-Code compilation (M2C) or Model-to-Code interpretation (M2I). UIDLC stakeholders do not easily perceive when some degree of freedom exists to allow alternative choices in the process [7] and when some degree of determinism constraints these choices. MDE is often considered as a straightforward process, where little or no degree of freedom is offered, even when multiple development paths are possible [11].

Facing the multiplicity of models, such as task, domain, abstract UI, concrete UI, context of use, in a particular development path, stakeholders in general, the designer in particular, are rarely provided with some guidance on when and how to produce such models [20]. The proliferation of models may even be considered as a hindrance to conduct the UIDLC in a realistic way.

When a particular step in the UIDLC should be executed, designers do not easily identify which software should be used for this purpose, especially when different pieces of software could support the same step, partially or totally [7]. When a particular software is selected, they often feel lost in identifying the right actions to execute in order to achieve the step required in the UIDLC [10].

The multiplicity of development paths conducted among or within various organizations, in particular software development companies [3], increases the feeling that executing an unsupported UIDLC requires extensive training to become effective and efficient. Typical development paths occur along the following lines [11]: forward engineering, reverse engineering, lateral engineering, cross-cutting, round-trip engineering [1], beautification [28], etc.

Although several standardization efforts (e.g., the international standard for describing the method of selecting, implementing and monitoring the software development life cycle is ISO 1220707) and official organizations promote the usage of process models in order to increase the productivity of the development life cycle and the quality of the resulting software, they do not often rely on an explicit definition and usage of a method in these process models.

The above observations suggest that MDE seems more driven by the software intended to support it, less by the models involved, and even less by a method that is explicitly defined to help UI stakeholders. Therefore, a UIDLC could rest on three methodological pillars: *models* that capture the various UI abstractions required
to produce a UI, a *method* that defines a methodological approach in order to proceed and ensure an appropriate UIDLC, and a *software* support that explicitly supports applying the method.

For this purpose, the remainder of this paper is structured as follows: Section 2 presents a characterization of these three pillars in order to report on some initial pioneering work conducted in the area of UI method engineering with the particular emphasis of methodological support. Section 3 introduces the dashboard model as a mean to define a method that may consist of one or many development paths by defining its semantics and syntax. Section 4 describes how a method could be enacted, i.e. how a development path can now be applied for a particular UI project by interpreting the dashboard model. Section 5 provides a qualitative analysis of the potential benefits of using this dashboard model for method engineering in the UIDLC. Section 6 discusses some avenues of this work and presents some conclusion.

RELATED WORK

In general in computer science, a Software Development Life Cycle (SDLC) [29] is the structure imposed on the software development by a development method. Synonyms include software development and software process. Similarly, in the field of UI, a UI development life cycle (UIDLC) consists of the development path(s) defined by a UI development method in order to develop a UI (Figure 1). Representative examples of include: the Rational Unified Process (RUP) or the Microsoft Solution Framework (MSF). Each development path is recursively decomposed into a variety of development steps that take place during the development path. Each step uses one or several models (e.g., task, domain, and context) and may be supported by some software. All pieces of software, taken together support the development method.

For instance, the development path "Forward engineering" may be decomposed into a series of development steps: building a task model, building a domain model, building a context model, linking them, producing a UI model from these models, then generate code according to M2C. *Method engineering* [16, 21] is the field of defining such development methods so that a method is submitted to *method configuration* [17] when executed.

The meta-method Method for Method Configuration (MMC) [16] and the Computer-Aided Method Engineering (CAME) tool MC Sandbox have been developed to support method configuration. One integral part of the MMC is the method component construct as a way to achieve effective and efficient decomposition of a method into paths and paths into steps and sub-steps and explain the rationale that exist behind this decomposition. Method engineering has already been applied to various domains of computer science such as, but not limited to: information systems [17], collaborative applications [26], and complex systems [8, 13]. Typically, method engineering is based on a meta-model [18, 30] and could give rise to various adaptations, such as situational

method engineering [16] and method engineering coupled to activity theory [21].

In Human-Computer Interaction (HCI), we are not aware of any significant research and development on applying method engineering to the problem of engineering interactive systems, part from M2Flex [7], Sonata [14], and Symphony [13]. Several HCI development methods do exist and are well defined, such as a task-based development method [34], method-user-centred design [20], activity theory [21], but they are not expressed according to method engineering techniques, so they do not benefit from its potential advantages.



Figure 1. Structure of a UI development life cycle.

Probably the first one to address method engineering in HCI was the MIDAS (Managing Interface Design via Agendas/Scenarios) [25] environment. In this software, a methodologist was able to define a method by its different paths that could be followed and the steps required for achieving each path. MIDAS was able to show at any time when a method is executed, what are the different paths possible (e.g., design alternatives, criteria) by looking at design intentions stored in a library. MIDAS is tailored to the HUMANOÏD environment [25] and does not rely on a meta-model for defining a method and to execute. But it was a real methodological help.

User Interface Description Languages (UIDLs) [12] do not possess any methodological guidance based on method engineering because they mostly concentrate on the definition and the usage of their corresponding syntaxes and less on the definition of the method [3, 7].

TEALLACH [11] offers some method flexibility by enabling the designer to start from a task model, a domain model or a UI model and to then derive or link other elements related to each other. This flexibility is not method-oriented though. A more recent effort used Service Oriented Architecture (SOA) to define and enact a method [32], but there was no real software for achieving the method engineering. In conclusion, very few works exist on applying method engineering to HCI, but several existing work could benefit from it.

A DASHBOARD META-MODEL FOR A METHOD

To adhere to method engineering principles, a metamodel [18] is defined that addresses its methodological concepts as outlined in Figure 2. The dashboard is based on a meta-model that allows the description of development steps via their decomposition in Tasks, Resources required in Tasks and Dependencies between Tasks. This Dashboard meta-model has been expressed using Ecore/Eclipse Modelling Framework (EMF) and implemented in the MOSKitt environment [33]. The main

entities, i.e. Task, Resource, Dependency and Action, are structured as follows.



Figure 2. The meta-model for a methodological dashboard.

A **NamedElement** consists of a common ancestor for all metamodel elements. With the experience of the definition of several meta-models, we have found very useful to have a common ancestor element that all other elements in the meta-model inherit from. It simplifies several tasks in the following steps in the MDE approach, such as allowing to identify whether any given element belongs to this meta-model by checking its ancestry, and providing several properties we need in all elements, such as the 'name' property.

A **DashboardModel** represents a complete development path and at the same time is the root element of the metamodel. It holds the visual configuration to be used in the interpreter/enactment view.

A **Task** represents one development step of the development path. A Task is always bounded by Dependencies, except for the Tasks involving the first and last steps of the process. A Task can produce or consume zero or many Resources. As an ActionContainer, a Task can perform Actions on selected Resources.

A **Dependency** represents a milestone in the development path, which means that a series of development steps should be achieved before proceeding to the next development step. The Milestone is introduced as a straightforward mechanism for synchronizing different types of development steps, whatever their purpose is. Each Dependency is a step in the development path (Process) that forces the preceding Tasks to synchronize. A Dependency can require zero or more Resources from previous Tasks to be completed. As an ActionContainer, a

Dependency can perform one or more Actions on selected Resources.



Figure 3. The meta-model of UIDLC Manager.

A **Resource** consists of a (im-)material entity, produced or consumed by a Task or a Dependency of this development path (Process): model definition files to meta-model..

An Action represents an action to be performed by the user when enacting the process. An Action can range from launching a transformation to opening a cheatsheet to visiting a web page. An ActionContainer represents any element in the meta-model that can hold and perform Actions. A CustomAction represents a custom Action allows the methodologist to specify uncommon Actions with an external specification of the Action. A RunWizardAction expresses a specialized Action that runs the wizard specified by the hint parameter of the Action.

The UIDLC Manager is the software that implements the methodological dashboard whose meta-model is depicted in Figure 2. Figure 3 graphically depicts the meta-model of a project.

METHOD DEFINITION AND ENACTMENT

In order to define a UIDLC based on one or many UI development paths (e.g., simplified, enhanced forward engineering, forward engineering with loops) as defined in Figure 1, methodologist has to create one Dashboard model based on the meta-model outlined in Figure 2. A Dashboard model therefore represents the definition of a particular development path, but may also contain several

development paths in one model thanks to the concept of milestone. A milestone consists of synchronization points between tasks (e.g., development steps) involved in a development path and is attached to a *synchronization condition*. Such a condition governs the contribution of each task to the milestone (AND, OR, XOR, NOT, *n* iterations). Once the synchronization condition is satisfied, the milestone is considered to be achieved and the development path can proceed to the next development step.

Definition

Figure 4 depicts in Moskitt how a Dashboard model is created for the development path "Forward Engineering" that consists of the following development steps (that are represented as tasks to achieve to complete the development step) [32]:



Figure 4. The Dashboard model for the "Forward engineering" development path.

1. Create Task Model. This task is aimed at creating a task model that is compliant with the task meta-model, whatever the task meta-model would be. This task has three resources:

- 1. One and only one task model that will result from this task.
- 2. An optional document containing a documentation of the task modelled.
- 3. An optional set of task formal specifications.

A "task model definition guide" is a cheatsheet provided for giving methodological guidance on how to define a task model. Figure 4 details some potential development steps and sub-steps for this purpose in a cheatsheet. A *cheatsheet* is hereby referred to as a methodological panel that is provided from the methodologist to the method applier with any rules, heuristics, principles, algorithms, or guidelines that are helpful for achieving the associated task (here, creating a task model that is correct, complete, and consistent). An action "Generate Task Documentation" is added in order to specify a task model would ultimately result from it. The tool allows passing parameters to customize the generation.

2. Validate Task Model. Once the task model has been created, its validity with respect to its corresponding task meta-model is checked by means of Eclipse model checking techniques. Therefore, only one action is triggered: "Validate Task Model". Note that this task

serves as a milestone: the method applier cannot proceed with the next tasks if the synchronization condition is not satisfied.

3. Create Domain Model. This task is aimed at creating a domain model that is compliant with the task meta-model, whatever the task meta-model would be. It contains three resources, one cheatsheet and one action that are similar to those introduced for the task model.

4. Validate Domain Model. Once the domain model has been created, its validity with respect to its corresponding domain meta-model is checked by Eclipse model checking.

5. Link Task and Domain models. This task is aimed at establishing a link from the nodes of a task model to the appropriate nodes of a domain model thanks to the set of mappings accepted between these two models (e.g., a task observes a domain class, a task supports input/output of a set of attributes taken from different classes, a task triggers a method belonging to a class). Note that there is a dependency between this task and the two previous ones in order to ensure that the linking will be applied on two syntactically valid task and domain models.

6. *Milestone: start the Abstract UI generation.* When the task model has been linked to a domain model, we have all the elements in order to initiate a generation of an Abstract UI [15]. Again, this serves as a milestone.

7. Generate AUI. This task is aimed at (semi-) automatically generating an Abstract UI (AUI). For this purpose, an input resource "Task and domain models linked" (coming from the previous milestone) will result into an output resource "AUI model" by means of the action "Transform into AUI". This action is related to a set of transformation rules that are automatically applied to the input resource in order to obtain the output resource. Only one set of transformations is defined, but several alternative sets of transformation rules could be considered, thus leaving the control to the method applier by selecting at run-time which set to apply. Furthermore, this action is related to a transformation step (here, a M2M), but it could also be attached to an external algorithm that is programmed in a software. When all these alternatives coexist, a cheatsheet could be added to help the method applier in selecting an appropriate technique for ensuring this action (e.g., a transformation or an external algorithm) and parameters that are associated to this action.

8. *Milestone "AUI to CUI"*. This milestone serves as a synchronization point for initiating the next development step through the task required for this purpose [1].

9. Generate CUI. This task is similar to the "Generate AUI" except that a CUI is produced instead of an AUI, but with parameters that govern the CUI generation.

10. *Milestone "CUI to FUI"*. This milestone serves for initiating the last step and corresponds to a transformation [1].

11. Generate FUI. This task is aimed at transforming the CUI resulting from the previous task into code of the Final UI (FUI) by means of M2C transformation. Again, we may want to specify here that the transformation could be achieved by code generation or by interpretation of the CUI model produced. In the first case, a code generator is executed while a FUI interpreter renders the CUI into a FUI in the second case. Again, one default interpreter could be specified or the method applier can pick another one from a list of potential interpreters or rendering engines.

Enactment

Once one or several development paths of a UI development method have been defined in a dashboard model, the method can be *enacted* [3.6] by instantiating the dashboard model. This instantiation results into a runtime representation of the Dashboard (Figure 7) that depicts the progression of tasks already achieved, future and pending tasks, all with their associated resources. For instance, if a task requires to output resources to be created, this task will only be considered finished when the corresponding actions will have been able to produce the required resources. The method enactment is then under the responsibility of the person who is in charge of applying the method defined, e.g. an analyst, a designer. In the next section, we review potential benefits brought by the MDA approach under the light of this dashboard approach.

THE PRISM UIDLC

The prism UIDLC is different from common. It is based on a linguistic perspective to the development of the GUI. It mainly addresses the integration between HCI and Software Engineering in the development of a software product with usable UIs, with focus on the evolution of the software.

The linguistic perspective to the GUI development considers the interaction between the human and the machine as a communication text that is written differently than the human language, based on Nielsen's virtual protocol [27]. This GUI text is analysed linguistically to identify what is exchanged on each linguistic level; what are the semantics, syntactical rules, lexemes and alphabets used. It re-arranges GUI concepts on these linguistic levels and defines communication interfaces between them, in order to realize (refine with more details) concepts from upper levels on lower levels. More details on this linguistic perspective can be found in [22,23].

The linguistic perspective defines 6 levels for the development of the GUI. These levels are presented in the table 1, with description of concepts on each level, in addition to defining the communication interface between levels. The first level is "goal and task", which should be separated into two levels: "goal" and "task". But because task analysis cannot make this separation, we merge both levels into one. This merge is less confusing to the HCI community who is familiar with task analysis.

An example on the linguistic UI development

UI development from the linguistic perspective is iteratively refined. At first, we identify task input elements: input elements that modify a task state. A task can pass through several states like: created, offered, started, completed, suspended, destroyed, and erred. More on task modelling from this perspective (a linguistic task model) is in [23]. State transitions define required input elements on the UI, for the task.

			GUI	Communication	interface
World	Level	Artifacts	Key (Concepts	Realize from upper level	Define for lower level
	Goal &			-	-Define goals and real objects
	Task	Goals, Tasks		goals	-Define tasks and relations amongst -Define task input elements.
Conceptual	Semantic	Detailed functions: System, Input, Output	UI Elements	-Realize tasks by defining needed detailed functions.	-define input and output elements
	Syntax- time	Time containers: groups of Navigation elements	Navigation	-Realize distribution of UI elements on time by defining time containers.	-Define time containers -Define navigation elements.
	Syntax- space	Space containers: Placement rules.	Placement	-Realize placement of UI elements in time containers on the screen.	-Define space containers - Place UI elements on space containers
Perceptual	Widgets	GUI widgets		-concretize UI elements by mapping with appropriate widgets.	-Select appropriate concrete widgets for UI elements
Physical	Widgets Propertie s	Properties of GUI widgets	GUI Widgets	-Realize widgets attributes.	-set attributes of widgets -define nothing for lower level

 Table 1. The linguistic perspective to UI development:

 levels, concepts, activities and the communication interface.

Take the example of a GUI for registration to a conference. The end user needs to fill registration information and then pay the fees. Registration information include the user's personal information, registration type (regular, student or discounted fees), and additional information if exists, and billing information. The goal of the user from using the GUI is: Register for a conference. This goal is further refined at the task level by performing two tasks: "Fill registration information" and "Pay conference fees". The task level should identify task input elements, which are in this case, input elements each completes the related task. Figure 5 presents the "Finalize Order" task input element that completes the first task. On the *semantic* level (Figure 5 refines only the task "Fill registration information"), we refine tasks by defining necessary detailed functions to carry out these

tasks. Detailed functions define needed input and output elements for the performance of the task. They also define the interface to communicate with software modules.

On the *syntax-time*, we define a correct distribution on time (that respects environment constraints): when each group of UI elements will appear on the screen. We may have two styles: (1) Display UI elements at the same time. (2) Define navigation as in Figure 6.



Figure 5. Outcome on the semantic level. "Finalize Order" is a task input element.

The *syntax-space* (Figure 7) refines only the syntax-time style 1 for concision: place elements on screen= vertical placement, horizontal, or any other form of placement. Figure 7 depicts only the output of this level for the first time container: Personal Info. according to syntax-time style 2 in Figure 6.



Figure 6. Distribute on time containers and define navigation elements for parts of the GUI.

On the *widget level* (Figure 8), map UI elements to concrete GUI widgets. Finally, on widget Properties level (Figure 9): Setting properties of widgets to get the final GUI. On the final GUI, we note that every element is related to the level of abstraction that defines it. The "next step" button in Figure9 is defined at the syntax-time

level, while other input elements (text boxes) are defined at the semantic level. The reader can foresee that the button "Finalize Order" that is defined at the task level will appear on a screen at a later step (Figure 6). Please notice that upper levels may impose constraints on the choices on lower levels (like selection of widgets).



Figure 7. Placement of elements on the screen.

Prism-DLC

Prism is the development life cycle that aligns the UI development (from the linguistic perspective) and the software development. The main difference from other DLCs activities is the use of a classification step to analyse and classify UI requirements in order to determine the level(s) of enactment.

The linguistic perspective allows perceiving the UI since the analysis phase. This allows defining the term "UI requirements": *any modification(s) on the UI is based on a UI requirement*. UI requirements may issue from usability (like adapt to the user's culture), software design decisions (software modules and interaction capabilities), and software detailed design modules (like allow the user to interact with a function in the system). UI requirements impact on the UI might be decomposed on several levels, which is grouped in a UI batch.



Figure 8. Mapping UI elements with concrete GUI widgets.



Figure 9. Setting widgets properties and the final GUI.

A UI patch then is the impact of a UI requirement on different levels. Different development paths can be enacted in Prism. It can be used in a real software development to align UI and software developments, or to create a UI prototype to elicit usability requirements. In this paper, we explain the first development path as it is the more interesting one. Other development paths can be figured out by the reader.

Prism does not impose any constraint on the software DLC to integrate the UI development in. Anyway, in order to explain how integration and development is performed in Prism, we show integration with general development phases: analysis, design, detailed design, coding and testing. The Prism DLC is graphically depicted in Figure 10.

The software leads the UI development: This approach may be preferred by software engineers. The system development starts with the analysis activity of the system. When the analysis is completed, requirements pass through the classification phase to create the UI patch. This UI patch allows creating the task model from the software analysis.

The UI patch created after the software analysis may

impacts other linguistic levels. Note that software requirements are expressed at different levels of details. A user may express very detailed requirements like the preference for a specific theme of colours. The classification activity identifies UI-related aspects in every requirement and maps them to the appropriate linguistic level.

While developing the task level, usability shortcoming in analysis might be identified. The feedback loop from the task level to the analysis phase, not only ensures that usability requirements are gathered, also it assess consistency between the task model and the system analysis.

After the analysis is completed, the UI is fixed. This version of the UI might be communicated with

the user as a premature version of what is expressing in 32

requirements. Later modifications on the UI should not affect this version, which we call: the analysis-UI version. Modifications to this version should he communicated/approved with the user first.

The design phase starts with immediate feedback from the UI. As the task level is fixed, the design should implement each task appropriately: mapping to the domain model (which is part of the software design phase) and identify required UI elements. Design decisions (as UI requirements) are also linguistically classified to identify their impact on the UI. A UI patch is also created to express this impact. Note that the feedback loop from the semantic level to the design level is present to ensure that the UI and the design are consistent.

If the UI patch at the design phase contains implications on the task level, this means a shortcoming in the requirements. Tasks were not identified properly. If the software life cycle can handle such incompleteness, an alert can be triggered.

At the detailed design phase, the same repeat as with the design phase. After completing the detailed design, the semantic level is fixed. The UI for carrying out tasks is completely defined. No further modifications can be performed on the semantic level without repeating the design and detailed design phases. This version of the UI is called the *design-UI*.

In parallel with the implementation phase, the UI can be refined on the navigation, placement, widgets selection and stylistics on the last level. This gives the UI design the freedom to manipulate these aspects with the guarantee that any implemented design is compatible with the semantic and the task levels. Both activities are synchronized to start the testing activity.



Figure 10. The Prism DLC.

Testing can be decomposed into two activities (not depicted in Figure 10). Validation is to assess the implementation conforms to the specification (the design), and Verification to verify that the product satisfies user's requirements. Note that validation testing can be done on the design-UI version and verification can be done on the analysis-UI version. UI Validation testing is to compare the design-UI version with the final UI version on the navigation design, placement, widgets and stylistics. Functionality is guaranteed. Verification might be possibly enacted before fixing the design-UI.

EXAMPLES ON OTHER UIDLC

The purpose of this section is to demonstrate that the dashboard model is independent of any method, any meta-model and any User Interface Description Language (UIDL). It could be used for defining any UIDLC, any method that supports UIDLC (such as [4, 5, 24] to name a few), any meta-model of a model involved in such a UIDLC, and any UIDL (see [6, 19] for some representative examples). The only requirement is that each model should be explicitly linked to its corresponding meta-model in order to check its validity and conformity with respect to the meta-model as it is typically the case in MDE. Transformations gathered in transformation steps [1] should satisfy the same requirement, unless they are executed outside the Eclipse platform. The advantage of this approach is that all models and transformations between are defined by their corresponding meta-models in Eclipse, but forces to define them beforehand.

We evaluated UIDLC Manager on several UIDLCs. Table 1 contains the list of evaluated DLCs with a comparative analysis from the method engineering point of view. The table shows the number of development

steps in each method. the number of check points and the number of connections among development steps. These DLCs differ in the coverage of development phases an in the distribution of activities on each



development phase. In order to illustrate this difference, we project activities in each DLC on the generic development phases, defined as: Requirements Analysis (R), Design (D), Detailed Design (DD), Coding (C), Testing (T) and Maintenance (M). The result is shown in the right-most column in Table 1. Due to space constraints in this paper, we only illustrate the modeling of V-Cycle DLC using our tool in Fig 11. We also illustrate the projection of V-Cycle on the generic development phases in Figure 12. For the other UIDLC that have been realised with UIDLC Manager, the reader can visit the web-page: https://sites.google.com/site/user interfacedevelopmentcycles/uidlcmanager. Screenshots are given of respective UIDLCs that are typically found in HCI.

SDLC			žti	Dist	tribut	ion n	neası	ıres	
	Dev. steps	Check points	Connec ons						
Collin's	7	7	17	R	D	DD	С	Т	М
Circle [4]				2	5	3	1	1	0
Curtis &	24	24	61	А	D	DD	С	Т	М
Hefley's [5]				6	3	3	3	6	3
Nabla [24]	25	27	68	А	D	DD	С	Т	М
				6	5	2	2	8	1
O Cycle	6	6	14	А	D	DD	С	Т	М
[31]				1	2	0	1	1	1
Spiral	22	22	43	А	D	DD	С	Т	М
Model [2]				9	6?	6?	4	4	?
Star Model	6	6	15	A	D	DD	С	Т	М
[15]				2	1	0	1	1	0
V Cycle [8]	8	8	19	А	D	DD	С	Т	М
				1	2	1	1	3	1
Waterfall	8	8	22	А	D	DD	С	Т	М
[29]				1	1	1	1	1	1
Prism-	Max=			А	D	DD	С	Т	М
SDLC	26			10	8	9	5	1	1

Table 2. A comparison between different DLCs.

Figure 11. The V-Cycle using UIDLC Manager.

CONCLUSION

In this paper, we presented the dashboard model as a way to support the method engineering of a user interface development life cycle. For this purpose, we first defined what such a development life cycle is and how to structure it according to the principles of method engineering [3, 16, 17].



Figure. 12 Distribution of the V-Cycle activities on generic development phases.

This development life cycle is then expressed in terms of the following concepts: one or several development steps are defined in one single dashboard in order to create one development method, a development (sub-)step becomes a task to be achieved in the dashboard, the models involved in a development step become resources to be created and consumed by a task in the dashboard, the software required to manipulate these models become associated to resources via their associated file extension and/or from a list of potential software (e.g., model editor, model validator, model checker, transformation engine). The next step of this research will consider the forthcoming ISO 24744 standard on method engineering [2] that defines a set of concepts that support the definition and the enactment of a method based on welldefined concepts along with a graphical notation that combines structural aspects (e.g., how a task is decomposed into sub-tasks) and temporal aspects (e.g., how tasks are related to each other through dependencies and constraints).

REFERENCES

- Aquino, N., Vanderdonckt, J., and Pastor, O. Transformation Templates: Adding Flexibility to Model-Driven Engineering of User Interfaces. In *Proc. of 25th ACM Symposium on Applied Computing SAC'2010* (Sierre, March 22-26, 2010). ACM Press, New York, (2010), 1195–1202.
- 2. Boehm, B. A Spiral Model of Software Development and Enhancement. *IEEE Computer 21*, 5, (1988), 61-72.
- Brinkkemper, S. Method engineering: engineering of information systems development methods and tools. *Information and Software Technology 38*, 4, (1996), 275–280.
- Collins, D. Designing object-oriented user interfaces. The Benjamin/Cummings Publishing Company, Inc, Redwood City, 1995.
- 5. Curtis, B. and Hefley, B. A WIMP no more, the maturing of user interface engineering. *Interactions 1*, 1, (1994), 22–34.

- Céret, E., Dupuis-Chessa, S., Calvary, G., Front, A. and Rieu, D. A taxonomy of design methods process models. *Information and Software Technology* 55, 5, (2013), 795–821.
- Céret, E., Dupuy-Chessa, S. and Calvary, G. M2FLEX: A process metamodel for flexibility at runtime. In *Proc. of IEEE Conf. on Research Challenges in Information Systems RCIS'2013*. IEEE Press, Piscataway, (2013), 1–12.
- Forsberg, K. and Mooz, H. System engineering overview. In *Software Requirements Engineering*, R.H. Thayer and M.Dorfman (eds.), Second edition. IEEE Press, Piscataway, (1997), 44–72.
- Gonzalez-Perez, C. and Henderson-Sellers, B. A work product pool approach to methodology specification and enactment. *Journal of Systems and Software 81*, 8, (2008), 1288–1305.
- 10. Göransson, B., Gulliksen, J., and Boivie, I. The usability design process - integrating user-centered systems design in the software development process. *Software Process: Improvement and Practice* 8, 2, (2003), 111–131.
- 11. Griffiths, T., Barclay, P.J., Paton, N.W., McKirdy, J., Kennedy, J.B., Gray, P.D., Cooper, R., Goble, C.A., and Pinheiro da Silva, P. Teallach: a model-based user interface development environment for object databases. *Interacting with Computers* 14, 1, (2001), 31–68.
- 12. Guerrero García, J., González Calleros, J.M., Vanderdonckt, J., and Muñoz Arteaga, J. A theoretical survey of user interface description languages: preliminary results. In *Proc. of LA-Web/CLIHC'2009* (Merida, November 9-11, 2009). IEEE Press, Piscataway, (2009), 36–43.
- Godet-Bar, G., Rieu, D., Dupuy-Chessa, S., and Juras, D. Interactional objects: HCI concerns in the analysis phase of the Symphony Method. In *Proc. of ICEIS*'2007, 5, (2007), 37–44.
- 14. Godet-Bar, G., Dupuy-Chessa, S., and Rieu, D. Sonata: Flexible connections between interaction and business spaces. *Journal of Systems and Software 85*, 5, (2012), 1105–1118.
- 15. Hartson, H.R. and Hix, D. Towards empirically derived methodologies and tools for human-computer interface development. *International Journal of Man-Machine Studies* 31, 4, (1989), 477–494.
- Henderson-Sellers, B. and Ralyté, J. Situational method engineering: state-of-the-art review. *Journal* of Universal Computer Science 16, 3, (2010), 424– 478.
- Hug, Ch., Front, A., Rieu, D., and Henderson-Sellers, B. A method to build information systems engineering process metamodels. *Journal of Systems and Software* 82, 10, (2009), 1730–1742.
- 18. Jeusfeld, M.A., Jarke, M., and Mylopoulos, J. *Metamodeling for Method Engineering*. The MIT Press, New York, 2009.

- 19. Pérez-Medina, J.-L., Dupuy-Chessa, S., and Front, A. A Survey of Model Driven Engineering Tools for User Interface Design. In Proc. of Int. Workshop on Task Models and Diagrams for User Interface Design TAMODIA 2007. Lecture Notes in Computer Science, vol. 4849. Springer, Berlin, (2007), 84–97.
- 20. Karlsson, F. and Ågerfalkb, P.J. Method-User-Centred Method Configuration. In Proc. of Situational Requirements Engineering Processes SREP'2005 (Paris, August 29-30, 2005). Ralyté, J., Ågerfalk, P.J., Kraiem, N. (Eds.), 2005.
- 21. Karlsson, F. and Wistrand, K. February. Combining method engineering with activity theory: theoretical grounding of the method component concept. *European Journal of Information Systems 15*, 1, (2006), 82–90.
- 22. Khaddam, I., Mezhoudi, N., and Vanderdonckt, J. Towards a linguistic modeling of graphical user interfaces: eliciting modeling requirements. In Proc. of 3rd Int. Conf. on Control, Engineering & Information Technology CEIT'2015 (Tlemcen, 25-27 May 2015). IEEE Press, Piscataway, (2015), 1–7.
- 23. Khaddam, I., Mezhoudi, N., and Vanderdonckt, J. Towards Task-Based Linguistic Modeling for designing GUIs. In Proc. of 27ème conference francophone sur l'Interaction Homme-Machine IHM'2015 (Toulouse, 27-30 October 2015). ACM Press, NY, Article #17, 2015.
- 24. Kolski, C. A "call for answers" around the proposition of an HCI-enriched model. *Software Engineering Notes 3*, 23, (1998), 93–96.
- 25. Luo, P. A human-computer collaboration paradigm for bridging design conceptualization and implementation. In *Proc. of DSV-IS'94* (Carrara, June 8-10, 1994). Focus on Computer Graphics Series. Springer Vienna, (1994), 129–147.
- 26. Molina, A.I., Redondo, M.A. and Ortega, M. A methodological approach for user interface development of collaborative applications: A case study. *Science of Computer Programming* 74, 9, (2009), 754–776.

- 27. Nielsen, J.A. Virtual protocol model for computerhuman interaction. *International Journal of Man-Machine Studies* 24, 3, (1986), 301–312.
- 28. Pederiva, I., Vanderdonckt, J., España, S., Panach, I., and Pastor, O. The beautification process in modeldriven engineering of user interfaces. In Proc. of 11th IFIP TC 13 Int. Conf. on Human-Computer Interaction INTERACT'2007 (Rio de Janeiro, September 10-14, 2007). Lecture Notes in Computer Science, 4662. Springer, Berlin, (2007), 409–422.
- 29. Pressman, R.S. Software engineering, a practitioner's approach. Fifth Edition. McGraw Hill, 2001.
- 30. Sauer, S. Applying meta-modeling for the definition of model-driven development methods of advanced user interfaces. In Proc. of Model-Driven Development of Advanced User Interfaces MDDAUI'2007. H. Hussmann, G. Meixner, D. Zuehlke (Eds.). Studies in Computational Intelligence, 340, (2007), 67–86.
- 31. Scapin, D., Leulier, C., Vanderdonckt, J., Bastien, Ch., Farenc, Ch., Palanque, Ph., and Bastide, R. Towards automated testing of web usability guidelines. In Proc. of 6th Conf. on human factors & the web HFWeb'2000 (Austin, 19 June 2000). Ph. Kortum & E. Kudzinger (Eds.). University of Texas, Austin, 2000.
- 32. Sousa, K., Mendonça, H., and Vanderdonckt, J. Towards method engineering of model-driven user interface development. In *Proc. of Int. Workshop on Task Models and Diagrams for User Interface Design TAMODIA 2007.* Lecture Notes in Computer Science, 4849. Springer, Berlin, (2007), 112–125.
- 33. The Moskitt Environment, ProDevelop, Valencia, 2015. Accessible at http://www. moskitt.org/eng/proyecto-moskitt/
- 34. Wurdel, M., Sinnig, D., and Forbrig, P. Task-Based Development Methodology for Collaborative Environments. In *Proc. of EIS'2008*. Lecture Notes in Computer Science, vol. 5247. Springer, Berlin, (2008), 118–125.

Increasing the Accuracy of Indoor Localization Applications by Using Predefined Markers and the Phones Camera

Szabolcs Orban Computer Science Department Technical University of Cluj-Napoca Cluj-Napoca, Romania orbanszabolcs11@gmail.com

ABSTRACT

This article shows our approach on determining the global position of an Android device user, who wants to navigate in an interior of a building. The application is based on our previous application, which offers the possibility to navigate through a 3D model using the movement of a mobile phone, determined by its own sensors.

Previously the application used the navigations starting position given by the user himself. The method presented in this article adds to the accuracy of the existing application by using predefined points detected with the phone's camera. These points are placed in the building to determine the initial position or to correct the placement errors of the user. After the location is determined the application switches back to the navigation using the phone's sensors.

INTRODUCTION

In our everyday life we use our cellphones more and more as time passes. Therefore, it is a perfect platform to implement different applications to ease everyday activities or to get information faster.

In this article we propose a solution to a scenario where a mobile phone user is situated in the interior of a huge building with a complicated infrastructure for the first time and wants to reach a specific destination also located in the building. The ideal method would be getting the exact location of the user and navigating him to the desired destination using the mobile device without any other information input besides the destination from the user's part.

To implement an application of this kind it is advantageous to select one of the two main and widespread platforms, Android or IOS. Because Android is more accessible and the development and code signing is easier we choose this platform.

The most popular and easy localization method for a mobile application is the use of the GPS (Global Position System) functionality. It can specify the user's location from a wireless network or just from activating the GPS directly from the mobile device. The main problem with this method is that it's transmitted frequency is weakened by the building's walls when used inside and consumes a lot of battery [1].

Teodor Stefanut Computer Science Department Technical University of Cluj-Napoca Cluj-Napoca, Romania teodor.stefanut@cs.utcluj.ro

Another possibility to determine the user's positon is by using the RSS (Received Signal Strength) of the wireless signals of routers placed in the building. This method needs routers to be placed to cover all the areas of the building and the resulted location is not very accurate. [1]

In our previous article about this subject we did not take into consideration the global or initial location of the application user. [2] The main goal was to navigate through the building using the predefined position introduced initially by the user. The navigation started after the given starting and end points and consisted in mapping the shortest path with the help of the predefined 3D model of the building. The navigation itself used the device's sensors to get the direction and movement of the user. The direction was given by the gyroscope and the magnetometer and the movement was given by a step counter implemented from the accelerometer's data.

The advantage of this approach was the constant mapping of the user's position and movement in the interior of the building, but was heavily based on the initial global position introduced by the user itself. It had no correction or misplacement detection implemented and this way if the user did not know his exact position, he could use the application wrong.

So, finding the current positon of the user automatically or resynchronizing it represents the main flaw of the previous implementation. In this article some possible solutions are highlighted and the chosen one is detailed and implemented.

The proposed solutions make use of the mobile device's own camera for detecting and identifying key points or objects that can be used to determine the current location. One approach is a marker based method, which consists in placing predefined images of markers in the building. The application detects these markers and determines the location on the map. One other approach is feature based. Using a local database on the device, the camera scans the surrounding area and maps it in this database. This approach needs a predefined database of the key points of the building.

The chosen method uses predefined markers in addition with an augmented reality framework, to automatically detect the markers and display different objects on the scanned area that guide the user to the desired destination. Augmented reality in this application is used to present the direct view of the real-world environment with added augmented elements that are computer generated in real time.

RELATED WORK

Global localization methods

Global localization methods help define the user's location from where navigation methods can be used. These methods consist in detecting a predefined point in the environment and searching its correspondent location already mapped in a database. The methods differ on the way of defining these key points.

A localization method on a smaller surface is presented in the article [3]. The authors use a simultaneous localization and mapping algorithm to construct a real time discovering and tracking method for global localization. This real time tracking after an initial detection can easily adapt to different scaling transformations of the object.

For this method the initial object must be first scanned and mapped into a local database. The end user will use the mapping to re-localize the object as a starting point. This object can easily be detected from any angle or point of view.

As stated by the authors of the article [4] localization done only by one tracking method is not reliable every time. Consequently, they propose a solution that uses two localization methods and uses a switching algorithm to change the one in use depending on different factors.

The two algorithms used a feature based method, which uses already extracted knowledge points as prerequisite and a marker based method, which needs the device's camera to track for markers and to estimate de 3D position of the user.

The application uses a client-server model. All the computing, deciding and method switching is made on the server side. The client side must ensure to send continuous information about the camera's view, to extract potential key points or to detect markers.

The algorithm on the server side uses a fail-safe method to determine which algorithm to use. If the marker based method does not find a marker or the feature based method cannot extract key points it changes to the other one. In addition to the switching they also use sensor based localization and real-time mapping for improved stability of the system.

In our earlier article about localization and navigation [2] we presented a solution for navigating in the interior of a building using its 3D model and the mobile phones sensors. In order to start navigation, the user needed to enter the current location and desired destination. The application assumes that the user introduced his global position correctly and displays the 3D model of the interior of the building. On the 3D model a path is mapped dynamically formed by connection points. These

connection points are predefined and only those composing the shortest path will be displayed.



Figure 1. Sample screen of our existing navigation application.

The user can interact with the 3D model by moving the phone around himself and the model shows the corresponding view synchronized with the real world. This has been implemented based on the phone's gyroscope, which keeps track of the rotation movement of the device.

Using the phone's accelerometer, the user's traveled distance could be determined through the implementation of a step detector. Analyzing recorded data, the algorithm decides in real time if the user took a step and maps the movement on the 3D model.

In order to bind the distance travelled with the facing orientation, the device's geomagnetic field sensor was used. This sensor monitors the earth's magnetic field and thus behaves like a compass. Combined with the accelerometer, it allowed the mapping of the user's exact movement on the 3D model.

The main flaws of this method were the initial global localization method and the possible errors that may occur while detecting user's movements, which can cause the application to lose track. To adjust or correct these errors, the user had the possibility to manually synchronize by indicating a connection point forward or backward on the path. The approach presented in this article will aim at resolving these flaws.

Existing tools

There are several tools which can be used for detecting markers. One of them is Vrui VR toolkit [5], which was designed to implement a toolkit for scalable and portable applications. Its most notable strength is the focus on interactivity and immersive display environments.

Another example of an augmented reality toolkit is Meta Developer Kit [6]. It is produced by a Silicon Valley company that provides a 3D object projection using an augmented reality headset. These objects are able to lock onto areas where specific markers are placed in the real world.

A final example of an existing tool is Vuforia [7]. It provides SDK for Android, IOS and Unity, which is capable of recognizing frame markers, images, text or 3D objects of different forms, such as cylinder or box. Vuforia uses a natural feature tracking algorithm, which makes augmented reality practicable on low performance devices. The exact algorithm used by Vuforia is not public for the community, but it can be used and improved as part of the samples they provide. Because of its good performance and range of samples we decided to use Vuforia for our application.

RESEARCH ON THE MARKERS

Markers

Vuforia uses a variety of marker types to detect objects. The main type is a plain image, which can be detected from a fair distance, with different luminosity and from a wide angle. In addition, there are cylinder type or cuboid type markers. The difference between these and the simple image marker is the mapping of the key points on the desired object.

The used algorithm scans the input image from the source code and determines important key points specific to the image. Then, while scanning the input image from the camera, it tries to find these key points. Once found it keeps track of them until they get out from the frame or the camera cannot get a clear image of them.

The Vuforia platform also provides for developers the possibility to deposit the marker images on their cloud storage. This way the client application functions only if it has internet connection and access to the database, but the input images can easily be added, or edited without the application being modified.

Marker definition

Vuforia has defined a number of predefined frame markers that are present in its source code. Every one of these markers contains a unique pattern on its frame with a series of black and white squares, as seen in Figure 2.



Figure 2. Example of predefined frame makers.

The interior of the marker is transparent and thus allows the integration with any desired image, without any influence on the detection of the marker. For these predefined markers the Vuforia framework uses a more efficient and precise algorithm for detection. In addition, the application developer can define its own images for detection, which must be added to the online database by specifying the marker's type. After the upload the marker is given a rating, which shows the value of the detectable key points. If the rating is low the desired image has a poor detectability and needs to be changed. An image with a high rating must have a high amount of detail, to allow the extraction of a high number of key points.

We tried to define our own markers and to make different variations of them. After the creation we uploaded them to the online database for ranking and concluded the results. The simplest pattern to create was a chessboard like 8x8 pattern only with black and white colors shown in Figure 3. From this we varied the two colors in each square to generate different markers. For the tests we also used a same type of marker with different color variations.





Figure 3. Proposed markers.

After the upload and gathering the rating, the results showed that the markers needed more detail, because the rating was average. Consequently, we concluded that if we want to keep the pattern and to generate more key points, we need to separate the marker's squares into squares of smaller sizes. The key point extraction algorithm used by Vuforia works only on greyscale images, so using colors as an approach to add more details to the image is not has not been successful.

Taking into consideration the presented results, the final design of the marker is very similar to the predefined markers of Vuforia. The only difference is the transparent interior, which is an advantage because it can contain any desired image and the marker can easily be added to any surface without standing out. Because of this and because the predefined markers have maximum rating, we use these in our application.

Marker tolerance

The marker's tolerance to modifications depends on the number of extracted key points and in case of a user defined marker it must have a high rating. In case of the predefined frame markers the tolerance is as optimized as possible.

We tested the tolerance of the user defined markers from Figure 2. The considerations were: distance of detection, angle of vision and the luminosity. For the tests we used a 10x10cm image and changed the mobile device's

				Normal Lumin	osity	Low Luminosi	ty
				Angle of Visibility	/	Angle of Visibility	
Distance (cm)	Distance Measured (units)	Distance Calculated	Error (cm)	90 - 0	0 - 90	90 - 0	0 - 90
20	100	20.40816327	0.4081632653	80	80	60	45
30	155	31.63265306	1.632653061	80	80	60	40
40	203	41.42857143	1.428571429	80	60	60	30
50	243	49.59183673	0.4081632653	80	60	60	30
60	290	59.18367347	0.8163265306	65	50	50	30
70	341	69.59183673	0.4081632653	60	45	45	30
80	385	78.57142857	1.428571429	55	45	45	30
90	434	88.57142857	1.428571429	50	30	45	30
100	474	96.73469388	3.265306122	45	30	40	30

Table 1. Distance and angle of visibility measurement.

distance to the image by 10cm every time. For every section of 10cm the angle in which the image was detected and lost was also observed. The measures were repeated until the image could not be detected by the device's camera. All the measurements were repeated in strong and poor luminosity.

At that moment we did not know the connection or the scale between the width of the original image in centimeters and the unit of distance measurement returned by the application. The virtual distance measurement was made with the translation and rotation components of the detected image.

According to the measured results, monitored properties and the actual distance in centimeters increase in a linear fashion, so we could determine a constant which helped to transform the result into centimeters. This constant was determined in such a way to have the smallest error in centimeters. In our case measured value is 4.9. This value varies for each marker that has a different width from 10cm. The resulted measurements can be seen in Table 1 and the comparison of the actual and measured distance can be seen in Figure 4.



Figure 4. The Graphic of the actual distance and the calculated distance in cm.

Of course the camera of the device has a huge influence on the measurements. We used an 8MP camera, which is the most common on the mobile cellphones.

The tests show that the angle of visibility at smaller distances reaches about 80 degrees and at bigger distances about 40 degrees. While the angle of detection is smaller than the angle on which the applications loses

track of the marker, the application provides a continuous tracking of the detected key points.

The maximum distance at which the 8MP camera could detect the marker was about ten times the width of the image. This of course scales with the size of the marker. The luminosity did not affect the measures on a very significant scale.

PROPOSED METHOD

Our proposed method uses the Vuforia framework to detect precisely placed markers inside the building to determine the current position of the user. After the location is computed the application automatically adjusts the 3D model of the building to match the viewer's sight. Synchronization between virtual and real worlds is further performed with the help of the device's sensors, while continuously scanning for further markers. Every time a new marker is identified, the position of the user is resynchronized on the device.

For a more precise location computing we use a distance measurement from the detected marker, based on the analysis of marker's distortion. Furthermore, making use of the Vuforia capabilities, we display a 3D object on the top of the marker. The arrow will point in the direction of the next connection point.

Proposed markers

Because of the difficulty of the definition of custom markers with a high rating and the difficulty of detection we do not use any user defined markers. For a better tracking performance, we decided to use the predefined frame markers which are detected more efficiently.

For the detection to go naturally without the user's need to specifically position the device towards a marker, we propose to place all the markers on the ground. The natural holding of a mobile phone is 45 degrees to the ground's normal vector, so this placement is easily justified.

Because the maximum detected distance is ten times the width of the marker we use 20x20cm markers. The markers are placed on the ground so it is no need for a maximum distance greater than 2m.

Due to the contrast difference of the colors of the black and white markers with transparent background, which can be any image, have high tolerance to luminosity change, they are a good choice for using in the interior of a building.



Figure 5. Frame markers with different luminosity.

Distance Measurement

When the marker is detected in the camera's frame we have only unverified information of the user's position (inferred using phone's sensors), but we have the known location of the marker. Because of this it is necessary to compute the distance between the user and the marker, which will give us the exact position of the former.

The distance is computed using the distortion of the detected marker to the sample image. The result is a transformation matrix, from which the translation and rotation values can be extracted. By knowing also the exact width of the marker, the distance can be determined.

The transformation matrix obtained from the detected image has the following format:

The extracted translation matrix is the following:

1	0	0	Translation.x
0	1	0	Translation.y
0	0	1	Translation.z
lo	0	0	1

Finally, the resulted rotation matrices are as follows:

Around X axis:

[1	0	0	0
0	$\cos(\theta)$	$-\sin(\theta)$	0
0	$\sin(\theta)$	$\cos(\theta)$	0
Lo	0	0	1.

Around Y axis:

$\cos(\theta)$	0	$\sin(\theta)$	0]
0	1	0	0
$-\sin(\theta)$	0	$\cos(\theta)$	0
Lo	0	0	1

Around Z axis:

$\cos(\theta)$	$-\sin(\theta)$	0	0]
$\sin(\theta)$	$\cos(\theta)$	0	0
0	0	1	0
LΟ	0	0	1

The formula for computing the distance from the user to the marker is the following:

 $distance_{units} = (Translation.x)^{2} + (Translation.y)^{2} + (Translation.z)^{2}$ $distance_{cm} = distance_{units}/constant$

Knowing the distance from the marker and the transformation matrix, the current position of the user can easily be determined. In this way the displayed 3D model can also be synchronized.

Mapping Plan

As presented in previous sections all the markers will be placed on the ground. Every one of the markers will be near or on a connection point of the building so the user has a very low chance to miss any of them.

In the case of a large building there is a problem with the number of the markers. The Vuforia framework has only 200 predefined frame markers. In the mapping process it is inevitable to use all the 200 markers or to place them more than once.

Our proposed solution for this problem is to leave a small number of markers as boundary markers that separate several marker sets. Each separated set contains the remaining markers, which can be repeated in every separated set only once. These boundary markers will always be placed between inevitable connection points that are detected every time the user passes next to them. In addition, the detection algorithm keeps track of the detected marker sets and after every boundary marker it determines in which wing or section of the building the user is located.



Figure 6. Boundary marker distribution example.

Limits

The used methods and algorithms all have their limits. The detection algorithm has a maximum distance and an angle in which the marker is still detected. The distance limit is solved by placing all the markers on the ground with a 20x20cm size, so the clarity is not a problem. The number of markers used in mapping is also limited, but it is solved by using boundary markers.

What is more, the number of markers detected in one frame is also limited by Vuforia to 5. In our case, the placement and the ambiguity on which 5 markers to track is also not an issue because all the markers are placed on connection points of the building, so they are not present on the same frame very often.

MAIN FLOW

Compared to the previous flow, where the user had to specify his current location and his destination to navigate through a building, in this version the user needs to specify only the destination and to scan the first marker he sees on the floor to get the shortest path.

After the initialization, the 3D model of the building is displayed with the dynamically drawn path and connection points marked with spheres. As the user moves in the building his movement is also tracked on the 3D model.

Without the need to search and position the mobile device on a marker, it automatically starts the camera in the background and scans for markers. If a marker is found, the user can switch to camera mode and on top of the marker an arrow type 3D object is placed inside the scanned camera frame indicating the way to the destination.

If the sensor based navigation has an error and misplaces the user in the virtual environment, after every marker reached by the user the application automatically resynchronizes and corrects its misplacements.

CONCLUSION

This approach is an improvement of the previous version of the application, which used only the mobile device's sensors for navigation. The user had to know and introduce himself the starting point. In this method the application only needs to scan a marker on the ground to compute the user's position. This method has a more precise determination of global localization and does not need any additional input or correction by the user.

There is no need for the user to search for markers, because of their placements in key connection points of the halls. If there are any navigation errors, after a marker is detected, the errors are corrected, so the use of the application is more practical.

The next step of the research will be testing the performance of the application with the implemented marker detection on the field and to improve any downsides, lacks or failures that may occur and were not dealt with in the design phase.

As a further improvement we could remove all the 3D model mapping from the application and leave only the augmented reality representation. This consists in mapping the dynamically computed path directly on the scanned camera frame.

REFERENCES

- 1. Shala, U. and Rodriguez, A. Indoor Positioning using Sensor-fusion in Android Devices, *School of Health and Society, Department Computer Science*, Kristianstad University, Sweden, 2011.
- 2. Orban, S. and Stefanut, T. Indoor Localization and Navigation Using Phone Sensors and a 3D Model of the Building, 2015.
- 3. Martin, P., Marchand, E., Houlier, P., Marchal, I. Decoupled Mapping and Localization for Augmented Reality on a Mobile Phone, 2014.
- 4. Miyagi, A., Yoshihara, D., Kusui, K., Kimura, A. and Shibata, F. Mobile Augmentation Based on Switching Multiple Tracking Method, Graduate School of Information Science and Engineering, Ritsumeikan University, Japan, 2014.
- Vrui VR Toolkit, http://idav.ucdavis.edu/~okreylos/ ResDev/Vrui/
- 6. Meta Developer Kit, https://www.metavision.com/
- 7. Vuforia, http://www.vuforia.com/

Using WSN and Mobile Apps for Home and Office **Ambient Monitoring and Control**

Catalin Damian Faculty of Electrical Engineering, Gheorghe Asachi Technical University of Iasi Dimitrie Mangeron, No. 21-23 cdamian@tuiasi.ro

Lenuța Alboaie Faculty of Computer Science, Alexandru Ioan Cuza University of Iasi General Berthelot, No. 16 adria@info.uaic.ro

Adrian Iftene Faculty of Computer Science, Alexandru Ioan Cuza University of Iasi General Berthelot, No. 16 adiftene@info.uaic.ro

ABSTRACT

This paper presents a complex system (hardware and software) that is used to monitor and control some ambient parameters (temperature, humidity, light intensity, etc.). Monitoring parameters are temperature, humidity and light intensity. The system provides remote and automated control of local devices that can be used to assure the security or to increase the comfort indicator of the air.

Author Keywords

Sensors; air quality; monitoring; wireless communication; mobile applications.

ACM Classification Keywords

C.2.1. Computer-communication networks: Wireless communication.

General Terms

Experimentation.

INTRODUCTION

Wireless connectivity of more home appliances was a target for many years [10]. Temperature and light control for each room, status of the doors and windows and even monitoring of the domestic applications as washing machine, refrigerator and air conditioning provide many benefits to many home or office building owners [1]. Since the global trend is to build homes more economical in terms of energy, it eliminates the traditional heating, ventilation, lighting because the costs are very high. In new buildings such systems can be easily implemented at a low price given the low energy consumption of wireless sensors. One system based on standard 812.15.4 - ZigBee [9] can provide fire detection, security and access control as well as monitoring of environmental parameters. There are now security surveillance systems [2, 7], home automation [6], smart home systems [8] based on wireless sensors and ZigBee technologies. Some recent and important attention has been paid to Mobile Devices as terminal for users and some interesting applications have been developed [5, 9].

Even if there are over 10 years since the appearance of the wireless modules that allow networks of wireless sensors (Wireless Sensor Networks - WSN) the presented application tend to be innovative by extending mobility, surveillance and command being done through a dedicated application for mobile devices like a 42

smartphone or tablet. It shows a connection of new software and hardware technologies, not older than a few years, thereby achieving a solution of the future supervisory and safety living or working spaces.

SYSTEM ARCHITECTURE

The presented system (see Figure 1) has two main features: displaying the information measured in real time and the possibility to control remotely multiple actuators in order to maintain the optimum intervals for the monitored parameters. The optimum intervals are related to low and high values for each measured parameter. Recent technological solutions are used for both hardware and software elements. As communication standards used there are: ZigBee, RS232, web protocols, for saving data a MySQL database was used and as mobile devices operating system Android was chosen.



Figure 1. System architecture.

The values of monitored parameters are retrieved using wireless modules. These modules are arranged in space monitored at various points, without any restriction. They are powered by batteries that provide autonomy to several months. The information of all measurement nodes is centralized by another measurement node that acts as coordinator of the network. It allocates addresses to other modules and centralizes data. The data is sent from the coordinator node to a computer connected to the Internet using a USB-RS232 serial protocol. The data read from the measurement nodes is saved in a database that is located on a server. The data can then be accessed by mobile devices through a native Android applications, which in addition to displaying the current parameters also offers other functionality (text-to-speech, sending e-mail, command of physical devices, etc.).

HARDWARE RESOURCES

In order to obtain a physical system easy to design, program and that can allow quick re-dimension a series of modules which construction method is suitable for such applications were chosen. Being based on the ZigBee standard and containing all the elements necessary for measurement, analogue-to-digital conversion, acquisition and transmission of information, JN5148 family circuits are ideal for prototyping of wireless monitoring systems.

Wireless Modules

The presented system contains five measurement nodes but can be extended as required up to 256 nodes. All communication modules work on the same microcontroller (JN5148), small differences being due to attached peripherals [4]. From this point of view the system comprises a coordinator and four measurement nodes. Features of communication modules are:

- Compatibility with IEEE 802.15.4, ZigBee PRO JenNet and the 2.4GHz;
- Current consumption in standby mode (with timer active) 2.6µA;
- Up to 1km and 4km away from outdoor communication (depending on the antenna used);
- Current consumption in transmission: 15mA;
- Current consumption data reception: 17.5mA;
- Voltage: 2.3-3.6V.

Measurement Nodes

Figure 2 presents an image with a measurement node. This module is a stable platform that allows rapid application testing and development. Radio Frequency (RF) transceiver is on a smaller board which allows it to be moved on another board after programming. A measurement module has the following features:

- 2 configurable LEDs and an LED for checking power supply;
- 2 configurable push-buttons and 3 buttons with predefined functions (ON-OFF, RESET, PROGRAM);
- Temperature/Humidity sensor;
- Light sensor;
- Serial EEPROM Memory;
- Two UART interfaces (Universal Asynchronous Receiver / Transmitter) used for communication and for programming the circuit;

• A set of pins extension for addition of other sensors or actuators.

The Coordinator

This module is similar with the measurement node, with the addition of some components that allow easy interfacing with a human operator (Figure 3). These components are:

- 2 configurable LEDs (totally 4);
- 2 configurable push-buttons (totally 4);
- a 128x64 pixels LCD.

Sensors types

The proposed system is based on the use of three types of sensors: temperature, humidity and light intensity. The first two are integrated into a single physical circuit. The system can be improved by adding more types of sensors given the complexity of initial data acquisition interfaces.

Temperature and Humidity Sensor

Each monitoring module is equipped with a temperature and humidity sensor type SHT11 (Sensirion). This is a single-chip sensor for temperature and relative humidity. The conversion can be scheduled on 8, 12 or 14 bits, and the measurement range is between 0-100% humidity and 85° C and -40°C temperature.



Figure 2. Measurement node.



Figure 3. The WSN coordinator.

Light Intensity Sensor

The module contains a light intensity sensor type TSL2550. This is a digital sensor with a 2-wire SMBus serial interface. It combines two photodiodes and an AD converter into a single integrated circuit and can be used to determine a dynamic range of 12 bits for luminous intensity, which is very close to the human eye.

SOFTWARE TECHNOLOGIES

The software part of the system can be divided in four parts:

• wireless measurement nodes programming;

- interface between WSN and the database;
- server side services used for uploading, recording and providing measurement data;
- Android app used by user to interact with the system (this part will be presented in the next section).

Wireless measurement nodes programming

Eclipse IDE (Integrated Development Environment) with a Jennic SDK (Software Development Kit) was used to configure the wireless measurement nodes and the coordinator. The following steps are performed in order to establish the wireless communication protocol:

- *Initialization Protocol*: physical levels and 802.15.4 MAC protocol must be initialized in each device on the network;
- *Network coordinator initialization*: each network must have a fixed coordinator device. Initialization of the coordinator is the first step to take in the idea of building a WSN.
- Set PAN identification and address: immediately after initialization coordinator chooses a unique identifier for the network. This identifier can be fixed before or coordinator can automatically choose an ID by "listening" to other networks in the neighborhood so as not to be the same identifier in two different networks. Scanning networks is made possible in several frequency bands. It is usually preferred choice of frequency bands and then only scans for other networks available that band which is selected by the ID.
- Selecting the frequency band for transmission: coordinator must choose a frequency channel for communications within the network. This channel can be selected from a scan of Energy Detection. This scan determines which channel is "quietest" (has less traffic).
- *Starting the network*: the network is turned on. Once it is initialized devices are expected to connect to the network.
- *Recording modules on the network*: from the moment devices may require network connections. In order to connect to the network device must first be initialized after which it must find a coordinator. To find a coordinator, the device performs Active Channel Scan devices. This involves sending connection requests on all available frequency channels. When a coordinator receives such a request responds to signal the existence of a network channel.

Interface between WSN and the database

Wireless sensor network coordinator is connected to a computer using a PC USB serial interface. JN5148 wireless system contains a USB-RS232 interface cable used for both programming and circuits for data communication between PC and modules. In order to upload data from the WSN to a database on a server a Virtual Instrument (VI) was developed using LabView. The VI has the following tasks: read data from USB serial interface, transmission of data to the database by accessing a web page, reading the values for control 44

equipment from the database and sending control data through the USB when it is required to control the equipment.

The Front Panel of this VI is presented in Figure 4.

Server side services

An http server with php support and mysql database service was used to store the measurement and configuration data. The WSN side is connected to the database using an ODBC connection as LabView resource. When a client service needs to read measurement data or to write some data for system's actuators, php scripts are called. The php scripts are used for GET and SET operations. JavaScript Object Notation (JSON) was used to intermediate data between server side and the mobile App.

USER INTERFACE OF THE SYSTEM

A user interface was developed using official Android Studio IDE. The application was tested in a Genymotion emulated device and a Samsung Galaxy II tablet.



Figure 4. Front Panel of the VI used to communicate between WSN and a network PC.

The application is based on the following tasks (background or foreground):

- reading data from a database;
- display parameters based on the location selected;
- calculation and display some statistic values;
- configuring global settings;
- Text To Speech TTS function;
- send e-mails;
- setting the values for the actuators.

To achieve the above requirements, the application is based on complex programming technologies: Android fragments, native operating system TTS engine, e-mail sending, volley library for connection to the server, etc.

The Android application (Figure 5) is described regarding its functionality. Application screen is divided vertically into two parts; the left representing a menu that allows the selection of the right side content.





Figure 5. Mobile application's display types: a) main window, b) settings screen, c) a specific room screen.

The menu side is designed using ListAdapter and contains entries:

- *Home:* allows to select the main window of the application;
- Room1, Room2, ..., RoomN: a list of locations;

Settings: displays configuration snippet.

CONCLUSION

The paper presents a hardware system and a mobile application that can be used to monitor and control a series of parameters from a home or office building. The system uses five ZigBee modules (four measurement nodes and one coordinator) that send data to a server using a local PC. The data is displayed on a mobile device using a dedicated Android application [3].

REFERENCES

- 1. Damian, C., Lunca, E., Ilinca, M. Remote administration of hardware resources using TCP/IP Protocol and WEB technologies, *EPE*, (2014).
- 2. Hou J., Wu C., Yuan Z. and Tan J. Research of Intelligent Home Security Surveillance System Based on ZigBee, *IITAW* (2008), 554-557.
- 3. http://thor.info.uaic.ro/~homenv
- http://www.digchip.com/datasheets/parts/datasheet/ 1019/ JN5148-EK010_596.php
- 5. Jabbar A.A. and Kawitkar R.S. Implementation of Smart Home Control by Using Low Cost Arduino & Android Design. *International Journal of Advanced Research in Computer and Communication Engineering*, 5(2), (2016).
- Longheu A., Carchiolo V., Malgeri M. and Mangioni G. An Intelligent and Pervasive Surveillance System for Home Security. *INT J COMPUT COMMUN*, ISSN 1841-9836, 7(2), (2012), 312-324.
- 7. Obaid T., Rashed H., Abou-Elnour A., Rehan M., Saleh M.M. and Tarique M. ZigBee technology and its application in wireless home automation systems: a survey. *International Journal of Computer Networks* & *Communications (IJCNC)*, 6(4), (2014).
- 8. Saravanan R. and Vijayaraj A. Home Security Using Zigbee Technology, IRACST, 1(2), (2011).
- Yusuf A. and Baba M.A. Design and Implementation of a Home Automated System based on Arduino, Zigbee and Android Application. *International Journal of Computer Applications* (0975 – 8887), 97(9), (2014).
- 10. Zet, C., et.al. Advances in Remote Monitoring. 5th International Conference MTC (2007), 147-156.

Sphero - Multiplayer Augmented Game (SMAUG)

Marian-Nicolae Pînzariu "Alexandru Ioan Cuza" University, Faculty of Computer Science General Berthelot, No. 16 marian.pinzariu@info.uaic.ro

ABSTRACT

The current paper exposes the development of a game application that combines augmented reality, modern robotics, mobile features and multiplayer networking in order to create an immersive collaboration system, where people can interconnect to have fun and complete entertaining tasks together. SMAUG is a game based on Sphero 2.0 robotic ball, where multiple players share the same driving control of the robot. Depending on the game session and quest, a player can influence a certain moving direction or the robot's speed. Furthermore, the playground is augmented on the driving surface, so each player is looking to the game scene through his/her mobile's camera. The main objective here is that players need to collaborate in order to accomplish various tasks: collecting game artifacts, drive on augmented tracks, break down walls/gates, escape mazes and so on. In order to accomplish this objective the game is demanding collaboration: the game artifacts are distributed to different players augmented views, so they will need to communicate often in order to drive Sphero in the right direction. Finally, the most important objective was to open a new perspective on the Sphero-driven application development, creating at the same time the possibility of using mobile devices to conduct team-building activities.

Author Keywords

Augmented Reality; Robotics; Multiplayer Networking; Mobile Application.

ACM Classification Keywords

H.5.2. Information interfaces and presentation (e.g., HCI): User Interfaces. H.3.2. Information Storage and Retrieval: Information Storage.

General Terms

Human Factors; Design.

INTRODUCTION

Sphero 2.0

Designed as a toy with an educational purpose in mind, the Sphero 2.0 robotic ball [1] has a series of particular characteristics: instead of using wheels, rails or spider legs as many other robots, this one has the shape of a sphere with no extra parts. As a consequence, all the hardware and mobility components are encapsulated inside the sphere's shell. The shell and chassis are made out of polycarbonate, raising the robot's resistance to impact and providing waterproof properties. Adrian Iftene "Alexandru Ioan Cuza" University,

Faculty of Computer Science General Berthelot, No. 16 adiftene@info.uaic.ro

Hardware specifications

Beside the special mobility system, Sphero is equipped with an inductive charging base, a Bluetooth device (10 m range), 1 RGB and 1 blue LED and the following sensors: (i) Three axis gyroscope; (ii) Three axis accelerometer; (iii) Battery voltage sensor. All these mechanical and electronic modules provide the robot with capabilities to roll horizontally in any direction (0 to 360 degrees), maintain its stability, communicate over Bluetooth and glow in various RGB colors.

Software toolkit

From a software perspective Sphero implements primitive mobility operations (like rolling in a specific direction with a provided speed, etc.), approximate location tracking, collision detection and color switch. However, for more complex tasks, for example describing a 1 meter square, the robot's platform offers two possibilities: (1) Robot-interpreted languages; (2) SDK API commands.

Robot languages

An easy way to create small programs/routines that control the robot, and which can be used also by non-programmers, are using one of the three robot-interpreted languages offered by the Sphero suite [3]: *Macros, OrbBasic* and *Oval.* The development and deployment of programs written in one of these languages are managed through mobile applications that provide text/visual programming capabilities.

Sphero SDKs

Unlike the interpreted languages, which are run locally on the robot, the Sphero toolkit provides mobile SDKs that can be used to send remote commands to the robot, using the Bluetooth communication capabilities. In this scenario, the Sphero acts like a server, while the mobile application is a client. One disadvantage to this approach is the communication overhead, for which the protocol allows the client to send only around 12 commands per second.

At the moment, the Sphero developers can choose between three major mobile platforms to develop for: Android, IOS and Windows 8.1. Considering our experience with the Android SDK and the handy tools that support the development for this platform, we chose it as the deployment platform for the current application.

Adrian Iftene, Jean Vanderdonckt (Eds.)

STATE-OF-THE-ART

As it was already mentioned, SMAUG is a game that combines multiple technical fields and paradigms, but its main objective is to demand collaboration for completing entertaining activities. However, another important characteristic is also the use of AR, which can be found in a wide range of applications today. Following these traits, the following items will present some carefully selected applications that have a good resemblance to the present project.

Color Grab - Based on the Sphero's color glowing and acceleration detection features, Color Grab [4, 5] is a mobile multiplayer game with a simple and entertaining task. The game consists in a few rounds where each player needs to score by grabbing the robotic ball while it randomly glows a certain color.

The Rolling Dead - While the two presented applications compare to SMAUG on social interaction side, The Rolling Dead [6, 7] similarity resumes to the use of Augmented Reality. As such, the main character of the game is again Sphero, but this time the robot is placed into an augmented scene, where it is attacked by virtual zombies. The quest of the game is to defeat as many zombies as possible by using virtual abilities of the augmented Sphero.

Temple Treasure Hunt - This AR multiplayer game [8] classifies the players in two parties: Treasure Protectors (which are creating treasure trails, place various guardians and set challenges) and Treasure Hunters (which take up the challenges and explore the treasure trails by locating the treasure guardians in given time limit). The game can be played either indoor or outdoor (using geo-location map).

Ingress - This is an AR MMO location-based game, available on Android and IOS platforms. The gameplay is based on capturing virtual portals in culturally significant places (landmarks, public arts, monuments etc.), and link them to create virtual triangular "control fields" over geographical areas. Depending on which faction is the player joining (Enlightened or Resistance), the map areas will be marked with green or blue color [9].

APPLICATION DEVELOPMENT

In this section we'll discuss in details the steps involved in developing the Sphero driven application, and the main features that contribute to creating the immersive experience of the game.

Game workflow design

As presented in Figure 1, most of the setup phase is handled by the *Main Menu Scene*, where the first displayed panel has the same name. All available options from the menu are represented by the arrow labels from the image, each arrow representing a transition to another panel or scene.

Furthermore, a brief presentation of each panel/scene's logic would be the following:



Figure 1. SMAUG Application/Game Workflow.

- **Sphero Panel** offers the possibility to open a new connection to Sphero and to calibrate its orientation;
- **Sphero Server Activity** for connecting to Sphero, the application starts a separate Android activity which handles the connection and runs a tiny TCP Server in background;
- **Calibration Scene** this scene contains the necessary UI for calibrating the Sphero orientation and perform a quick drive test;
- Map Selection Panel allows players to navigate between available maps, select the one to play and proceed;
- **Multiplayer Panel** allows a player to choose between *Host* and *Guest* roles when starting a multiplayer session;
- Guest Connection Panel before connecting as guests, players are asked to enter the IP of the host server, which they do using this panel;
- **Team Setup Scene** after connection to the game, either as host or guest, the players will choose a role in the team (one or two driving directions);
- Map Scene this is a generic name for the scenes corresponding to each map, being the place where the actual play is happening.

Map/Level design

This is the most important part of the SMAUG game, as it is the place where the players will spend most of their time in the application. This part links all the pieces together in order to create the entertaining experience for the users. As exposed in Figure 2, every map consist of four main pieces: driving layer, UI, game logic and AR components.

The **Driving Layer** offers the necessary functionality for collecting direction and speed input from the players and transferring this parameters to the robot. In order to accomplish this task, there is a direct collaboration with

the UI, where each player interacts with a virtual joystick, and with the Network Manager.



Figure 2. Map/Level System's Architecture.

The **UI Layer** holds graphical controls that facilitate communication between the game and the users. Beside the virtual joystick that we already mentioned, the UI contains also components like status bars and various message/dialog boxes.

The third, but not last piece, is the **Game Logic**. The central business block of this layer is the *Multiplayer Level Manager*, which orchestrates all the activity that takes place within a game scenario (map). The **AR Layer** is a very important one as it provides the necessary means to detect the *Sphero playground* surface and augment the game objects over the video feed.



Figure 3. Main Menu UI.

USER EXPERIENCE

In order to display an attractive look and feel, as presented in Figure 3, the menu's UI provides a colorful 3D scene view, with a transparent panel and symbolic elements which support the game's story.

Game Stories – Diamonds Hunting

In this quest, the wizards sent the "magical Orb" to the "Valley of Stones", to find and collect as many "enchanted diamonds" as it can (see Figure 4), before they run out of energy and teleport Sphero back with the collected amount (this motivates the players to collaborate efficiently in order to return with the highest amount possible).



Figure 4. Collecting diamonds with Sphero.

Game Stories – Portals Maze

With a similar plot-line as the previous quest, this time the Orb needs to find only one artifact: "The Sword of Victory" (see Figure 5). By teleporting around and paying attention to their path, the players will finally arrive at the destination point, their goal being to obtain the shortest time possible.



Figure 5. Portals Maze - finding the Sword of Victory.

Game Stories – The Dragon's Treasure

This scenario is about using one player to distract the dragon, while the others are guiding Sphero to steal its treasure. Timing is an important aspect of this mission, as if the dragon comes close to the Orb the game would be over.



Figure 6. Sphero facing the furious dragon.

As it can be observed in Figure 6, the current version of the game exposes only the dragon and the treasure, while the diversion logic is yet to be implemented.

USABILITY TEST REPORT

"SMAUG makes a better world!" (Participant quote) On 23 June 2016 we invited four users to a specifically

arranged place and during the playing sessions, we've recorded their experience with the game while they were performing usual tasks. Recordings analysis clearly showed that:

- The UI looks very attractive, despite a few points where it's a bit confusing;
- The multiplayer feature presents minor performance and stability issues;
- Sphero driving went smoother than expected;
- Social interaction, visual effects, game dynamics and quest plots are the main selling points of the game.

Participants

We've recruited four participants for the present tests. Their selection was random, the group being formed of 2 women and 2 men. The occupations varied from teacher and accountant to student and QA engineer, while only one of them didn't have previous experience with multiplayer games. We also took into account their social interaction activity, having 2 participants with high activity, 1 with moderate and 1 with reduced.

Methodology

The conducted usability test consisted of an introduction, six tasks, a short interview and a post-test questionnaire. We instructed the participants to think out loud and express their thoughts during the test. After the task series that we communicated verbally to the participants, we gathered their assessment of the overall experience using the *QUIS scale*.

The tasks that users performed covered the operation of Sphero (connect, calibrate and drive) and playing the available quests in both, singleplayer and multiplayer modes.

Results

UX element	User1	User2	User3	User4	AVG
Menus	9	8	9	8	8.5
Selection controls	9	9	9	8	8.75
In-game messages	8	8	9	8	8.25
Game status bar	8	7	7	8	7.5
Sphero driving	8	9	9	8	8.5
Singleplayer	9	8	9	8	8.5
Multiplayer	7	7	6	8	7
Quest engagement	8	8	7	7	7.5
Overall look and feel	9	9	9	9	9
Team-building experience	8	7	7	7	7.25

Table 1: Post-test questionnaire results.

From our observation during the test sessions, the users had the best experience while performing the first four

tasks. They found the necessary menu actions very quickly and were able to successfully connect and drive Sphero around, as well as play any map in singleplayer mode. On the other hand, the participants had some trouble while trying to complete the last two tasks.

The main issue was that the game was not synchronizing well between devices, the users not being able to spot the objects from the augmented scene. Also the application seemed to work a bit slowly in this mode which indicates some performance issues that need to be addressed.

Another data point that confirms the above results comes from Table 1, which summarizes the user responses to the post-test questionnaire. The participants were asked to rate their experience with the elements from the left column, giving a note from 1 to 9, where 1 stands for confusing/frustrating experience, and 9 for clear/pleasant experience.

CONCLUSION

The paper presents a multiplayer augmented game called SMAUG, based on Sphero platform. The user can control the Sphero with an intelligent device, with augmented reality, via a lane, with game objects and a virtual joystick, which can be seen only on devices, by all players.

In the current state the application offers only a basic experience, but as the testing users agreed, there are plenty of areas that can be improved in order to transform SMAUG into a great game, such as: (1) Enabling players to compete each other in the arena, controlling one Sphero each; (2) Develop other game scenarios, with increasing complexity and difficulty; (3) Introduce player/team profiles, special abilities, rewarding system and ranking; (4) Connect with social networks, allowing players to share their best scores; (5) Introduce UI themes and rich audio content.

REFERENCES

- 1. Sphero 2.0 robotic ball: http://www.sphero.com/ sphero
- 2. Inside Sphero: http://www.howitworksdaily.com/wpcontent/uploads/2013/11/Inside-Sphero.jpg
- 3. Sphero Suite: http://sdk.sphero.com/sphero-robotbasics/robot-tech/
- 4. ColorGrab: https://play.google.com/store/apps/details? id=com.orbotix.colorgrab&hl=en
- 5. Sphero ColorGrab: http://technoitem.blogspot.ro/ 2014/08/childs-toy-robot-or-video-game-sphero.html
- 6. Rolling Dead: https://play.google.com/store/apps/ details ?id=com.orbotix.rollingdead&hl=en
- Sphero Releases The Rolling Dead: http://news. decision.com/orbotix/r/sphero-releases-the-rollingdead-augmented-reality-app-for-ios-andandroid,c9424861
- 8. Temple Treasure Hunt: http://www.indiedb.com/ games/temple-treasure-hunt-game/images/templetreasure-hunt-game-screenshots5#imagebox

Reducing Gestural Interaction Physical Stress in Virtual Environments: an Experiment

Sobhi Ahmed

Paragraphe - Université Paris 8 2 Rue de la Liberté, 93200 Saint-Denis, France subhigham@gmail.com Laure Leroy Paragraphe - Université Paris 8 2 Rue de la Liberté, 93200 Saint-Denis, France laure.leroy02@univ-paris8.fr

Ari Bouaniche

Paragraphe - Université Paris 8 2 Rue de la Liberté, 93200 Saint-Denis, France abouaniche@yahoo.com

ABSTRACT

Despite the advantages of gestural interactions, they involve several drawbacks. One major drawback is their negative physical impact. To reduce them, it is important to go through a process of assessing risk factors to determine the interactions' level of acceptability and comfort so as to make them more ergonomic and less tiring. We propose a method for assessing the risk factors of gestures based on the methods of posture assessment in the workplace and the instructions given by various standards. The goal is to improve interaction in virtual environments and make it less stressful and more effortless. We present our experiment which aims at validating our approach of evaluating gestural interactions.

Author Keywords

Gestural interaction; gesture assessment; musculoskeletal disorders; gestural stress.

ACM Classification Keywords

H.1.2 [Models and Principles]: User/Machine Systems -Human factors; H.5.2 [Information interfaces and presentation]: User Interfaces - Evaluation/methodology; Ergonomics; Interaction styles (e.g., commands, menus, forms, direct manipulation).

INTRODUCTION

One of the purposes of gestural interactions is to facilitate interaction with virtual environments. They aim at being intuitive, easier to use and learn, since lots of them are based on the emulation of natural gestures [22]. Some can fulfill specific needs (such as those of physically disabled people [15], etc.). These interactions are supposed to entail less cognitive and physical effort than 'traditional' interactions: for example, the use of a mouse, which demands a physical effort because of its distance from the user, calls for the user's arm to be outstretched while requiring a very accurate gesture when pointing [16].

However, musculoskeletal disorders associated with gestural interactions can be caused by movements requiring substantial physical effort. What is more, the extended and/or frequent use of such systems can result in an overuse of the muscles in charge of performing those gestures [25].

There exists stressful, tiring, illogical gestures and some might be impossible to perform for certain people (physically disabled people, for instance, but not only): the interaction with some gestured-controlled TV sets is considered stressful [10] because of the high position of the hand during use; interaction with touchscreens also affects user comfort negatively because of the need to keep one's arm outstretched [16]; the use of big screens is sometimes considered stressful to the neck because of frequent movements of the head and eyes [5].

Few studies have been conducted on how to reduce the physical impact of gestural interactions on the human body, and this sometimes resulted in the creation of nonergonomic, stressful gestures that are difficult to use [3]. Interaction with such systems can lead to various musculoskeletal injuries.

To the goal of analyzing and assessing the health risks associated with gestures, we have studied task assessment methods in the workplace. Just like gestural interactions, those work tasks consist of movements repeated frequently.

The physical impact of gestures is affected, for example, by the angle of the joint used in the gesture, the gesture's duration, its repetition, etc. The evaluation of such factors allows the assessment of gesture quality and consequently of their physical impact. This allows the design and implementation of ergonomic gestures that will cause neither pain nor stress, and that will be easier to use. We aim to implement a gesture assessment method based on certain criteria and factors stated in current studies.

In a first part, we present the medical problems related to gestures used in videogames and the workplace and the existing assessment methods of physical movements. The second part presents a synthesis and an analysis of these methods as well as our own approach. In a last part, we detail our experiment as well discuss the results we have obtained.

PRIOR WORK

As mentioned before (cf. Introduction), gestural interfaces are used more and more frequently in numerous domains. The use of such interfaces implies the performance of certain types of movements, sometimes repeatedly and/or for a long time, necessitating some effort. The overuse of the muscles in charge of these gestures can cause musculoskeletal disorders (MSDs).

"The term MSD groups some fifteen diseases acknowledged as work-related pathologies. These pathologies represent more than 70% of known workrelated pathologies." [1] MSDs affect the muscles, tendons and nerves of upper and lower limbs, at the level of wrists, shoulders, elbows or knees. A lot of MSDs have resulted from the frequent use of gestural interactions, such as those included with the Wii® gaming console [14].

Painful gestures

Painful gestures are often caused by being subjected to an external or internal force and by exceeding the standard angle range at which joints are normally used. Those outof-range angle values can be occasioned by numerous movements such as extension, flexion, abduction, adduction, pronation, etc. The movement range determines whether the joint is overly used and if the gestures resulting from the movement could be painful. Besides, static and dynamic constraints on some parts of the human body impact movement range and interdependence. [8, 25].

Injuries related to videogames based on gestural interactions

The repeated use of videogames can cause musculoskeletal injuries: for example, the use of the Wii® gaming console involves repeated physical movements that occasion sore muscles and knee, shoulder and heel injuries (DOMS: Delayed Onset Muscle Soreness) [25].

Videogame-related injuries can be classified in four categories: Tendinopathy, which means tendon injuries; Bursite, which means swelling and irritation of one or several bursa; Enthesitis, which refers to inflammation of the sites where tendons and ligaments are inserted into the bone and Epicondylitis (tennis elbow), which is a painful inflammation of the tendon on the outside of the elbow.

The main cause for such injuries and inflammations is the repeated stress undergone by involved muscles. According to the National Electronic Injury Surveillance System (NEISS), a high percentage of MSDs (67%) involve the use of Wii® in playing virtual sports [14].

Work-related injuries

The movements used during gesture interactions are extremely similar to those performed in the completion of some work-related tasks at the level of repetitions, extended time span, involved muscles, postures and the force exerted [20, 25]. These movements could occasion injuries called "Repetitive Strain Injuries" (RSIs). Several diseases have been associated with RSIs such as tendinitis, bursite, tenosynovitis, carpal tunnel syndrome, etc. [26]. Symptoms such as pain, discomfort and a sensation of localized fatigue of an overused joint can all point to RSIs.

According to [26], the risk factors associated with the onset of RSIs and their level of severity depend on time span, frequency and intensity, and have been classified in six categories: awkward postures, force, efforts and musculoskeletal load, static muscular work, exposure to certain physical stressors, repetition and the unvarying nature of the work and organizational factors.

Effort depends on the joints involved, movement direction, posture, type of grip and individual characteristics [1].

Risk factors can be decreased by adapting workstations and improving several elements: physical environment, task characteristics, technical aspects, individual, etc.

In gestural interactions, most gestures are deemed natural (natural user interface) [22], and require certain spatial movements, which in turn demand some effort as well as an internal or external force which can over-exert muscles and tendons affected by these activities [25]. What is more, these movements are repetitive, and occur over a long time span [3]. It is therefore possible to speculate that videogame- and work-related injuries are similar to those resulting from gestural interactions. It is rather clear that movements with extended arms, device vibrations and activities involving one's arm are very similar.

According to Nielsen [21] the basic principles of gesture ergonomics are: avoiding external positions, avoiding repetition, muscle rest, favoring neutral, relaxed positions, avoiding static positions, avoiding internal and external forces on joints and the interruption of the natural flow of bodily fluids.

GESTURE ASSESSMENT

Assessment methods

The reduction of the negative physical impact of gestures requires an assessment procedure. This procedure would allow determining the level of comfort and the stress they cause by measuring risk factors related to these movements. Assessment methods are classified in two categories:

Subjective methods:

Most studies on the assessment of the negative impact of gestures and physical movements generally resort to subjective methods [20, 21]. Amongst those can be found:

a. The Body Discomfort Diagram method (BDD), which assesses the level of discomfort in different parts of the body, using a diagram of the body and an assessment scale. The diagram allows identifying and assessing the places of discomfort by marking the affected areas [6].

b. Scoring methods, where a number of points is assigned to each single movement and criterion, resulting in a final score which determines the gesture's level of comfort. Each single score is decided either by the users [21] or by experts (ergonomists, etc.) [19].

c. Other methods are used, such as questionnaires [11], interviews, open-ended questions [20].

Other methods and angle measurements:

There exist methods and standards which allow assessing physical movements in a more objective way:

a. Electromyogram

The electromyogram can measure muscle activity through the detection and recording of electric signals sent by muscle motor cells used during activity. The electric signal is amplified and processed to determine the level of muscle force exerted. Electric activities vary according to the number of muscle motor units involved, which vary according to force. [18, 9]. This technique is used by [20] to measure muscle activity pertaining to the gestures and effort when interacting with touch-enabled devices.

b. RULA (Rapid Upper Limb Assessment)

RULA is a fast risk-factor assessment technique for upper limbs, geared towards individuals subjected to postures, forces and muscle loads potentially leading to MSDs [19]. The assessed factors for the selected tasks are: number of movements, static work, force, work posture, working time. RULA allows the observation of work posture, the identification of muscular stress and the attribution of a final assessment score for each posture ranging from 1 to 7. This score indicates the level of discomfort for the posture: the higher the score, the higher the risk.

c. The ISO 11226 standard

The ISO 11226 standard [12] aims at assessing health hazards for workers involved in manual labor. This standard defines comfortable and uncomfortable workrelated postures and allows their assessment. The assessment process involves specifying and classifying posture conditions as acceptable or not. These conditions comprise joint angle, time-related aspects and movement repetition. Each body part and joint is assessed separately. The assessment procedure is a one- or twostep process. The first step measures joint angles. If the angle doesn't exceed a given limit, the posture is deemed 'acceptable'. If not, the second step focuses on the time span of the movement.

d. The AFNOR NF EN 1005-4 standard (Safety of machinery – Human physical performance).

NF EN 1005-4 is an AFNOR standard [7] that defines a posture and movement assessment procedure related to working with machinery. The goal of this assessment is ensure machinery design matches to the recommendations aiming at avoiding postures and stressful movements leading to MSDs. The assessment can either be 'acceptable', 'acceptable under conditions' or 'unacceptable'. In situations determined as 'acceptable under conditions', other risk factors must be considered and additional measurements are needed. Factors which can be assessed are: duration, repetition, period of recovery, the presence of a support to the body, etc. The risk factors considered in this standard are: movement angle, gesture time, frequency, etc. This standard refers to other standards, amongst which ISO 11226, presented above.

Creating non-stressful gestures

There exist several approaches to creating gestural interfaces that take user preferences and needs into

account. Sometimes, predefined gestures are created, in which case the gesture vocabulary is derived by observing, collecting and assessing gestures done by some test subjects. Generally the assessment is done only by those subjects [21, 23, 27, 28]. On the other hand, in some methods, users are requested to create their own user-defined gestures. In this approach, the user defines the gestures they want to use in a preliminary step, before they start using the system [15]. Both approaches only take subjective assessment into account to define the gesture set. Our goal is to develop an automated, more objective method which could replace the latter subjective methods at the initial gesture definition step.

ANALYSIS AND COMPARISON OF APPROACHES, PROPOSITION

We aim to design an assessment method for gestures used during interaction that would minimize their negative physical impacts. This method is supposed to be more objective and allow automatic assessment of gestures, which can help designers to choose nonstressful gestures without the need to refer to subjective assessment every time. A complete gesture consists of a set of single gestures whose assessments result in an overall assessment of the gesture. Assessments of these gestures are done through the assessment of certain conditions and variables of the postures and physical movements effected. These conditions are: joint angles, posture duration, frequency, muscle load and external force. Variables will be assessed based on specifications for acceptable and unacceptable movements in various studies and standards [2, 7, 12, 19]. These specifications assess movement variables, thereby evaluating the quality of the gesture.

The data related to each joint is organized in tables specifying all possible movement types for that joint and giving 'acceptable' or 'unacceptable' values for the various criteria and variables of movement. The angle of movement is a key factor in the assessment process, since it indicates the level of joint stress and, consequently, the potential discomfort to which that stress could lead. The various levels of acceptability and comfort for shoulder movements (Figure 1) are shown in Table 1. In this table, the acceptability of postures and gestures is mainly determined by joint angles. What is more, gesture duration, movement frequency, and other factors potentially affecting the level of comfort are assessed, such as supports for the body, an even distribution of weight on both legs and feet, etc. Joint ranges are classified in 'acceptable', 'acceptable under conditions' or 'unacceptable' categories. The acceptability of movements is always connected to tasks with enough variation at the mental and physical levels [12]. Similar tables for each joint have been compiled and are not printed here because of space constraints. To understand the data shown in the table, we refer to the different evaluation strategies used by the aforementioned approaches:

• In RULA, acceptability is determined using a posture scoring system which works by adding points. Only the

final score can rate the gesture on an acceptability scale.

- In the AFNOR 1005-4 standard, each movement close to the limits of mobility is unacceptable if frequent (that is, if repeated as little as twice per minute). The assessment of gesture duration depends on ISO 11226 specifications.
- A static posture is a posture maintained for more than 4 seconds according to [12], more than a minute according to [19].



Figure 1. Shoulder movements [2], modified.

The measurement of time is crucial in the assessment of the acceptability of work postures: the longer the gesture and the higher number of repetitions, the more stressful the movement is. The different approaches use various strategies to measure time. Some measure movement frequency (repetition) [7, 19], others measure gesture duration [12], etc. In ISO 11226, the assessment of gesture duration is necessary when one gets a result that 'acceptable under conditions' (owing to the is movement's exceeding joint angle limits). In that case, time is of the essence in the assessment process. The standard comprises graphs which plot the relationship between joint angle range and the maximum acceptable gesture duration. According to these curves, the movement is deemed acceptable if it does not exceed the maximal time according to the joint angle [12, 13].

Some approaches use a scoring system based on an accumulation of points [19, 21]. In this case, the gestures' final score is determined by adding the scores of the single movements which the gesture consists of: the more stressful the gesture, the higher the score. Besides, other approaches depend on joint angle testing followed by gesture duration to determine its acceptability [12]. The information about the levels of acceptability of joint ranges, duration and other risk factors (such as repetition, force, muscle load, etc.) are collected and organized so as to be used in the assessment process. This process aims to determine the level of acceptability of the gesture according to the information collected.

Our approach uses Microsoft Kinect SDK to detect joint positions [24]. From these data, the system deduces joint angles. Duration and repetition of the movement are also calculated. Other variables, such as the presence of body supports, which can't be detected by Kinect, are entered manually by the evaluator. The application's output is a binary assessment (acceptable or not) of a body posture. The result obtained is considered 'acceptable' only if all gesture assessment results according to all standards and methods implemented within the system are 'acceptable'. It will however be deemed 'unacceptable' if any standard or method yields an 'unacceptable' result for the gesture. We have adopted such an approach to ensure a maximum level of safety. In addition, the system finds the average fatigue level of each joint according to each method. This can be used to compare joint fatigue levels over different tasks, as is done in our data analysis (see Results and discussion).

In the perspective of maximum safety, the software was designed to detect the maximum angles reached in the course of a gestural interaction, and to compute its assessments from these maximums, according to the standards and methods stated above. The measurements of time and frequency, used for the assessment of gestures following the AFNOR 1005-4 standard are only triggered when the threshold value (which requires time span to be taken into account) is exceeded (an angle of 20° for arm abduction, for example [7]).

METHOD AND EXPERIMENT

Our experiment aimed at validating our approach by evaluating the system's results. Such an evaluation was performed through a comparison between the system's evaluations and subjects' evaluations of their fatigue levels after performing some tasks using gestural interactions.

Participants

Twenty-six participants (aged 29 ± 10 years old), 17 males and 9 females, who all had beginner level with gestural interactions, were tested (we were aiming to test a gestural interface destined to the general public).

Tasks and Procedure

Participants were asked to perform —in a random order— two tasks in a virtual environment. One task was deemed "difficult" when the other was deemed "easy". The task was about arranging objects: the subject would pick an object from a stock box, and then move and drop it into the appropriate box. Subjects completed the task in three steps, using gestural interaction:

- 1. picking the object from the stock box (at a height of 90 cm), by pointing at it with their right hand and then closing it.
- 2. moving the object by moving their right hand towards the appropriate, illuminated box.
- 3. dropping the object in the appropriate box by opening their hand.

There were a total of six boxes; only three were visible in each task. In each task, a light indicated to subjects where to put their object. In each condition, the task was repeated 30 times in order to move 30 objects from the stock box to the other boxes. The number of times was chosen after pretests. Subjects were asked to return to a resting position between each task. The order in which tasks were performed was random. During the interaction, our system detected joint positions and angles, as well as evaluated them.

Movement	Source	Acceptable limit (1)	Acceptable under conditions – Not recommended (2)	Unacceptable limit (3)
	AFNOR	0° -20°	- 20° - 60° if static: (supported arm)	$- > 60^{\circ}$ if static
	1005-4		ou (short duration + recovery time).	- > 60° if
			- 20° - 60° if frequent.	frequent
			- 20° - 60° if:	
Antenulsion			- frequency <10 per min	
(Flexion- front)			- short duration	
(Tiexion front)			$- > 60^{\circ}$ II: - short duration.	
	INRS	0° -20°	20°-60°	-> 60°
	Tab Reg G	0 20	20 00	-> 60°
	RULA	20° (1 pt)	$20^{\circ} - 45^{\circ} (2 \text{ pts}) = 45^{\circ} - 90^{\circ} (3 \text{ pts})$	$->90^{\circ}$ (4 pts)
	AFNOR	0°	$> 0^\circ$ if : - not frequent	$\rightarrow 0^{\circ}$ if static
	1005-4	0	- short duration	$-> 0^{\circ}$ if frequent
Retropulsion	ISO 11226	0°		> 0°
(Extension- back)	INRS	0°		> 0°
	RULA	0° -20° (1 pt)	> 20° (2 pts)	
	AFNOR	0°	> 0° if : - not frequent	-> 0° if static
Adduction	1005-4		- short duration	$\rightarrow 0^{\circ}$ if frequent
	INRS	0°		> 0°
	AFNOR	0° - 20°	- 20°-60° if static: (supported arm)	$-> 60^{\circ}$ if static
	1005-4		or (short duration + recovery time)	$\rightarrow 60^{\circ}$ if frequent
			-20° - 60° if not frequent.	
			$-20^{\circ} - 60^{\circ}$ II:	
			- short duration	
Abduction			$- > 60^{\circ}$ if $-$ short duration.	
			- not frequent	
	ISO 11226	20°	$20^{\circ}-60^{\circ}$ (with support or check max	> 60°
			time)	
	INRS	20°	60°	> 60°
	RULA	stressful (1 pt)	
	AFNOR		stressful : if not frequent	stressful if
Flevated shoulder	1005-4			frequent
	ISO 11226			stressful
	RULA	stressful (1 pt)	1
Hyper adduction of				
the arm	ISO 11226			stressful
Extreme external				
Arm support		1 nt		
forward	NULA	- 1 pt		
forward				

Table 1. Recommendations for shoulder joint angles [2, 7, 12, 19].

The task was performed in two conditions, cf. Figure 2.

• Condition 1, "difficult": the task was supposed to be tiring; the levels of destination boxes were above shoulder level. Boxes' heights were respectively 160, 180, and 170 cm.

In this case, subjects had to raise their hand in order to move the object and drop it into the appropriate box. In addition, this condition requested more precision than the other. • Condition 2, "easy": the task was supposed to be easier; the level of boxes was around body center level. Their heights were respectively 85, 80, and 90 cm.

We began by introducing our work and the experiment steps orally. The subject started their first task in first condition, then filled a questionnaire about the fatigue level felt during a 10-minute break between the two conditions. Next, the subject performed the task in the second condition, which was followed by a second questionnaire about their levels of fatigue. Experiment duration was about 28 minutes, including 8 minutes in average for performing the tasks, and 20 minutes for both breaks.

Apparatus

We used two PCs running Windows 7 combined to two Microsoft Kinect for Xbox® motion sensors.

The first computer was used to run the test application comprising the gestural interface; a video projector was attached to this computer. Unity3D was used to develop our test interface. This computer was connected to the first Kinect which allowed users to manipulate the gestural interface. The Kinect was placed 130 cm to the left of the middle of the active zone (second Kinect). Its height was about 65 cm, its rotation angle was about 30°, cf. Figure 2.



Figure 2. Kinect locations and active zone.

The second computer was used to track user movements, process data, and display results. This computer ran C# code, developed in Microsoft Visual Studio 2013. This second computer was connected to the first Kinect. This Kinect was placed in the middle of the active zone, between the stock box and the first destination box, facing the active zone directly. Its height was about 70 cm.

The active zone was the zone in which the subject was allowed to move to manipulate the gestural interface. Within this area, subjects could be tracked by both Kinects. It was located in front of the second Kinect. Physical markers indicated its limits.

Kinect sensors were used to allow users to interact with the system in conditions as natural as possible. To that effect, we used a Wizard-of-Oz technique [4] to simulate picking and dropping objects. When the subject tried to pick an object (by closing their hand), the experimenter pressed a button, and when they tried to drop it (by opening the hand), the experimenter pressed another button. Subjects were unaware, only knowing that the opening / closing movements were responsible for picking and dropping objects. We used this technique to overcome the limitations of Kinect in detecting accurate movement of the wrist.

Data collection

The validation of our approach was done through analyzing and comparing two categories of measurements.

• Data collected by the system: the system detected positions, angles, duration and repetition for each gesture, and analyzed these data to evaluate the level of tiredness associated to the gestures, determining whether they were acceptable. Evaluation results for each joint were logged every 0.5 second, thus yielding results for ISO and RULA (assessments by the ISO11226 and AFNOR NF EN 1005-4 standards have been merged into a single test and result called ISO. Indeed, AFNOR 1005-4 uses the exact same angle measurements as ISO 11226 and adds to them a frequency factor.) Additional data was also collected by the software, such as information about the subject (name, date of birth, etc.), and information about the task (duration, order, etc.)

• Subjective data: Subjects filled in a questionnaire about their level of tiredness in each joint, the technical and cognitive difficulties they experienced, as well as their physical exercise capabilities. They could also add comments and remarks about the experiment. We used a six-point Likert-type scale for the subjective evaluation (0 for absence of fatigue and 5 the extreme fatigue) [17].

Data processing and statistical analysis

We studied whether means for the detected level of fatigue in each joint were different in the "easy" and "difficult" tasks, and whether the results given by the system were in accordance to those described by subjects. The data outputted by our system consisted of RULA and ISO assessment results. This data was compared to the subjective data for each joint. We calculated the average level of fatigue for every joint in each task, as well as the average subjective fatigue level for all subjects. We used a Wilcoxon test with a p-value equal to 0.05, whose null hypothesis was that there was no difference between the fatigue levels in both tasks. A Wilcoxon test was used because results did not follow a normal distribution.

RESULTS AND DISCUSSION

Table 2 shows the average levels of fatigue for the right shoulder, right wrist, and neck in both 'difficult' and 'easy' tasks, according to our various evaluation methods (subjective, RULA and ISO). For the shoulder and neck, the average level of fatigue in the 'difficult' task is higher than that in the 'easy' task. The difference between these two levels in both systems (ISO and RULA) and the subjective data is significant according to the Wilcoxon test. This means that our system's evaluation matches the subjective evaluation for shoulder and neck.

Tatat	Mahad	Mean (SD)	n voluo	
Joint	Method	Difficult Task	Easy Task	<i>p</i> -value
D. I.	Subjective	0.45 (0.24)	0.19 (0.12)	0.00003
Right Shoulder	ISO	0.77 (0.12)	0.34 (0.18)	0.00001
Shoulder	RULA	0.40 (0.08)	0.21 (0.03)	0.00001
	Subjective	0.11 (0.14)	0.04 (0.08)	0.00833
Neck	ISO	0.41 (0.3)	0.20 (0.28)	0.00022
	RULA	0.33 (0.16)	0.26 (0.13)	0.00322
Right Wrist	Subjective	0.15 (0.17)	0.15 (0.17)	0.71270
	ISO	-	-	-
	RULA	0.54 (0.1)	0.51 (0.09)	0.37400

Table 2. Average levels of fatigue for right shoulder, neck and wrist in both "difficult" and "easy" tasks given by subjects, as well as system data (ISO and RULA), p-value yielded by Wilcoxon test when comparing means for both tasks.

As for the wrist, according to RULA results, the average level of fatigue in the 'difficult' task is a little higher than that in the 'easy' task. The difference between them is not statistically significant according to the Wilcoxon test. Subjective results show that levels in both 'difficult' and 'easy' tasks are also almost the same.

Our objective was to verify whether the data produced by our system's evaluation matched subjective assessment in detecting gestural fatigue. We found that the system's and the subjective data for fatigue levels in shoulder and neck were compatible. We think that our system can detect gestural fatigue for both those joints. The higher fatigue level given by the system for the right shoulder in the 'difficult' task is logical, because in this task, the subject raises their hand to a higher level than the shoulder, which is considered stressful according to ergonomic standards and methods. Additionally, for the neck, in the 'difficult' task the gaze level is higher than that in the 'easy' task, so the subject had to raise their head more than what was necessary in the 'easy' task. Time spent performing the 'difficult' task also affected stress levels: the average time for the 'difficult' task was 4.2 minutes versus 3.2 minutes for the 'easy' one.

For the wrist, subjective and system results showed that there was no significant difference between both tasks. The wrist was used almost in the same way in both tasks. We noticed that the fatigue level according to RULA results is much higher than the one yielded by subjective results. We think that this difference is due to various reasons, among which the inability of Kinect to detect the exact wrist position and some of its movements, such as rotation, thus yielding some inaccurate evaluation results. Another reason was the nature itself of wrist movements: such movements are often used in daily life; we therefore posit that subjects underestimated their fatigue. In addition, it was apparent to the experimenters that subjects enjoyed performing such a task of picking and dropping objects using a freehand gesture. We think that our system could do much better if it used a more accurate detection setup, such as a multi-Kinect system 56

and/or ART tracking for example. We are planning to use such a system in our future work. Our system was originally designed using Kinect devices for Kinect's portability and ease of use [29]. It can be implemented easily in workplaces or a laboratory environment without needing a complex and time-consuming setup. There is however a clear substantial tradeoff between the quick and easy setup of a readily usable system and the detection precision expected of its use.

Generally, we think that our system was able to detect fatigue levels in some joints and that these results were in accordance with subjects' evaluations, which means that this system is valid for evaluating these joints. Other joints will be studied later. We think that this method for assessing gestures represents a potentially valuable approach to detecting gestural fatigue. It performs better for some joints than others, depending on the accuracy of movement detection and whether specifications about acceptability levels for these joints are available.

CONCLUSION AND FUTURE WORK

In spite of undeniable advantages to gestural interactions, the latter still exhibit several weaknesses, amongst which their negative physical impact on the subject performing them. In order to reduce that impact, it is important to implement a risk-factor assessment procedure to determine the levels of acceptability and comfort of the suggested gestures. This will ensure that the interactions created are more ergonomic and less stressful.

We propose a semi-objective assessment method of gestural risk factors based on the assessment of workrelated tasks and the specifications found in certain standards. We have validated our approach for some joints with a conclusive experiment.

Our objective is to try to improve interaction in virtual, augmented and mixed environments, so as to make it easier and less detrimental to subjects.

As future work, we envision validating our approach for other joints and testing it with more complex tasks. We also plan to integrate other specifications for acceptability. In addition, we are thinking of integrating other factors to the evaluation process, such as accuracy and duration of task as well as psychological influences (pleasure, familiarity, etc.), and study their effects on the evaluation process. Furthermore, we are interested in using a more accurate motion detection system such as ART tracking and integrating additional sensors which could detect some other important variables to the evaluation process such as limb rotations.

REFERENCES

- Aptel, M., Cail, F. and Aublet-Cuvelier, A. Les troubles musculosquelettiques du membre supérieur (TMS-MS)- Guide pour les préventeurs. Musculoskeletal disorders of the upper limb (TMS-MS) - Guide for OSH, Brochure de 96 pages, INRS ED 975, France (2011).
- 2. Aptel M., Gerling, A. and Cail, F. Méthode de prévention des troubles musculosquelettiques du

membre supérieur et outils simples: *la revue Documents pour le médecin du travail*, INRS Dossier TC78 (DMT 83), France (2000).

- 3. Aimaiti, N. and Yan, X. *Gesture-based Interaction* and Implication for the Future, Master thesis, Umeå University (2011).
- Beringer, N. Evoking Gestures in SmartKom Design of the Graphical User Interface. In: gesture based communications in Human Computer Interaction. Lecture notes in computer science. 2915. (2002), 409-420.
- Bowman, D.A., Chen, J., Wingrave, C., Lucas, J., Ray, A., Polys, N., Li, Q., Haciahmetoglu, Y., Kim, J., Kim, S., Boehringer, R. and Ni, T. New Directions in 3D User Interfaces, *International Journal of Virtual Reality IJVR*, 5, 2, (2006), 3–14.
- Cameron, J.A. Assessing work-related body-part discomfort: Current strategies and a behaviorally oriented assessment tool. *International Journal of Industrial Ergonomics*, 18, (1996), 389 –398.
- CEN (European Committee for Standardization). Evaluation of Working Posture in Relation to Machinery - prEN 1005-4. In: Safety of machinery – Human physical performance – Part 4: Evaluation of working posture in relation to machinery (1998).
- 8. Charles Eaton, M.D. "*Electronic Textbook on Hand Surgery*", http://www.eatonhand.com/. (1997).
- 9. Freivalds, A. Biomechanics of the upper limbs: Mechanics, modeling, and musculoskeletal injuries. 1 ED CRC Press, (2004).
- 10. Freeman, W.T., Weissman, C.D. Television Control By Hand Gestures, *from IEEE Intl. Workshop on Automatic Face and Gesture Recognition*, (1995).
- 11. Ha, V., Inkpen, K.M., Mandryk, R.L. and Whalen, T. Direct Intentions: The Effects of Input Devices on Collaboration around a Tabletop Display. *Proc of the First IEEE International Workshop on Horizontal Interactive Human-Computer Systems*, (2006).
- International Standards Organization (ISO), ISO11226. Ergonomics — Evaluation of static working postures. (2000).
- International Standards Organization (ISO). Technical corrigendum 1 to standard ISO 11226:2000- ISO 11226/AC1. (2006).
- Jones, C. and Hammig, B. Case report: injuries associated with interactive game consoles: preliminary data. *Phys Sports Med 37*, (2009).
- 15. Jégo, JF, Paljic, A and Fuchs, P. User-Defined Gestural Interaction: a Study on Gesture Memorization, *IEEE 3D User Interfaces*. (2013).
- 16. Lalumière, A. and Collinge, C. Revue de littérature et avis d'experts sur les troubles musculosquelettiques associés à la souris d'ordinateur», Rapport R-220, 74 pages, IRSST, (1999).

- 17. Likert, R. A. Technique for the measurement of attitudes. Archives of Psychology, 140 (1932), 1-55.
- 18. Long, C., Conrad, P.W., Hall, E.A. and Furler, S.L. Intrinsic-extrinsic muscle control of the hand in power grip and precision handling: an electromyographic study, *Journal of Bone and Joint Surgery*, 52, (1970), 853–867.
- 19. McAtamney, L. and Corlett, N.E. RULA: A Survey Method for The Investigation of Work-Related Upper Limb Disorders. *Applied Ergonomics*, 24, 2, (1993), 91-99.
- 20. Muse, L and Peres, SC. The Ergonomic Implications of Gesturing: Examining Single and Mixed Use with Appropriate Placement, on Texas Regional Human Factors & Ergonomics Society Symposium. (2011).
- 21. Nielsen, M, Störring, M, Moeslund, TB and Granum, E. A Procedure for Developing Intuitive and Ergonomic Gesture Interfaces for HCI, *In Proc. of the* 5th Internation Gesture Workshop, GW (2003).
- 22. Rauterberg, M. From gesture to action: Natural User Interfaces. Eindhoven University of technology; Mens-Machine Interactie, Diesrede (1999), 15-25.
- 23. Ruiz J., Li Y. and Lank E. User-defined motion gestures for mobile interaction, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Canada, (2011).
- 24. Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R., Kipman, A., and Blake, A. Real-time human pose recognition in parts from a single depth image. *IEEE CVPR*, (2011).
- 25. Sparks, D., Coughlin, L. and Chase, D. Did too much Wii cause your patient's injury?, *the Journal of family Practice*, 60, 7, (2011).
- 26. Simoneau, S., St-Vincent, M. and Chicoine, D. Les LATR: mieux les comprendre pour mieux les prévenir. RSI: better understanding for better prevention. IRSST et ASP métal-électrique, University of cubec, (1996).
- 27. Vatavu R. D. and Zaiti, I. A. Leap gestures for TV: insights from an elicitation study. In Proceedings of the 2014 ACM international conference on Interactive experiences for TV and online video (TVX '14). ACM, (2014), 131-138. DOI=http://dx.doi.org/10.1145/ 2602299.2602316 -27
- 28. Wobbrock, J. O., Morris M. R., Wilson A. D.: Userdefined gestures for surface computing, *Proceedings* of the ACM Conference on Human Factors in Computing Systems (CHI '09ACM Press, (2009), 1083-1092.
- 29.Zerpa, C., Lees, C., Patel, P. and Pryzsucha, E. The Use of Microsoft Kinect for Human Movement Analysis, *International Journal of Sports Science*, 5, 4, (2015), 120-127. doi: 10.5923/j.sports.20150504.0

Oculus Rift 3D Interaction and Nicotine Craving: Results from a Pilot Study

Ioana Monica Ciolan

Alexandru Ioan Cuza University Alexandru Ioan Cuza University Alexandru Ioan Cuza University of Iasi, Romania

Carol I, 11 - Iasi, Romania ioana.monica.ciolan@gmail.com Sabin C. Buraga

Ion Dafinoiu

of Iasi, Romania

of Iasi. Romania

Berthelot, 16 - Iasi, Romania busaco@info.uaic.ro

Carol I, 11 - Iasi, Romania dafinoiu@uaic.ro

ABSTRACT

Virtual reality is a new technology used in the treatment of nicotine addiction that involves immersing smokers in environments with specific stimuli in order to reduce the level of craving. The present study aims to present a new virtual reality technology (in this case, Oculus Rift) used to assess craving of Romanian smokers. Results showed that smokers feel a higher craving after the exposure to virtual reality environments with specific stimuli than neutral stimuli.

Author Keywords

Smoking Addiction; Nicotine Craving; Virtual Reality; User Testing; Interaction Experiment; Oculus Rift.

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms

Human Factors; Design.

INTRODUCTION

Nowadays technology is advancing fast, and the concept of virtual reality is no longer unknown for the general public. Most people have experienced 3D cinema. Also, vounger people are informed of all the devices which can be attached to smartphones or computers. Although initially virtual/augmented reality technology was created for recreation, today it is use in various fields: medicine (Szekely & Satava, 1999), education (Pribeanu, Balog & Iordache, 2016), psychology (Riva, 2003). Two important characteristics of virtual reality should be emphasized: immersion and presence. The immersion consists of all design and technological means of surrounding an individual or group of users with a virtual context (a tridimensional computer-generated space). The presence is a sensorial (visceral) reaction to a convincing immersive user experience (i.e. the human body reacts instinctively to the virtual environment as though it is the real one).

One of the issues that come to the attention of specialists is tobacco consumption which is currently responsible for the deaths of 5 million people (Mathers & Loncar, 2006) worldwide each year, and many of these deaths occur prematurely and at a young age. In addition, passive smokers are affected and there are approximately 600,000 people who die from exposure to passive smoking (Oberg et al., 2011). If the number of smokers continues to rise, at the end of this century there will be around one billion deaths caused by smoking, most of them in medium or underdeveloped countries, and half of these deaths will occur before the age of 70 years (www.who.int/tobacco/ global report/2013/en/). Even if there are a lot of smokers looking to quit smoking, only a small proportion have come to succeed without relapses.

In Romania, the prevalence of tobacco consumption is 27% according to the Eurobarometer from 2014 (ec.europa.eu/

public_opinion/archives/ebs/ebs_429_en.pdf), thus registering a decrease of 3% compared to 2012.

The Advantage of Romania compared to other European countries is that here is one of the highest percentages of people who never smoked, 60% (compared with the percentage of 54% across Europe). In contrast, Romania has one of the lowest percentages of people who have managed to quit smoking and continues to be abstinent, compared to other countries from Europe.

The present study aims to present a new virtual reality technology used to assess craving of Romanian smokers. In order to accomplish this goal, we designed various 3D environments to be explored by using Oculus Rift (Davis, Bryla, & Benton, 2015), a head-mounted display (HMD) which offers users a unique experience through the high level of immersion in virtual reality.

VR Applications in Psychology

Virtual reality (VR) is a concept that refers to artificial environments created by a computer system which can offer the user an almost real impression of physical presence.

The idea of virtual reality was first presented by Ivan Sutherland, in 1965: "make that (virtual) world in the window look real, sound real, feel real, and respond realistically to the viewer's actions". It has been a long time since then and a lot of research has been done in different fields. Another definition of virtual reality is offered by Cruz-Neira in 1993: "Virtual reality refers to immersive, interactive, multi-sensory, viewer-centered, three-dimensional computer generated environments and the combination of technologies required building these environments."

One of the recent promising devices in this area is Oculus Rift (Davis, Bryla, & Benton, 2015), a personal headset especially designed for people who are interested in the 3D virtual reality, in general, and videogames, in particular. Oculus Rift allows its user to feel like they are actually in a virtual environment just by wearing it. The screen displays two images adjacent to each other, one for the left eye and one image for the right eye. The combination of lenses is placed above the screen, enabling the zoom in-out and re-shaping the picture for both eyes, thereby creating a stereoscopic 3D image. These Rift devices monitor the wearer's head motions by the embedded sensor and accordingly adjust the image. The user's head movement is continuously analyzed and it is not necessary to use a mouse to control the direction of the view.

Using virtual reality in psychological studies

Besides the entertainment role, virtual reality has begun to play an important role in clinical psychology and it is used increasingly more often in treating phobias, eating behavior disorders, obesity, erectile dysfunction and posttraumatic stress disorder.

The main advantages of using virtual reality in clinical researches are: the potential of a precise control; the ability to adjust the virtual environments to the individual needs of the participants; the opportunity to expose the client to a particular environment that normally would not have access to, or would not be safe; providing increased confidentiality by replacing in vivo desensitization with virtual reality desensitization.

Until now, some researches were conducted in order to test the effectiveness of using virtual reality to treat various conditions: fear of height – Hodges *et al.* (1995), Emmelkamp *et al.* (2002); fear of flight – North, North, & Coble (1997); Rothbaum *et al.* (1996); Rothbaum *et al.* (2002); fear of driving – Wald, Taylor (2000); PTSD (Posttraumatic stress disorder) – Rothbaum, Hodges, Alarcon *et al.* (1999); eating disorders – Riva *et al.* (2001); sexual dysfunctions – Optale, Nasta, Marin, & Ipianon (2003); ADHD (Attention deficit hyperactivity disorder) – Rizzo *et al.* (2000); pain management – Hoffman *et al.* (2011); panic disorder and agoraphobia – Vincelli *et al.* (2003); social anxiety – Pertaub, Slater, & Barker (2001).

Virtual reality in nicotine dependence

Regarding smokers, there are already various studies who started to use virtual reality in order to develop strategies for nicotine cessation. They use cue exposure therapy, which involves a repeated exposure to stimuli which were previously associated with an addiction in order to extinct the conditioned response to those stimuli.

The argument for using exposure to stimuli to nicotine addiction is based on classical conditioning learning model. Nicotine is the unconditional stimulus and the effects of nicotine are the unconditioned responses. The conditions or contexts in which a person smokes very often become conditioned stimuli that determine the appearance of conditioned responses, and this leads to craving and nicotine consumption.

Craving is a state which is felt in all addictive behaviors and diseases, like drug use or gambling. This state can be regarded as a phenomenon, especially that many of the addicts invoke craving as a reason for encouraging addiction and as an excuse for treatment failure. Additionally, craving can be considered as a way to escape from painful reality (whether internal or external), so that the risk of relapse to addictive behaviors increase (Higley, Craney, Spadoni, 2011).

Some authors have identified a number of craving features (Nespor, Matanelli, Pekarkova, Gregor, 2011): (a) Craving is caused by specific stimuli and triggers that can be categorized into two main groups: internal (fatigue, pain, etc.) and external (for example, environment, straining situation, and others). Craving can be triggered also by negative stimuli (anger), as well as positive stimuli (joy). (b) Usually, craving is accompanied by subjective states as stress, tension and exhaustion. (c) When someone is experiencing craving, cognitive functions are altered, like the decision-making process. (d) When someone is experiencing craving, the ability of self-control is affected.

Although most cases, craving leads to maintaining the addictive behavior, it can also help to identify triggers, environments, specific stimuli that can contribute at the development of intervention strategies.

Various studies (Araujo, Oliveira, & Mansur, 2006; Araujo *et al.*, 2007; Cox, Tiffany, & Christen, 2001; Sayette *et al.*, 2000; Tiffany, Drobes, 1991) promotes the idea that the concept of craving is not just an intense desire to consume a substance, but it includes more elements like: the intention to satisfy a desire, the anticipation of the positive effects of substance consumption, avoiding negative symptoms of withdrawal.

Given that low levels of craving are considered to be predictors of long-term abstinence, the main goal of cue exposure therapy is to determine a decrease in the level of craving.

Several studies developed in this field (Baumann, Sayette, 2006; Bordnick, Carter, 2012; Rodrigues, Valverde, Maldonado, Ferrer-Garcia, Secades-Villa, 2012) concluded that virtual reality is able to recreate situations and everyday environments that are associated with smoking. These environments can be used in intervention programs that are based on cue exposure therapy.

Another interested direction of research is to use specific smoking cessation applications for mobile devices in order to self-track unhealthy behaviors or to unlearn smoking habits – for examples, see (Matei *et al.*, 2014). In this case, the presented information is only bidimensional, unless special devices like Google Cardboard are used.

DESIGN METHOD AND EXPERIMENTATIONS

Our project – currently available as a prototype – is using a truly immersive device to study the behavior of smokers in various 3D virtual environments.

The actual purpose is to test if the level of craving will increase when smokers will be exposed to a set of virtual environments (genuine 3D generated scenes) with specific stimuli *versus* virtual environments with neutral stimuli by using the Oculus Rift Development Kit 2 computing device.

To ensure that all participants in our study will have the same meaning of the concept of craving, we operationalized it using a simple item: "On a scale from 0 to 10, where 0 is no desire and 10 strong desire, how strong is your desire to smoke right now?"

Technological considerations

Obviously, having a 3D-enabled device is not enough. Specific software able to create various 3D interactive scenes – in our case, familiar controlled VR environments for the target users (especially smokers) such as home apartment, office, café, and club – was developed by using the Unity game engine (https://unity3d.com/).

For user modeling, the persona method was adopted by following an ontological approach (Negru & Buraga, 2012).

To access various data regarding the user behavior (e.g., movements, object selected from a given environment), a service-oriented architecture was implemented.

This approach is enabling the researcher to study the correlation between the need to smoke and the context the person is in. Secondly, the modular architecture is suitable for further processing of gathered data, including statistics, complex visualizations, data mining, etc.

User testing phase

After the prototype of the 3D environment was completed, we conducted a user testing controlled experiment on a number of 9 subjects (post-graduate students, age between 22 and 42 years, only two smokers) during the lab classes of the *Human-Computer Interaction* discipline (http://profs.info.uaic.ro/~busaco/teach/courses/hci/) in the second semester of the 2015/2016 academic year, Master of Software Systems Engineering, at the Faculty of Computer Science, Alexandru Ioan Cuza University of Iasi, Romania.

The purpose of this test phase was to ensure us that the virtual environments are perceived to be realistic and we used the feedback in order to improve the technological aspects. Each person had to explore a set of 5 interactive scenes (home apartment – alone, home apartment – party, office, café, crowded noisy club) in 10—15 minutes, with external guidance at request – see Figure 1.

All these virtual scenes contain neutral stimuli and stimuli associated with smoking. Depending on each virtual environment, specific objects were designed (TV, tables, couch) and some objects that are related with smoking such as lighters, ashtrays, and cigarette packs. Several scenes also include computer-animated characters (Figure 2).

After finishing the testing of the 5 virtual environments, each participant was asked to complete a questionnaire and answer a series of questions about this experience. Participants had the opportunity to explore each scene separately, to view those objects and even to interact with some of the specific stimuli.



Figure 1. User testing session of interacting with Oculus Rift.



Figure 2. Various interactions within the café 3D scene.

The most significant results regarding the assessment procedure of the user experience with the VR environments are available in Table 1.

Pilot study

The actual study consisted of the main phases depicted by Figure 3.

Participants

Daily smokers were invited to attend a testing session with the Oculus Rift virtual reality system.

In order to participate in this experimental session, smokers had to be 18 years old and had to be at least one-year smoker.

Items of the questionnaire	Result
Already interacted with Oculus Rift Feeling of presence	Yes – 2 persons No – 7 persons Other form of VR interaction (Google Cardboard) – 3 Average score: 6.67
(immersion, embodiment)	(on 1—10 scale)
Number of specific stimuli	Averagescore:7.44(on 1—10 scale)
Physiological side-effects	Vertigo – 2 persons Nausea – 2 persons; vomiting – 1 Fear – 2 persons Post-experiment reality perception difficulty – 1 Novelty-caused arousal – 1
VR design issues	 Difficult multimodal interaction (Rift + mouse) Non-realistic 3D graphics, physical proportions and collisions, depicted scenery Strange movements of computer-generated characters Break-in-presence (Steed et al., 2005) Too few stimuli for the actual participating smokers (2 of 9 subjects)

Table 1. Important user feedback concerning the VR design.

To recruit potential participants, we posted a message on a social network and those who were interested in this subject filled out a registration form with personal data.

A number of 48 people were scheduled for a testing session, but only 35 persons have successfully completed this task that lasted about 30 minutes.

Both men and women attended the testing session – from a demographic point of view, the average age was 30.2 years, all subjects were Romanian citizens and many of them were employed in a multinational IT company.

Procedure

The participants were informed at the beginning that they will test two virtual environments using Oculus Rift, because we want to observe differences in perceiving the virtual reality.

They had to smoke 30 minutes before starting the test. In the first stage, they were invited to become familiar with the VR device and they tested a virtual environment in the demo. After the adjustment phase, subjects had to complete a questionnaire – the Fagerstrom test – which measured their level of nicotine dependence.

It was found that most participants had an average level of nicotine addiction which means they are daily smokers, but the amount of nicotine consumed daily is not very high.





Subjects were randomly distributed into two experimental groups:

Group I – those who started with the VR environment with specific stimuli (a fully 3D modeled apartment – Figure 4) and continued with neutral stimuli environment (forest – Figure 5).



Figure 4. Specific stimuli 3D environment (as seen by the user interacting with the Oculus Rift device – left and right eyes).



Figure 5. Neutral stimuli 3D scene.

Group II – who started with the environment with neutral stimuli (forest) and continued with the environment with specific stimuli (apartment).

The apartment scene contains two rooms including a kitchen and a living room connected by a hallway. In each of these rooms there are some common items that can be found in any home: various pieces of furniture, TV, refrigerator, and in addition were placed a number of specific items which can be related to smoking – for example, packs of cigarettes, lighters, and ashtrays.

The forest scene was created exclusively with stimuli that are not directly related to smoking, especially to see if the perceived level of craving is different after the exposure to the apartment scene and the forest scene.

In order to simplify the pilot study, we chose that the other four designed environments will be used in further experiments.

The length of each exposure lasted between 5 and 7 minutes. Before and after each exposure, the subjects had to rate their level of craving using a visual analogue scale.

At the end of the exposure, the subjects had a debriefing phase, in which they were informed about the purpose of our study and had the possibility to tell how was that experience for them, because many of them have not tested Oculus Rift until then.

Results

We had 48 subjects that managed to complete all the phases from this process, but we had to exclude the data from

13 subjects due to initial confinement violation or due to the side effects of the exposure to virtual reality (e.g., nausea, headache).

Thus, results from 35 smokers entered in the statistical analysis. Of the subjects, 20 are men and 15 women.

To test whether there are differences between the initial level of craving and the level of craving after each exposure to virtual reality, we used ANOVA – analysis of variance (Bailey, 2008) method with repeated measurements and obtained an effect of F(2, 68) = 33.32, p < 0.01, which means that there are significant differences between the initial level of craving and the level of craving after exposure to the virtual apartment and after the virtual forest.

The Bonferroni Post Hoc Test (Bailey, 2008) showed that there are significant differences (p < 0.01) between the initial level of craving and the level of craving felt after the exposure to the virtual apartment, and between the level of craving felt after exposure to the virtual forest, meaning that subjects felt a higher levels of craving after the exposure to the virtual apartment environment (virtual environment with specific stimuli). No significant differences were found (p = 1.00) between the initial level of craving and the level of craving felt after the exposure to virtual forest environment. In addition, we conducted similar analysis to capture the effect of gender on the level of craving felt by subjects from exposure to virtual environments.

Results emphasize that there is no significant interaction between gender and the level of craving experienced after the exposure to the VR scenes: F(1, 33) = 0.04, p = 0.83.

CONCLUSION AND FURTHER DIRECTIONS

As we expected, smokers have felt a greater need to smoke when they were exposed to virtual environment with specific stimuli compared to the virtual environment with neutral stimuli. These results are congruent with those obtained in other studies (Baumann, Sayette, 2006; Bordnick, Carter, 2008; Acker, MAcKillop, 2013; Gamilto *et al.*, 2011; Lee *et al.*, 2003; Paris et al., 2011; Pericot—Valverde *et al.*, 2011; Traylor *et al.*, 2009; Steed *et al.*, 2016) which showed that exposure to virtual environments with specific stimuli increases the level of craving felt by smokers. These results can confirm that a cue-exposure process can be a useful strategy in smoking cessation.

Smokers must face every day different situations that either cannot or should not be avoided. The majority of smokers affirm that the biggest difficulty in the process of quitting smoking is linked to the inability to resist, even if they are aware of the negative effects of this behavior. That is why craving is an important element of quitting strategies. Through cue exposure therapy, smokers are exposed repeatedly to smoking-related cues in order to reduce the reactivity.

The argument for using exposure to stimuli to treat nicotine dependence is based on the classical conditioning learning model.

Thus, in the nicotine addiction, the nicotine is an unconditional stimulus and the effects of nicotine are the unconditioned responses. The conditions or contexts in which a person smokes very often become conditioned stimuli that determine the appearance of conditioned responses, and this leads to craving and nicotine consumption.

Given that low levels of craving are considered to be predictors of long-term abstinence (Shiffma, Ferguson & Gwaltney, 2006; Killen & Fortmann, 1997), the main goal of cue exposure therapy is to determine a decrease in the level of craving.

Regarding the limitations of this study, we can mention that we had a reduced number of participants, which does not allow us to extend the results obtained so far. It would have been useful to have participants at this study with different levels of nicotine addiction to see if the perceived level of craving after the exposure to specific stimuli would be different.

From the human-computer interaction point of view, we used in this pilot study a novel virtual reality device – Oculus Rift, and it has been proven that it can be used in our research field.
We aim to investigate more realistic VR interactions that contain specific stimuli for smokers, and then to implement an intervention program based on cue exposure therapy using virtual reality for smokers who want to quit.

ACKNOWLEDGMENTS

We thank *Răzvan-George Brezulianu*, *Tudor Carare* and *Alin-Ioan Lupu* for their invaluable support regarding the development of the system prototype. Also, we are grateful to the post-graduate students for their important feedback.

REFERENCES

- 1. Acker, J. and MacKillop, J. Behavioral Economic Analysis of Cue-Elicited Craving for Tobacco: A Virtual Reality Study. *Nicotine Tab. Res.* 15, (2013), 1409-1416.
- Araujo, R. B., Oliveira, M. S. and Mansur, M. A. Brazilian Validation of the Questionnaire of Smoking. *Cad Saude Publica.*, 22 (10), (2006), 2157-2167.
- Araujo, R. B., Oliveira, M. S., Moraes, J. F., Pedroso, R. S., Port, F. and Castro, M. G. Validation of the Brazilian Version of Questionnaire of Smoking Urges- Brief. *Rev Psiquiatr Clin.*, 34 (4), (2007), 166-175.
- 4. Bailey, R.A. *Design of Comparative Experiments*. Cambridge Press, 2008.
- Baumann, S. B. and Sayette, M. A. Smoking Cues in a Virtual World Provoke Craving in Cigarette Smokers. *Psychology of Addictive Behaviors* 20 (4), (2006), 484-489.
- Bordnick, P. S., *et al.* A Feasibility Study of Virtual Reality-Based Coping Skills Training for Nicotine Dependence. *Research on Social Work Practice*. 22 (3), (2012), 293-300.
- Cox, L. S., Tiffany, S. T. and Christen, A. G. Evaluation of the Brief Questionnaire of Smoking Urges (QSU brief) in Laboratory and Clinical Settings. *Nicotine Tob Res.* 3 (1), (2001), 7-16.
- 8. Cruz-Neira, C. Virtual Reality Overview. *SIGGRAPH'93 Course*. No. 23, 1993.
- 9. Davis, B. A., Bryla, K. and Benton, P. A. Oculus Rift in Action. Manning, 2015.
- Emmelkamp, P. M., *et al.* Virtual Reality Treatment versus Exposure in Vivo: A Comparative Evaluation in Acrophobia. *Behaviour Research and Therapy.* 40 (5), (2002), 509-516.
- 11.Gamito, P., et al. Virtual Reality Exposure on Nicotine Craving. Stud. Health Technol. Inform. 167, (2011), 63-68.
- 12. Garcia-Rodrigues, O., *et al.* Validation of Smoking Related Virtual Environments for Cue Exposure Therapy. *Addict Behav.* 37 (6), (2012), 703-708.
- Heatherton, T. F., Kozlowski, L. T., Frecker, R. C. and Fagerstrom, K. O. The Fagerstrom Test for Nicotine Dependence: a Revision of the Fagerstrom

Tolerance Questionnaire. *British Journal of Addiction*, 86, (1991), 1119-1127.

- 14. Hodges, L. F., *et al.* (1995), Virtual Environments for Treating the Fear of Heights, *IEEE Computer*, 28 (7), 27-34.
- 15. Hoffman, H. G., *et al.* (2011). Virtual Reality as an Adjunctive Non-Pharmaceutical Analgesic for Acute Burn Pain during Medical Procedures. *Ann Behav Med.* 41 (2), 183-191.
- 16. Higley, A. E., Crane, N. A. and Spadoni, A. D. Craving in Response to Stress Induction in a Human Laboratory Paradigm Predicts Treatment Outcome in Alcohol-Dependent Individuals. *Psychopharmacology* (*Berl*). 218 (1), (2011), 121-129.
- 17. Killen, J.D., Fortmann, S.P. Craving Is Associated with Smoking Relapse: Findings from Three Prospective Studies. *Exp Clin Psychopharmacol.* 5 (2), (1997), 137-142.
- Lee, J.-H., *et al.* Experimental Application of Virtual Reality for Nicotine Craving through Cue Exposure. *Cyberpsychol. Behav.* 3, (2003), 275-281.
- Mathers, C. D., Loncar, D. Projections of Global Mortality and Burden Disease from 2002 to 2030, *PloS Medicine*, 3 (11), 2006.
- 20. Matei, S. et al. Refreshing Quantification and other Ploys to Give Up the Habit. *Lecture Notes in Computer Science* 8512, Springer, (2014), 265-276.
- 21. Negru, S., Buraga, S. A Knowledge-Based Approach to the User-Centered Design Process. Communications in Computer and Information Science, Volume 415, Springer, (2012), 165-178.
- Nešpor, K., Matanelli, O., Pekárková, G. and Gregor, P. Classification of Anticraving Techniques, *Prakt. Lék.*, 91 (12), (2011), 703-706.
- 23. North, M. M., North, S. M. and Coble, J. R. Virtual Reality Therapy for Fear of Flying. *American Journal of Psychiatry*, 154 (1), (1997).
- 24. Oberg, M., et al. Worldwide Burden of Disease from Exposure to Second-Hand Smoke: A Retrospective Analysis of Data from 192 Countries. *The Lancet*, 377 (9760), (2011).
- 25. Optale, G., et al. Male Sexual Dysfunctions and Multimedia Immersion Therapy. CyberPsychology & Behavior, 6 (3), (2003), 289-294.
- 26. Paris, M., *et al.* Cue Reactivity in Virtual Reality: the Role of Context. *Addict. Behav.* 36, (2011), 696-699.
- 27. Pericot-Valverde *et al.* Evolution of Smoking Urge during Exposure through Virtual Reality. *Stud. Health Technol. Inform.* 167, (2011), 74-79.
- 28. Pertaub, D. P., Slater, M. and Barker, C. An Experiment on Fear of Public Speaking In Virtual Reality. *Studies in Health Technology and Informatics*, 81, (2001), 372-378.
- 29. Pribeanu, C., Balog, A. and Iordache, D. Measuring the Perceived Quality of an AR-based Learning

Application: a Multi-Dimensional Model. *Interactive Learning Environments*, (2016), 1-14.

- Riva, G. Virtual Environments in Clinical Psychology. Psychotherapy: Theory, Research, Practice, Training, 40 (1-2), (2003), 68-76.
- 31. Rizzo, A., *et al.* The Virtual Classroom: A Virtual Reality Environment for the Assessment and Rehabilitation of Attention Deficits. *Cyberpsychology* & *Behavior*, 3, (2000), 483–499.
- 32. Rothbaum B.O., *et al.* Virtual Reality Exposure Therapy in the Treatment of Fear of Flying: A Case Report. *Behaviour Research and Therapy*, 34 (5-6), (1996), 477-481.
- 33. Rothbaum, B.O., *et al.* Twelve-Month Follow-up of Virtual Reality and Standard Exposure Therapies for the Fear of Flying. *Journal of Consulting and Clinical Psychology*, 70 (2), (2002), 428-432.
- 34. Rothbaum, B.O., *et al.* Virtual Reality Exposure Therapy for PTSD Vietnam Veterans: A Case Study. *Journal of Traumatic Stress*, 12 (2), (1999), 263-271.
- 35. Sayette, M. A., Shiffman, S., Tiffany, S., Niaura, R., Martin, C.S. and Shadel, W. The Measurement of Drug Craving – Methodological Approaches to Craving Research. *Addiction*. 95, (2000), 189-210.
- 36. Shiffman, S., Ferguson, S. G. and Gwaltney, C. J. Immediate Hedonic Response to Smoking Lapses: Relationship to Smoking Relapse, and Effects of Nicotine Replacement Therapy. *Psychopharmacology* (*Berl*). 184 (3-4), (2006), 608-618.

- 37. Steed, A., et al. Breaks in Presence as Usability Criteria. HCI International 2005, the 11th International Conference on Human Computer Interaction, 2005.
- 38. Steed, A., et al. An 'In the Wild' Experiment on Presence and Embodiment using Consumer Virtual Reality Equipment. *IEEE Transactions on* Visualization and Computer Graphics, 22 (4), (2016), 1406-1414.
- 39. Szekely, G. and Satava, R. Virtual Reality in Medicine. *BMJ*, 319, (1999).
- 40. Sutherland, I. E. A Head-Mounted Three Dimensional Display. *AFIPS '68 Proceedings*. (1968), 757-764.
- 41. Tiffany, S. T. and Drobes, D. J. The Development and Initial Validation of a Questionnaire on Smoking Urges. *Br J Addict*. 86 (11), (1991), 1467-1476.
- 42. Traylor, A., et al. Using Virtual Reality to Assess Young Adult Smokers' Attention to Cues. Cyberpsychol. Behav. 12, (2009), 272-278.
- 43. Vincelli, F., *et al.* Experiential Cognitive Therapy in the Treatment of Panic Disorders with Agoraphobia: A Controlled Study. *CyberPsychology and Behavior*, 6 (3), (2003), 312-318.
- 44. Wald, J. and Taylor, S. Efficacy of Virtual Reality Exposure Therapy to Treat Driving Phobia: A Case report. *Journal of Behaviour Therapy and Experiential Psychiatry*, 31 (3-4), (2000), 249-257.

Brain Computer Interface using Machine Learning

Cristian-Valeriu Soare University Politehnica of Bucharest, Faculty of Automatic Control and Computers, Computer Science Department cristian.soare@cti.pub.ro

ABSTRACT

This paper presents the design and development of a complete hardware and software solution for a brain computer interface (BCI). It consists of a non-intrusive multiple channel data acquisition device which captures the electrical brain wave signals and passes the data to a computer. The computer then uses signal processing and machine learning algorithms to identify patterns in the signals received from the BCI. The goal of the device is to be a highly adaptable BCI, able to be used in a multitude of applications ranging from object recognition to basic control functions. Currently, the system is work in progress.

Author Keywords

Brain-computer interface; Data acquisition; Machine learning; EEG; Feature extraction

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous; I.2.m. Artificial intelligence: Miscellaneous

INTRODUCTION

Brain computer interfaces have been a subject of research for decades. Their primary use has been mind prosthetic control for patients with severed limbs, neurological diseases study and treatment (such as epilepsy) and the study of the brain itself. They are generally composed from an EEG (electroencephalography) system which records brain wave electrical signals by the means of electrodes either placed on the scalp, in a non-intrusive manner, or implanted in the brain, in an intrusive manner. Brain implanted computer interfaces have far greater accuracy when compared with non-intrusive techniques, and thus, they are mainly used in medical research and treatment. Lately, however, non-intrusive solutions have seen a rise in popularity as they can be easily used in various applications: basic mind-controlled devices such as drones, remote controlled (RC) cars as well as computer games. Therefore, nowadays there are commercially available BCI devices such as NeuroSky [2] or Emotiv [3] headsets which provide basic EEG data acquisition and offer programming SDKs for easy software development. This paper focuses on implementing such a device, able to handle a multitude of use cases. From the simple control of a computer, RC car or drone maneuvering to medical use and brain research, this BCI aims at providing the hardware and software necessary for a complete easy to configure and enhance system. One other notable non-invasive BCI system can be achieved by using functional magnetic resonance imaging (fMRI) scans of the brain. Through this technique a 3D representation of the blood pressure distribution within the brain can be obtained. In 2008 scientists at the Advanced Telecommunications Research (ATR) Computational Neuroscience Laboratories in Kyoto, Japan successfully reconstructed 10x10 black and white images from the brain using such a system.

BRAIN WAVES AND REGIONS

Brain waves are caused by neural oscillations within the brain. They have been studied by measuring the cumulative signals of large groups of neurons using EEG devices. Brain waves are classified according to their frequency and amplitude. Each frequency band is shown to correspond to a different mental activity and since each cortex has a distinct function, these signals can be more prominent in some parts of the brain when compared to others. Therefore, a time-frequency analysis of the signal is one of the first steps in decoding their information. The signal frequency bands are as follows [1]:

- Delta (less than 4 Hz) are high amplitude waves that in the frontal part of the brain. They are present during slow-wave sleep (dreamless NREM stage).
- Theta (4 7 Hz) are the waves most often present during states of drowsiness or idling thoughts.
- Alpha (8 15 Hz) waves are located in the posterior regions of the brain, on both sides. They are present during a relaxed state of the brain.
- Beta (16 31 Hz) are low amplitude waves located most prominently in the front, but also on both sides of the brain and manifest during active thinking.
- Gamma (more than 32 Hz) are located in the somatosensory cortex and are displayed during the perception of two different senses, such as sound and sight.
- Mu (8 12 Hz) are sensorimotor waves located in the sensorimotor cortex.



Figure 1: One second of typical EEG signal.

BCI HARDWARE DESIGN

Electroencephalography (EEG)

Electroencephalography devices consist in a number of electrodes placed on the scalp; each electrode providing a data channel for further processing. The most important features of such a device are the number of channels, the sampling rate and the sample resolution of each channel. Since data acquisition scenarios are expected to be reproducible to compare subject studies over time, international standards for scalp electrode placement have been defined. One of the most popular ones is the 10-20 system 3.1 which specifies a positioning such that the actual distances between adjacent electrodes are either 10% or 20% of the total front-back or right-left distance of the skull. There are also the 10-10 higher resolution system and a proposed 10-5 system.





Commercial solutions

Commercial BCI solutions have risen in popularity due to their ease of use and available SDKs. Such products are:

- NeuroSky [2] It has one electrode with a sampling rate of 512 Hz. Its frequency range is 3 to 100 Hz. It sends its data through an UART connection.
- Emotiv [3] a more powerful headset. It has 14 channels of 16 bits resolution and a sampling rate of 128 samples per second. It sends its data through a WiFi connection.
- OpenBCI [4] available since 2013, this is a DIY open source BCI of high performance. It uses an ADS1299 a 24 bit, 8 channel Analog to Digital Converter (ADC), reaching a maximum of 16 kHz sampling rate (distributed evenly on the number of used channels). The kit also comes with an open source signal monitoring and frequency analyzing software.

IMPLEMENTATION

Hardware Architecture

As EEG signals are of 10 to 100 microvolts in amplitude, they require amplification and filtering before reaching the digital acquisition device. For testing purposes, electrocardiogram (ECG) signals are better suited as they reach 1 to 5 millivolts and are easier to identify. 66 Therefore, each data acquisition board channel has amplification stages with configurable gains. One other issue encountered in bio potential data acquisition is the common mode signal. This parasitic wave forms are typically caused by power line electromagnetic interference. The common mode signal can be rejected using a driven-right-leg[6] circuit designed to feed the amplified and inverted signal back into the subject's leg, effectively cancelling the electric noise.

In this paper, we are using an ADS1258-ep 16-channel ADC with 24 bit of resolution and a maximum of 23.7 kHz sampling rate (distributed evenly on the number of used channels). It is also able to provide 125 ksamples per second in a fixed channel operating mode. The ADC can communicate its data through the standard SPI interface. For acquiring this data a RaspberryPi 2[12] is used, as it is a high performance embedded system capable of an Ethernet connection to send the data to a computer for further processing in real time. Ethernet was the preferred mean of data connection as it is high speed and eliminates the electric noise encountered in UART communication.

Each time a sample is acquired the ADC emits an interrupt on a dedicated pin and 4 bytes of data have to be read by means of SPI (3 bytes of channel data and 1 byte of channel details). If the train of data is not read within the 40 µs window between two sample conversions, the data might be corrupted. Although the RaspberryPi 2 itself is a performant system, the Linux scheduler adds high (milliseconds) interrupt latencies when trying to communicate with the ADC from a user-space written code. To counter this problem, interrupt handling and SPI communication has been achieved within a kernel module which buffers the data to the user-space by a character device interface. The Pi then forwards the data to the computer through an UDP communication. The interrupt response latency has been, therefore, reduced to an average of 4 µs (21 µs worst case scenario), as seen in Figure 3.



Figure 3. Interrupt response latency test - oscilloscope plots. Yellow plot represents the SPI SCLK signal and the blue plot represents the falling edge interrupt test signal. The second SPI data train is a captured worst case scenario interrupt response latency.

Even though it is not complete yet, the analog hardware component is aimed at obtaining microvolt level accuracy, while the digital one outclasses most of the available EEG sets. Compared with OpenBCI, which uses the ADS1299 (24 bit, 8 channels, 16 ksps), the used ADS1258-ep has 8 more channels, with a sampling rate of 23.7 kHz.

Software Architecture

Data acquisition

The data acquisition and analysis software provides basic plotting functionality (time, frequency and bar plots) as well as plugin interfaces for custom data source and data altering plugins. It is fully written in C++ and uses the Qt Gui toolkit [13] and the QCustomPlot [14] library for graphs. This highly modular framework handles all data transfers and thread management allowing the user to focus on the actual signal processing algorithms as one only needs to implement a few callbacks for passing data through the plugin system. As a simple proof of concept, the framework has been coupled with an audio streaming plugin and a Fast Fourier Transform (FFT) plugin which returns the frequency plot of the signal.

For the BCI application, the source plugin gets the data from the RaspberryPi through the UDP Ethernet communication protocol. The data is then passed through the FFT plugin and through the data interpretation plugins which are used for classifying the signals. Compared to other BCI software solutions, the configurable signal processing pipeline makes this framework more flexible and professional use ready. Its functionality is similar to the OpenVibe [5] software.

Algorithms

Once the data has been recorded, the signals have to be classified. For this task machine learning offers promising techniques for pattern matching and identification.

Convolutional Neural Network

Neural networks are mathematical models inspired by biological neural networks which are used to estimate or approximate functions that can depend on a large number of inputs. Their architecture is a system of interconnected neurons which exchange values between each other. These models are suitable for pattern matching since they can be trained to fit non-linear and arbitrary functions. Convolutional neural networks [7] use kernels mapped on the input to create feature maps on higher order layers. The weights between the kernel and every linked neuron in the upper layer are shared. This architecture allows identifying features in patterns in an offset-independent manner, a property of great use in signal and image classification. It is very popular for hand writing recognition tasks.

As seen in Figure 4, the first algorithm prototype consists in a convolutional neural network with 3 layers: the convolutional, subsampling and fully connected layer. Its input is formed by n spectrograms, where n is the number of signal channels. Although the architecture is in theory promising, because of the high dimensionality of the problem, the network may fail to converge if not provided enough training samples. Since the order of the input is around 105, it is hard to provide a similar number of training examples.



Figure 4. Convolutional neural network.

Feature extraction

Spatial filtering

To counter this problem, one can use a spatial filter to extract features from the signals and reduce the number of inputs in the classifier. The role of this filter is to output signals of high and low variances signals, according to the class of the original signal. Spatial filters can only be applied on binary classification problems.

To train the parameters of such a filter, the Common Spatial Pattern (CSP) algorithm is used.

Other features that may prove of use for the classifier are:

Power Spectral Intensity and Relative Intensity Ratio [8]. For a time series $[x_1, x_2, \ldots, x_N]$, and its Fast Fourier Transform result $[X_1, X_2, \ldots, X_N]$, a continuous frequency band from f_{low} to f_{high} is sliced into K bins. Boundaries of bins are specified by a vector $[f_1, f_2, \ldots, f_K]$ such that the lower and upper frequencies of the i^{th} bin are f_i and f_{i+1} . The Power Spectral Intensity (PSI) of the k^{th} bin is:

$$PSI_k = \sum_{i=|N(f_k/f_s)|}^{|N(f_{k+1}/f_s)|} |X_i|, \qquad k = 1, 2, \dots, K-1$$

where f_s is the sampling rate and N is the number of samples. Commonly used bins for EEG are δ (0.5 – 4Hz), θ (4 – 7Hz), α (8 – 12Hz), β (12 – 30Hz), γ (30 – 100Hz).

Relative Intensity Ratio (RIR) is defined as:

$$RIR_j = \frac{PSI_j}{\sum_{k=1}^{K-1} PSI_k}, \qquad j = 1, 2, \dots, K-1$$

Spectral entropy [8]. The spectral entropy is defined as:

$$H = -\frac{1}{\log(K)} \sum_{i=1}^{K} RIR_i \log RIR_i$$

EXPERIMENTS

Open datasets

In order to test the algorithms, openly available data sets can be used. PhysioNet [10] offers a bank of EEG and ECG signals in the standard European Data Format (EDF). It has a multitude of signal types from epilepsy seizure recordings to usual brain activity patterns. A particularly interesting dataset is the EEG motor-imagery set [11]. It contains data from 109 subjects who have been instructed to perform several tasks. Each subject opens and closes each of his fists or moves his feet when being signaled by visual cues on a screen. They also perform the same tasks but only imagining them.



Figure 5. Channel data.

The recordings have 64 channels of data with the electrodes positioned according to the 10-10 system, each with a sample rate of 180 Hz. They have been successfully classified using the MNE [15] Python library. After frequency and spatial (CSP) filtering, the signals have been classified using an SVM algorithm. The classification was 94% accurate. It was tested using cross-validation with 20% of the data used for the test set.

Future experiment

Once the equipment is complete, the following experiment will be performed. A microcontroller will be programmed to blink several LEDs at distinct frequencies (16-30 Hz). The subject will look at each of the LEDs while having his EEG activity monitored. After several sessions, the subject will picture the blinking LEDs in his mind while still being monitored. The goal of the experiment is to see if the visual LED frequencies can be easily found embedded in the signals picked up by the BCI. The pattern identification could prove to be a viable way for physically impaired people to control a computer. Although it is slow to control, the technique allows quick algorithm training thanks to the few features needed for the classification.

CONCLUSION

Although it is still a work in progress, the high precision of the hardware and modularity of the software already form the shape of a successful brain computer interface solution. The software framework is complete and the CSP based classification algorithm has proven reliable. The final goal of the project is to provide the tools a user needs for a ready to use BCI system and the freedom and configurability that would empower one to add new functionality. As for the improvements in the near future, once the system is completed, small computer or RC-car mind driven applications can be achieved.

ACKNOWLEDGMENTS

This project has been developed under the supervision of Prof. Adina Magda Florea and Dr. Traian Rebedea whom I thank for all their support and guidance. Special thanks to Eng. Ion Ciobanu for providing great assistance on the hardware component.

REFERENCES

- 1. Teplan, M. Fundamentals of EEG measurement. Measurement science review 2.2 (2002), 1-11.
- 2. NeuroSky headset official specifications http://neurosky.com/biosensors/eeg-sensor/
- 3. Emotiv Epoc headset official specifications http://emotiv.com/epoc-plus/
- 4. OpenBCI website http://openbci.com/
- 5. OpenVibe websie http://openvibe.inria.fr/
- 6. Winter, B.B. and Webster, J. G. Drivenright-leg circuit design. IEEE Transactions

on Biomedical Engineering, 1, (1983), 62-66.

- LeCun, Y. and Bengio, Y. Convolutional networks for images, speech, and time series. The handbook of brain theory and neural networks, 3361(10), (1995).
- Forrest Sheng, B., Liu, X. and Zhang, C. PyEEG: an open source python module for EEG/MEG feature extraction. Computational intelligence and neuroscience 2011, (2011).
- Xiaoou, L. et al. Classification of EEG signals using a multiple kernel learning support vector machine. Sensors 14.7, (2014), 12784-12802.
- 10. Physio Toolkit Physio Bank. Physionet: Components of a new research resource for complex physiologic signals. Circulation, 101(23), (2000), 215–220.
- Schalk, G., McFarland, G. J., Hinterberger, T., Birbaumer, N. and Wolpaw, J. R. Bci2000: a generalpurpose brain-computer interface (bci) system. IEEE Transactions on biomedical engineering, 51(6), (2004), 1034–1043.
- 12. Raspberry Pi 2: https://www.raspberrypi.org/products/ raspberry-pi-2-model-b/
- 13. Qt http://www.qt.io/developers/
- 14. QCustomPlot http://qcustomplot.com
- 15. MNE library http://martinos.org/mne/stable/index.htm

Methodology for Identification and Evaluation of Web **Application Performance Oriented Usability Issues**

Mihaela Ciugudean Computer Science Department, Technical University of Clui-Napoca

Str. G. Baritiu, 28, 400027, Cluj-Napoca ciugudeanmihaela@yahoo.com

ABSTRACT

This paper aims to illustrate a methodology for identifying and assessing a set of performance issues encountered in a particular web application, with impact on the usability level. Throughout this methodology, several visual techniques are determined and investigated by taking advantage of the functionalities offered by available performance monitoring tools such as JMeter and VisualVM. This way, the major performance concerns and their impact on the web application usability are easily determined. Therefore, by identifying the problematic parts of a system, corrective measures could be taken, particularly designed to operate on the root cause of the problem, thus leading to the targeted objective. Moreover, some performance improvement recommendations are further on presented in order to enhance the overall usability and the user experience, the actual goal of the research and development activities conducted. These performance concerns might be designed and implemented after fixing the critical parts causing usability problems and consequently, maximizing the user satisfaction and comfort when exploring the functionalities offered by a particular web application.

Author Keywords

Interactive software application; usability; performance improvement; JMeter; VisualVM; user experience; visual techniques.

ACM Classification Keywords

H.3.4 Performance evaluation; D.2.2 Design Tools and Techniques; H.5.2 User Interfaces.

General Terms

Visual techniques; Interactive application; Usability; Measurement.

INTRODUCTION

Over the last past decades, the web has evolved into a growing universe of interconnected web pages and applications, offering the end user a wide variety of services and functionalities. With the web development, the interactive web applications have become easily accessible regardless of place and time concerns and without any installation requirements [1]. As a consequence, end users have begun to favor web applications over traditional desktop applications. On the other hand, the rapidly increasing mobile application

Computer Science Department, Technical University of Clui-Napoca Str. G. Baritiu, 28, 400027, Cluj-Napoca dorian.gorgan@cs.utcluj.ro

Dorian Gorgan

market has shown that web applications can be easily installed on devices.

In particular, the web enables marketers to get in touch with visitors of their websites or applications and start communicating with them. Furthermore, the web might be perceived as an excellent sales channel for millions of applications, either large or small.

More precisely, end users take advantage of the wide variety of tools and services offered by web applications, by means of web browsers. Therefore, the customers have the opportunity to access, retrieve and interact with data and content located on different software applications available over the web, as stated by Mohler et al. [2].

However, due to the fast growth of the web, an intensive competition has been created among web applications. According to Stoneham and Dastbaz [3], the web, perceived as an application platform, has raised the standard for the web applications it provides, establishing a norm for highly-interactive dynamic user interfaces with real time collaborative features. In case a certain web application fails to completely fulfill the customer's needs and gets him unsatisfied, he will start using services offered by a similar website. Therefore, application developers have begun to enhance the accessibility, usability and user experience when providing specific services and products in order to meet end users' demands.

When creating a web application, it is not sufficient to provide an intuitive, self-explanatory user interface that will allow even an unexperienced computer user to perform the desired actions. As specified by Shivakumar [4], designing a consistent and standardized user interface may still not satisfy the customer, determining him to abandon your website and switch to a similar one that completely fulfills his demands. All this frustration and annoyance that might lead a user to discard your application is very often caused by a series of usability problems occurring in the website.

Considered one of the most important quality factors for web applications, usability has been receiving great attention, being recognized as a fundamental property for the success of web applications [5]. As previously said, it is not sufficient to satisfy the functional requirements of a web application in order to ensure its success. According to Purba [6], the ease or difficulty experienced by users of these web applications is largely responsible for determining their success or failure. Consequently, usability evaluations and technologies that support the usability design process have therefore become critical in ensuring the success of web applications.

The scalability should be an important characteristics of the web application. The main requirement on the usability concerns with keeping a high level of usability at any level of scalable application. It is not enough to check the usability level for single user, but to check usability of the application for a great or huge number of users.

The aim of this study is to present the way by which the visual techniques allow to recognize the root cause of the usability issues encountered in a scalable web application, thus allowing for immediate identification of the technical reason. The paper presents a way of highlighting the usability issues and a directly causal link to the correspondent technical issues.

By taking advantage of the various performance indicators when automatically testing a web application, a particularized methodology might be applied in order to remove, or at least reduce, the encountered system's weaknesses. This way, time and effort invested in unsuccessful operations to improve the overall usability of a web application are saved by starting to design and implement a solution centered on the specific issue that acts as a bottleneck in the system. By means of technical and technological solutions implemented to fix the determined root cause, the overall application usability will be enhanced.

The rest of this paper proceeds as follows. In Section 2, state of the art is presented and the contribution of the current work is highlighted. Section 3 details the proposed methodology by presenting and exemplifying each step. Finally, some concluding remarks and recommendations are proposed.

RELATED WORKS

The exponential growth of the internet and of web applications lead to the necessity to evaluate them from a quantitative view point. In the past years, valuable methodologies have been used to evaluate the quality of specific web applications and discover possible issues. More precisely, inspection methods such as Cognitive Walkthrough for the Web (CWW) and Web Design Perspectives (WDP) are proposed by Haralambos and Colette [7] to identify usability problems. According to them, the aforementioned methods are characterized by a high degree of subjectivity in usability evaluations. In order to overcome this drawback, these methods are sometimes replaced by inspection metrics-based ones which are likely to reduce the subjectivity degree, such as WebTango and Web Quality Evaluation Method (WebQEM), [8]. More precisely, Web Tango provides quantitative metrics, based on empirically validated metrics for user interfaces to build predictive models in order to evaluate other user interfaces.

On the other hand, WebQEM performs a quantitative evaluation of usability aspects, aggregating them to 70

obtain usability indicators. As stated by Singh, the purpose of this method is to systematically assess characteristics, sub-characteristics and attributes that influence products' quality, finally yielding to global, partial and elementary quality indicators that can help different stakeholders in understanding and improving the assessed product. According to Olsina and Rossi [9], by implementing the four major technical steps of WebQEM methodology, it can be employed in assessing and comparing quality requirements in the operative phase of web sites and applications as well as in early phases of web development projects. By using the methodology, either absent attributes, absent sub-characteristics, or requirements poorly implemented might be easily discovered. As conclude by Mendes and Mosley [10], WebQEM can be used to assess diverse application domains according to different user views and evaluation goals.

Nowadays, more and more companies are concerned with addressing usability aspects when designing their website. This concern has developed as a consequence of the competitive market in which each organization strives to continuously enlarge the variety of users by providing an intuitive, self-explanatory and accessible website. Usability is an even broader goal than accessibility, which refers to how easily a website can be used, understood, and even accessed by people with disabilities.

Starting with 2015, even Google has started to evaluate the usability of web applications, mostly their mobile version, and announced by email their owners of the poor rated usability results. Moreover, there was included a list of aspects to be addressed in order to overcome these usability concerns.

One example of receiver of such an acknowledgement was WordPress, which begun to assess the concerns presented in the email received from Google and check their behavior in the mobile version of the web application.

A similar case study of a website trying to enhance its usability and accessibility features is The American Foundation for the Blind. Their strategy was to undergo a major redesign in order to improve the site's usability and create a more logical, user-friendly information architecture.

However, these are only two examples of websites dealing with usability and accessibility concerns in order to maximize the user experience, maintain and enlarge the number of customers. Many other web applications have realized that a consistent and intuitive user interface is sometimes not enough to please the customer's needs and therefore, started to analyze and fix their usability issues.

METHODOLOGY FOR IDENTIFYING AND SOLVING USABILITY ISSUES

The purpose of this paper is to briefly describe the methodology used for identifying the usability issues in a specific web application and inferring the technical reason. Therefore, the considered context is the following one: starting from a particular web application, its usability is defined as the system' capability to successfully handle a variable number of concurrent users. This case the usability is evaluated by the reported error percentage for each case scenario. Therefore, the main step is to measure the application's responsivenessoriented behavior when successively increasing the number of users, by taking advantage of the functionalities offered by JMeter tool] [11]. After concluding the existence of such usability issues, their root cause is to be determined by means of VisualVM [11], performance monitoring tool, thus obtaining accurate information regarding the sources of the problem. The further step is to analyze the situation and propose and implement specific solutions to overcome the usability problem. Finally, by measuring the application's behavior after applying the established solution, the followed methodology is validated and conclusions related to its accuracy may be stated.

Usability

Being one relevant component of web applications' quality, defined as the extent to which a system is able to satisfy its customers in efficiently achieving their goals, usability is recognized as a fundamental property for software applications' success. In the context of software lifecycle, usability is perceived as being relevant to all its stages, not only at the end of the product development.

According to Osina and Rossi [9], a software product quality might be defined in terms of fundamental characteristics (usability, functionality, reliability, efficiency, portability, and maintainability) as defined in the ISO/IEC 9126-1 standard, usability being one of them.

Specifically, for the considered web application, usability metrics are assessed by monitoring the system's capacity to handle a reasonable amount of concurrent usage. More precisely, the Application Performance Monitoring tests are executed in the context of successively increasing the workload expressed as the number of concurrent active users and monitoring the system's behavior. Therefore, the exemplified attribute in usability evaluation is the number of concurrent users that is successively increased and obtained results interpreted from the reported error percentage values.

Use case scenario: Online Scrum tool

To begin with, the web application under analysis represents a system capable of providing a tool used as an agile project organizer. More precisely, the application is intended to offer the functionalities of an online tool which allows the persons involved in the Scrum process to perform it regardless of any inconveniences such as physically distributed teams, unreachable customer or any other factors that may negatively affect the Scrum process and thus the corresponding software project, [13, 14].

In this context, one use case that has been studied mainly consists of the following actions to be undergo by a privileged user, say a Scrum Master: login, add a new developer to the project, delete another developer (that might have left) and logout.

Methodology for identifying and solving usability issues

In order to identify and evaluate the issues encountered in a web application, by means of visual techniques offered by performance monitoring tools such as JMeter and VisualVM, several use case scenarios were considered. Consequently, for these use cases, a series of experiments were conducted and the visual results were analyzed so that to determine the root causes of the usability issues and propose technical solutions to successfully solve them.

We may define visual techniques as the ensemble of differently shaped graphical representations of the executed testplans that allow evaluation based on various diagnosis indicators while adjusting the variable parameters and performing successive measurements. These visual techniques apply in the context of using dedicated Application Performance Monitoring (APM) tools, such as JMeter and VisualVM, used as a basis in building the methodology for detecting potential performance weaknesses of the system under test and subsequently propose particular solutions to overcome them.

Step 1: Determine the existence of usability problems

In order to assess the non-functional system's characteristics such as usability, availability, scalability, serviceability and other performance oriented metrics, JMeter monitoring tool was firstly used and the reported results analyzed by means of different visual techniques [11].

The aim is to determine potential usability issues in the context of a high workload obtained from a large number of concurrent users that access the web application. The basic idea is that some usability issues are not revealed at low scale of application. The goal is to identify the context in which such issues become detectable. It is difficult and sometimes almost impossible to manually perform all these performance tests, by considering a large number of online users. As a result, these cases were performed by means of the JMeter tool which automatically executes the desired scenarios by simulating the existence of the desired number of users.

The first step of the proposed methodology was to conduct JMeter performance tests, simulating a number of 100 users that simultaneously execute this use case. In the followed context, the considered web application as well as all the system's configuration are the fixed elements, while the number of concurrent users represent the variable elements. Finally, the measurable attributes are the reported error percentage, standard deviation, throughput, number of KB/sec, etc. Specifically for the considered use case, the evaluation is performed by analyzing the reported error by successively increasing the number of concurrent users.



Figure 1. View Results Tree – 100 users.

Finally, the obtained test results are assessed by means of different available listener components which offer visual representation of the output. More precisely, the investigation's objective dictates the number and variety of output listeners that are attached to the executed JMeter test plan.

To begin with, for 100 concurrent threads which simultaneously execute the sample case scenario (login, add a new developer to the project, delete another developer and logout), the application responds with error for some of these 100 HTTP requests. This result might be visualized from the View Results Tree image (which was obtained from JMeter tests) in which the output of each of the 100 user requests to the application (HTTP requests) is displayed, as well as relevant information for all of them (Figure 1). The precise error percentage that was reported for this case scenario is 36.36%, as illustrated in the Summary Report (Figure 2).

As the particular web application is intended to serve a large number of concurrent online users, it has to successfully respond to all of them, or to as much as possible. Therefore, the purpose of this step was to assess the system's behavior in the context of simulating a large users' number. If these results seem worrying for a number of 100 users concurrently accessing the system, the application performs even worse when increasing the number of users to 300. This outcome was determined by performing the same JMeter performance tests with an increased number of 300 users and evaluating the results. The inspected View Results Tree illustrated more HTTP

Summary R	eport									
Name: Summary Report										
Comments:										
Write results to file / Read from file										
Filename Browse Log/Display Only: Errors Successes Config									Configure	
Label	# Samples	Average	Min	Max	Std. Dev.	Error %	Throughput	KB/sec	Avg. Bytes	
http://localhost/	100	7015	416	9689	3809.09	0.00%	9.3/sec	1437.85	158249.0	
http://localhost/	300	6010	10	9447	3639.71	0.00%	4.6/sec	99.38	22036.1	
http://localhost/	200	5750	269	15688	3313.77	50.00%	6.6/sec	163.19	25292.5	
http://localhost/	200	4857	1330	17182	2692.20	100.00%	6.1/sec	49.71	8404.0	
http://localhost/	100	4797	3311	9952	1066.11	0.00%	4.8/sec	11.25	2417.0	
http://localhost/	200	1290	3	5605	1216.51	50.00%	10.9/sec	487.62	45620.5	
TOTAL	1100	4876	3	17182	3444.26	36.36%	16.7/sec	572.03	35037.1	

Figure 2. Summary report - 100 users.

request that were not handled by the application as compared to the case of 100 concurrent users. Moreover, the reported error percentage was 54.55%, which denotes the system's disability to handle a reasonable usage workload.

Furthermore, in case we increase the users' number, the error percentage increases as well, leading to almost an unresponsive application. The graphic below visually depicts the relationship among the number of concurrent users and the measured error percentage (Figure 3).



Figure 3. Reported error for different numbers of users.

Consequently, by means of these visual techniques and of the proposed methodology, the system's behavior might be evaluated by subsequently changing the number of users concurrently accessing the application. As bottom line of this first step, we may conclude the erroneous results obtained for a large number of customers, thus leading to poor serviceability and customer experience when taking advantage of the system's functionalities.

Step 2: Determine the root cause of the usability issues

The next step in the proposed methodology is to successfully determine the root cause of the encountered responsiveness-oriented performance issues so that to further on establish concrete solutions to be implemented in order to increase the usability and overall performance of the application. To accomplish this, the simulated test cases executed by taking advantage of the functionalities offered by JMeter, were associated by the ones provided by VisualVM tool [12].

VisualVM as an application performance monitoring tool which freely comes when installing Java JDK, offers several monitoring options, among which the profiling and sampling ones are of interest for our research. The profiler and sampler tools accessible from the VisualVM's IDE are available for both CPU and memory load as performance indicators.

By instrumenting the web application under test, profiling adds a constant amount of extra execution time to every single method call. Sometimes, this results in adding large amount of time to the execution which may even last hours. VisualVM's sampler works by taking a dump of all of the threads of execution on a fairly regular basis, and uses this to determine how much CPU time each method spends. Usually, sampling takes a constant amount of time each second to record stack traces for each thread, thus only adding 5 - 10 minutes of execution time in total, while still succeeding to provide information regarding the source causes of potential performance vulnerabilities. However, sampling has the drawback that the number of invocations recorded is not necessarily accurate, since a short method could easily start and finish between stack dumps. Therefore, the recommendation is to use sampling for recording the amount of CPU time or Memory consumption, rather than the number of invocations, thus having the chance to identify performance problems, much faster than the standard profiler.

Specifically for the scope of our research, only the CPU diagnosis indicator's results were analyzed and, through a series of visual techniques, the system's bottlenecks were identified and further on, specific solutions proposed so that to overcome the performance problems.

As a result, for the presented use case scenario (login, add a new developer to the project, delete another developer and logout), VisualVM tool offered the ability to perform profiling and sampling activities in order to establish the main bottleneck(s) of the system. Figure 4 presents the CPU usage information for sampling of the considered use case.

These two images offer information related to the bottleneck point of the application, by listing the exact functionalities that consume most of CPU's time. Consequently, the association between the JMeter and VisualVM tools, provide accurate information regarding the precise source of failure for the considered web application and enable specific solutions to be designed and implemented.

Step 3: Analyze results and propose solutions to be implemented in order to overcome the usability issues By analyzing the results obtained from VisualVM, visual techniques might be used to determine the root cause of the usability issues. Furthermore, from the two images (obtained after experiencing sampling and profiler activities), it can be determined that most of the CPU time is spent on calling functionality at the persistence layer. This outcome is provided by the database method (findAll(), findByName(), call findAllFromPanelForDevelo-per(), ...) that consume the most CPU when the considered use case scenario is performed. Therefore, the main CPU bottleneck is caused by methods from the persistence layer (when querying the database) and, thus, further improvements should be implemented at the database level.

Consequently, by means of the visual techniques experienced, the root cause of the encountered usability problems were briefly identified and, further on technical solutions might be proposed and corresponding approached implemented, thus enhancing the overall system's performance. 🐻 Overview 🗰 Monitor 🚍 Threads 🔬 Sampler 🚫 Profiler

O Tomcat (pid 6740)			
Sampler			E Setting:
Sample: CPU Memory Stop			
CPU samples Thread CPU Time			
0 🖏 🗐 Snapshot			Thread Dum
Hot Spots - Method	Self time [%] 💌	Self time	Self time (CPU)
com.iquestgroup.advancedframeworks.ScrumTaskboard.persistence.jpa.JpaDeveloperDao.findAll ()		6,470 ms (75.7%)	6,470 ms
com.iquestgroup.advancedframeworks.ScrumTaskboard.persistence.jpa.JpaDeveloperDao.findByName ()		1,264 ms (14.8%)	1,264 ms
com.iquestgroup.advancedframeworks.ScrumTaskboard.persistence.jpa.JpaTaskDao.findAllFromPanelForDeveloper ()		208 ms (2.4%)	208 ms
com.iquestgroup.advancedframeworks.ScrumTaskboard.persistence.jpa.JpaTaskDao\$\$EnhancerByCGLIB\$\$1d1c1641.findAllFromPanelForDeveloper ()		205 ms (2.4%)	205 ms
com.iquestgroup.advancedframeworks.ScrumTaskboard.persistence.jpa.JpaDeveloperDao\$\$EnhancerByCGLIB\$\$5b2545e4.findAll()		127 ms (1.5%)	127 ms
com.iquestgroup.advancedframeworks.ScrumTaskboard.persistence.jpa.JpaDeveloperDao\$\$EnhancerByCGLIB\$\$5b2545e4.findByName ()		99.5 ms (1.2%)	99.5 ms
com.iquestgroup.advancedframeworks.ScrumTaskboard.domain.Developer. <init> ()</init>		90.4 ms (1.1%)	0.000 ms
com.iquestgroup.advancedframeworks.ScrumTaskboard.persistence.jpa.JpaUserDao\$\$EnhancerByCGLIB\$\$5ea27cfb.create()		77.3 ms (0.9%)	77.3 ms
$com.iquestgroup.advanced frameworks. Scrum Taskboard. persistence.jpa. Jpa Developer Dao \$Fast Class By CGLIB \$a098 afdc. \\ invoke ()$		0.000 ms (0%)	0.000 ms
$com.iquestgroup.advanced frameworks. Scrum Taskboard.service.impl. Developer Service Implementation. compute {\it NumberOfWorkingHours} () and the service interval of the service of the service interval of the service inte$		0.000 ms (0%)	0.000 ms
com.iquestgroup.advancedframeworks.ScrumTaskboard.service.impl.DeveloperServiceImplementation.isDeveloperBusy ()		0.000 ms (0%)	0.000 ms
com.iquestgroup.advancedframeworks.ScrumTaskboard.service.impl.DeveloperServiceImplementation.findAll ()		0.000 ms (0%)	0.000 ms
com.iquestgroup.advancedframeworks.ScrumTaskboard.controller.DeveloperController.developers()		0.000 ms (0%)	0.000 ms
com.iquestgroup.advancedframeworks.ScrumTaskboard.controller.DeveloperController.processAdd ()		0.000 ms (0%)	0.000 ms
com.iquestgroup.advancedframeworks.ScrumTaskboard.persistence.jpa.JpaTaskDao\$\$FastClassByCGLIB\$\$6c7c9ad9.invoke()		0.000 ms (0%)	0.000 ms
com.iquestgroup.advancedframeworks.ScrumTaskboard.controller.DeveloperController.showFormDelete()		0.000 ms (0%)	0.000 ms
com.iquestgroup.advancedframeworks.ScrumTaskboard.service.impl.UserServiceImplementation.create ()		0.000 ms (0%)	0.000 ms

Figure 4. VisualVM Sampling results.

More precisely, for removing the database bottleneck, there are several techniques that may be explored and appropriate solutions implemented. By studying documentation related to database enhancements and methods to be applied on a web application to improve the persistence layer access, a series of rules and guidelines were considered. As one of the main strategy to be applied at persistence layer is adding database indexes where appropriate, this topic was studied in detail and best practices taken into account [15].

According to Singh [16], in order to fix the database bottleneck which determined the application to become unresponsive when a large number of users were trying to access it, the persistence layer strategies were applied where appropriate and the results analyzed after each step. Specifically for the described use case scenario, by means of SQL syntax, slow queries were determined and database indexes implemented in the adequate manner.

Step 4: Evaluating the solutions implemented

By implementing the established solution, the encountered bottleneck at database level was completely removed for the considered use case and the initial error percentages significantly reduced. The developed technical approach was based on an algorithm to determine the database queries which are reported as slow and evaluate the usefulness of adding an index. More precisely, after adding an index on a database table's column, the initial reported error was half reduced. Further on, several analysis were conducted in order to determine the suitability of database indexing for the problematic SQL queries, and indexes added where appropriate. This strategy brought successive performance improvements as a result of reducing serviceability and operability issues (obtained when increasing concurrent usage workload), fact denoted by the minimized reported error percentage (for a large number of concurrent users).

Validation of the implemented methodology

Therefore, by using the aforementioned visual strategies by following the described methodology steps, the root cause of the usability issue was determined and technical solutions were considered and successfully applied. The overall result was an increase in the user experience, satisfaction and response time metrics throughout the system. Consequently, the application will manage not to slow, block or frustrate concurrent customer requests' fulfillment, which determine client loss.

Furthermore, apart from the database bottleneck, the source code was revised and analyzed for the functionalities reported as taking most of CPU's time (in VisualVM's sampler and profiler experiments) and some refactoring strategies proposed. Consequently, the next step in the usability enhancement process was the development of these strategies or even re-implementing of the actual solution (where appropriate) so that to remove possible time or memory consuming algorithms and replace them with a cleaner, cost-effective approach. Both performance improvement methods (the database indexing and the source code refactoring) were considered and applied to other use case scenarios (apart from the one exemplified in this paper) and the system's behavior was definitely enhanced.

To conclude this chapter, the proposed methodology was of great help to determine the existence of usability issues, identify their root cause and allow proposal of particular technical solutions to overcome them. This way, the main operation was performed at the very bottleneck of the system and immediate recovery occurred. As a result, this boosts the overall system performance and offered the desired user experience because of improved usability.

CONCLUSIONS AND FURTHER WORK

As briefly presented in the previous chapter, by following the proposed methodology, the web application's performance problems, oriented towards serviceability, operability and user responsiveness, were spotted and the context clearly defined. Furthermore, by using tools such as JMeter and VisualVM, the obtained results were visually analyzed and the root cause of the issue determined. This way, particular strategies were proposed, centered on fixing the bottleneck, rather than offering a general system improvement (which would only slightly enhance the performance, the root cause not being handled yet).

This methodology proved successful for the considered context and it aimed to enhance the overall system's performance by focusing on the problems identified, i.e. its response time capabilities when increasing the usage workload (defined as the number of concurrent active users). However, attention must be paid to the specifics of a certain system whose performance is to be evaluated and further on improved, because they may differ and therefore, several adjustments have to be applied to the proposed methodology in order to fit particular needs.

Therefore, by applying the described methodologies to fix the encountered problem, system's usability was enhanced, fact clearly proven by the conducted experiments. Not only did the conducted results denote the system's improved performance-centered features by the considerably reduced reported error, but also the methodology validation indicates its efficiency. More precisely, the obvious advantages and strengths of the presented methodology are related to speed in analyzing the system's non-functional requirements from the customer's experience and serviceability viewpoint, as well as accuracy in determining the main bottleneck. The specified speed capabilities refer to the automatization of the testing process by using the JMeter tool which allows to simulate the number of concurrent users, otherwise being almost impossible to assess the system's behavior in this context (as manual testing would be more than cumbersome). As far as the accuracy concerns, the presented methodology offers precise diagnosis of the encountered response time-based user operability problems by identifying their root cause. This way, specific, problem-oriented solutions might be designed and implemented. By focusing on the source of failure, corrective measures are taken at an early stage in the development process.

However, there are still several strategies to be implemented that would obviously boost the entire application performance, and offer a better user experience. More precisely, technical solutions such as adding caching to the developed functionalities and setting up a load balancer are only some examples that involve further study, analysis and implementation effort. All these strategies have to be tested to determine their suitability for our particular web application and establish practical solutions to be developed in order to successfully apply them.

REFERENCES

- 1. Rossi, G. et al. *Web Engineering: Modelling and Implementing Web Applications*. Springer Science and Business Media, 2007.
- 2. Mohler, J. L. and Duff, J. M. *Designing Interactive Web Sites*. Delmar/Thomson Learning, 2000.
- 3. Stoneham, R. and Dastbaz, M. *Building Interactive Web Applications*, Addison-Wesley Longman, Limited, 2006.
- 4. Shivakumar, S. K. Architecting High Performing, Scalable and Available Enterprise Web Applications. Morgan Kaufmann, 2014.
- 5. Barish, G. Building Scalable and High-Performance Java Web Applications Using J2EE Technology. Addison-Wesley Professional, 2002.
- 6. Purba, S. Architectures for E-Business Systems: Building the Foundation for Tomorrow's Success. CRC Press, 2001.
- Haralambos, M. and Colette, R. Advanced Information System Engineering, 23rd International Conference, CAiSE 2011, London, (2011).
- 8. Singh, K. K. A Quantitative Method for Evaluation of Websites Quality using WebQEM Tool, Journal of Global Research Computer Science and Technology (JGRCST).
- 9. Olsina, L. and Rossi, G. Towards Web-site Quantitative Evaluation: defining Quality Characteristics and Attributes. *Proceedings of IV Int'l WebNet Conference*, World Conference on the WWW and Internet, Hawaü, US, 1, (1999), 834-839.
- 10. Mendes, E. and Mosley, N. *Web Engineering*. Springer Science & Business Media, 2006.
- 11. Online documentation for JMeter tool: http://jmeter. apache.org/
- 12. Online documentation for VisualVM tool: https://visualvm.java.net/
- 13. Schwaber, K. and Sutherland, J. *The Scrum Guide.The Definitive Guide to Scrum: The Rules of the Game.* July, 2013.
- 14. Rubin, K. S. Essential Scrum: A Practical Guide to the Most Popular Agile Process, Michigan, USA, Pearson Education, 2013.
- 15.Garner, S. R. Data Warehouse Implementation Strategies: A Mixed Method Analysis of Critical Success Factors, ProQuest, 2007.
- 16. Singh, H. Data Warehousing: Concepts, Technologies, Implementations, and Management, Prentice Hall PTR, 1998.

Testing the technology acceptance model with Romanian university students

Alexandru Balog, Costin Pribeanu

ICI Bucharest, Romania Bd. Mareşal Averescu No.8-10, Bucureşti {alexb,pribeanu}@ici.ro

ABSTRACT

Social networking websites, in general, and Facebook, in particular, gained a lot of popularity among university students. The increasing rate of the Facebook adoption stimulates the researchers to analyze and explain its usage. A well known model able to do this is the technology acceptance model (TAM) that has been tested on a large diversity of computer systems. This paper presents a case study in testing TAM to explain the adoption of Facebook by Romanian university students. The model estimation results show that the perceived ease of use and the perceived enjoyment have a stronger influence on the intention to use Facebook than the perceived usefulness. This study found that the most important determinant is the perceived enjoyment.

Key words

Social networking sites, technology acceptance, perceived ease of use, perceived enjoyment, perceived usefulness.

ACM Classification

D.2.2: Design tools and techniques. H5.2 User interfaces.

INTRODUCTION

Social networking websites are interactive environments for communication, socialization, collaboration, exchange of information / resources, and self-advertising. Facebook is a social networking site that gained a lot of popularity in recent years and is widely used by university students [12, 15, 17, 18, 20].

Several studies showed that for many users Facebook became part of their everyday life, being both useful [3, 12, 15, 20] and enjoyable [14, 17]. The widespread usage of social networking websites stimulated the research aiming to explain their adoption [12, 14, 17]. However, few approaches exist that analyze the main determinants of the Facebook adoption by university students.

A well-known model aiming to explain and predict the technology acceptance on a large variety of technologies is TAM (Technology Acceptance Model), developed by Davis [7], and Davis et al. [8]. TAM focuses on two main drivers, the perceived ease of use (PEU) and the perceived usefulness (PU), that determine the intention to use a technology. In further studies, the perceived enjoyment (PE) has been added as an intrinsic motivation to use a technology [9, 13].

In this paper, a case study of testing TAM is presented to explain the adoption of Facebook by Romanian university students. The model considers three factors: the perceived ease of use, perceived usefulness, and the perceived enjoyment. The sample consists of 414 Romanian university students from a university of economics.

The rest of this paper is organized as follows. In the next section we present the theoretical background and model conceptualization. Then, the empirical validation of the model and the estimation results are discussed. The paper ends with conclusions and future work.

THORETICAL BACKGROUND AND HYPOTHESES

Usage of Facebook in educational contexts

The explosion of social networking websites is closely related to the social learning featuring meeting, active participation, critical thinking, information and content sharing. The proliferation of Facebook in university contexts creates many opportunities for learning. Students benefit from sharing information and resources in their social network as well as from the debate on various issues of interest [15, 18, 20].

Lampe et al. [15] investigated the usefulness of Facebook as an information source. Their study shows how FB users convert the social capital in information (another form of capital). Lee et al. [16] explored the perceived community value of Facebook and found that the experiential value was the most important outcome of information sharing in a social network.

In a recent study, the Facebook usefulness for students has been modeled as a multidimensional construct featuring three dimensions: social, information, and collaboration usefulness [3]. The multidimensional model have been then used to analyze the differences between two university profiles [19].

Park et al. [18] identified four primary needs for joining Facebook groups: socialization, entertainment, self-status seeking, and information. The study of Hart et al. [12] as regards the usage patterns, shows that several Facebook functions were rated for positive experiences. Most often selected experiences are the enjoyment and the curiosity.

Related work in the technology acceptance

The first technology acceptance model considered two main drivers of the behavioral intention to use a computer system: the perceived ease of use and the perceived usefulness [7]. Later on, TAM has been extended with the perceived enjoyment (PE).

Heijden [13] distinguished between two kinds of computer system: production-oriented (or utilitarian) and pleasure-

oriented (or hedonic). His study concluded that in the case of hedonic systems, perceived ease of use and perceived enjoyment (intrinsic motivation) are stronger predictors of the intention to use than perceived usefulness (extrinsic motivation) [13].

Perceived enjoyment is an important determinant of the intention to use since users who experience pleasure or enjoyment are more likely to form a positive attitude and intention to use it than others [9].

The e-learning systems that are using novel technologies, such as Augmented Reality, aim to increase the students' motivation to learn. Testing TAM with these e-learning systems showed that the perceived enjoyment is a stronger predictor than the perceived usefulness [9]. Similar results have been reported in the study of Lee et al. [16] that tested TAM with an Internet-based learning medium.

In their study, Lin & Lu [17] integrated a motivational model to analyze why people use the social networking websites. They found that the perceived enjoyment is the main driver of the continuation intention. More recently, Iordache et al. (14) tested the motivational model with 152 Lithuanian university students and found that perceived enjoyment is a stronger predictor of the continuation intention of using Facebook than perceived usefulness.

Research model and hypotheses

This study proposes aims to empirically validate the acceptance model in the Facebook context based on the theoretical framework of TAM and the literature review.

In TAM, the extrinsic motivation is conceptualized as perceived usefulness, defined as "the degree to which a person believes that using a particular system would enhance his or her job performance" [7].

Intrinsic motivation is conceptualized as perceived enjoyment, defined as "the extent to which the activity of using a specific system is perceived to be enjoyable in its own rights, aside from any performance consequences resulting from system use" [9].

Perceived ease of use was defined as "the degree to which a person believes that using a particular system would be free of effort" [7].

The research model in Figure 1 illustrates the hypotheses.



Figure 1. The research model.

Previous research showed that the perceived ease of use (PEU) is supposed to positively influence the perceived usefulness (PU) [8], the perceived enjoyment [13], and the intention to continue using (CI) [21].

In an e-learning context, the perceived enjoyment (PE) is supposed to positively influence the perceived usefulness [3, 14, 21, 22]. Previous research shows that PE has both a direct and indirect influence on the intention to continue using the system [8, 13, 14].

There is extensive empirical evidence in the literature that the perceived usefulness has a direct positive influence on the intention to use [7, 8, 21].

METHOD

Data collection and sample

The model has been tested on 414 university students, from which 156 men and 258 women. The age of participants is varying between 18 and 37 years with a mean of 21.28 years (SD=2.78). Most of the participants (313) are undergraduates.

Item	Description	Μ	SD
PEU1	It is easy to learn how to use Facebook	6.10	1,27
PEU2	Facebook is easy to use	6.21	1.17
PEU3	My interaction with Facebook is clear and understandable	5,69	1,38
PU1	Using Facebook improves participation in collective activities	4.71	1.52
PU2	Using Facebook I can better present my university work to other people	3.81	1.60
PU3	Using Facebook I am better informed about events of interest in my university	5.05	1.52
PU4	Using Facebook I get useful information from university people	5.15	1.47
PU5	On Facebook I can find useful resources for my university work	4.96	1.59
PU6	Using Facebook improves communication between colleagues	4.99	1.55
PU7	Using Facebook encourages the creation of academic groups based on similar interests and needs	4.99	1.63
PU8	Using Facebook improves the student group work	4,92	1,49
PE1	I have fun using Facebook	4,56	1,73
PE2	Using Facebook is enjoyable	4.91	1.53
PE3	Using Facebook is entertaining	4.00	1.72
PE4	Using Facebook is pleasant	4.72	1.63
CI1	I intend to continue using Facebook in the future	5.34	1.59
CI2	It is likely that I will continue using Facebook in the future.	5.54	1.51
CI3	I will regularly use Facebook in the future.	5.08	1.73

Table 1. Variables.

RoCHI 2016 proceedings

Students were asked to answer general questions (faculty, enrollment, age, and gender), questions regarding the use of Facebook (network size, frequency, and duration of use), and then to evaluate several items on 7-points Likert scale. The variables are presented in Table 1.

The constructs PEU, PE, and CI have been measured by using and / or adapting existing scales in the literature [8, 9, 14]. For PU a multidimensional scale developed and validated in [3] has been used. The scale is featuring three dimensions: social usefulness, information usefulness, and collaboration usefulness.

Analytical procedures

Data analysis has been carried out using the statistical package SPSS 16.0 for Windows. Structural Equation Modelling (SEM) with AMOS 7.0 software was used to test the model. Testing was carried out in accordance with a two-step approach [1] including the validation of the measurement and structural models.

Based on the recommendations from the literature [10, 11], the following goodness-of-fit measures were used: normed chi-square (χ^2/df), Tucker-Lewis index (TLI), comparative fit index (CFI), standardized root mean square residual (SRMR), and root mean square error of approximation (RMSEA).

The fit between the model and the data is supported if the following conditions hold: the normed chi-square is less than 3, TLI and CFI exceed 0.95, SRMSR is less than 0.05, and RMSEA is less than 0.08.

The three dimensions of the perceived usefulness have already been validated in two previous studies, as firstorder factors, for unidimensionality and reliability [3, 19]. According to Bagozzi & Edwards [2], in this case the items could be averaged and the resulting constructs could further be used as items of a first-order construct [2] in other models. The descriptives are presented in Table 2.

Item	Description	М	SD
mUS	Social usefulness	4.26	1,38
mUI	Information usefulness	5.05	1.33
mUC	Collaboration usefulness	4.97	1,37

Table 2. Descriptives for the dimensions of PU.

ANALYSIS AND RESULTS

Descriptive statistics

The standard deviations ranged from 1.27 to 1.73, indicating a fairly narrow spread of scores around the mean. Univariate and multivariate outliers were searched in the data set and since none of the cases appeared to be extreme, all the data were kept for analysis.

Data normality was investigated in terms of skewness and kurtosis. The values are within the recommended level [11], supporting the moderate departure from normality for all variables.

All variables in the model have mean value over 4.00 (neutral value). The items related to the perceived ease of

use were highly rated by the students, which suggests that that Facebook is very easy to use.

Measurement model

We examined the convergent and discriminant validity of the model using the procedure outlined by Fornell and Larcker [10]. The results of model testing are presented in Table 3.

	Alpha	CR	AVE	PEU	PE	PU	CI
PEU	0.855	0.885	0.722	0.849			
PE	0.935	0.914	0.728	0.376	0.853		
PU	0.935	0.784	0.550	0.402	0.533	0.742	
CI	0.917	0.920	0.793	0.404	0.619	0.478	0.891

noles. The bola alagonal numbers are the square root of Avi	Notes:	The	bold	diagonal	numbers	are the	square	root o	f AVI
---	--------	-----	------	----------	---------	---------	--------	--------	-------

Table 3. Results of discriminant validity.

All standardized item loadings were statistically significant (*t*-values > 1.96). The item reliability (R^2) values are above the suggested standard of 0.50 [11], with exception of PU2 (0.40). Cronbach's alpha values are acceptable for all three constructs.

The composite reliability (CR) values ranged from 0.827 to 0.920, above the minimum level of 0.70 [11], indicating an adequate reliability. The values of the average variance extracted (AVE) are all above the minimum level of 0.50 [11], ranging from 0.715 to 0.794, confirming the convergent validity.

The discriminant validity of the constructs was examined through the squared correlation test [10]. The results in Table 2 show that the square root of the AVE is greater than the correlations between constructs, thus providing evidence of an adequate discriminant validity.

Structural model

A structural equation modeling (SEM) was carried on to test the fit between the research model and the data. The model testing results presented in Figure 2 show the item loadings, the standardized path coefficients, and the explained variance (\mathbb{R}^2) for each variable.



Figure 2. The structural model estimation results.

The measurement model fits acceptably with the data. Although the χ^2 test is significant ($\chi^2 = 191.46$, df = 59, p = 0.000), the other fit indices are very good: $\chi^2/df = 3.245$, CFI = 0.964, TLI = 0.952, SRMR = 0.052, RMSEA = 0.074 and its 90% confidence interval is fairly narrow (i.e., 0.062-0.086).

The analytical results show that all hypotheses are supported. The perceived usefulness has the smallest influence on the continuation intention ($\beta = 0.12$, p < 0.047). The rest of path coefficients are significant at p <0.001 level.

The model explains 33% of the variance in the perceived usefulness, 14% in the perceived enjoyment, and 43% in the continued intention to use Facebook.

DISCUSSION AND CONCLUSION

The main contribution of this study is an empirical validation of the technology acceptance model, measuring the contribution of each factor to the continued intention to use of Facebook by university students.

All hypotheses in this study have been confirmed, thus showing that the intention to continue using Facebook is influenced by three main factors: ease of use, usefulness, and enjoyment. The perceived ease of use has both a direct effect and an indirect effect, mediated by the perceived enjoyment and perceived usefulness.

This study found that for the Romanian university students the perceived enjoyment is the strongest determinant of the Facebook adoption. The perceived ease of use and the perceived enjoyment are stronger determinants of the intention to continue using Facebook than the perceived usefulness, which confirms the findings of Heijden [13].

The results are also confirming the findings from a previous study that tested the motivational model on a sample of university students from Lithuania [14]. Overall, the findings of this study confirm the hedonic nature of the Facebook.

Future work will focus on extending the model, in order to analyze the external variables that are influencing the Facebook adoption. Then the evaluation instrument will be administrated in several universities in order to crossvalidate the model on several samples.

ACKNOWLEDGMENTS

This work was supported by the Romanian grant financed by ANCS under COGNOTIC 1609 0101 / 2016.

REFERENCES

- Anderson, J.C. and Gerbing, D.W. Structural Equation Modelling in Practice: A Review and Recommended Two-Step Approach. *Psychological Bulletin 103 (3)*, (1988), 411-423.
- Bagozzi, R. and Edwards, J. A general approach for representing constructs in organizational research. *Organizational Research Methods* 1(1), (1998), 45-87.
- 3. Balog, A. and Pribeanu, C. A multidimensional model of the Facebook usefulness for students. *Revista*

Româna de Interactiune Om-Calculator 8(2), (2015), 157-172.

- Balog A. and Pribeanu C. An extended acceptance model for Augmented Reality educational applications. Chapter 22 in *Handbook of Research on 3-D Virtual Environments and Hypermedia for Ubiquitous Learning*, Neto, F. M., de Souza, R., & Gomes, A. S. (Eds), Hershey, PA: IGI Global. (in press), 2016.
- Cheng, M. Antecedents and consequences of elearning acceptance. *Information System Journal* 21 (3), (2005), 269-299.
- Chiu, C. M., Hsu, M. H., Sun, S. Y., Lin, T. C. and Sun, P. C. Usability, quality, value, and e-learning continuance decisions. *Computers & Education*, 45(4), (2005), 399-416.
- Davis, F.D. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13 (3), (1989), 319-340.
- Davis, F.D., Bagozzi, R.P. and Warshaw, P.R. User acceptance of computer technology: A comparison of two theoretical models, *Management Science*, 35 (8), (1989), 982-1003.
- 9. Davis, F.D., Bagozzi, R.P. and Warshaw, P.R. Extrinsic and intrinsic motivation to use computers in the workplace. *Journal of Applied Social Psychology*, 22 (14), (1992), 1111-1132.
- 10. Fornell, C. and Larcker, D. F. Evaluating structural equations models with unobservable variables and measurement error. *Journal of Marketing Research*, 18 (1), (1981), 39-50.
- Hair, J. F., Black, W. C., Babin, B. J. and Anderson, R. Multivariate Data Analysis. 6th Ed., Prentice Hall, 2006.
- 12. Hart, J., Ridley, C., Taher, F., Sas, C. and Dix, A. Exploring the Facebook experience: a new approach to usability. *Proceedings of the 5th Nordic conference on Human-computer interaction –NordiCHI 2008*, ACM, (2008), 471-474.
- Heijden, H. van der. User acceptance of hedonic information systems. *MIS Quarterly* 28 (4), (2004), 695-704.
- 14. Iordache, D.D., Balog, A., Pribeanu, C. and Lamanauskas, V. A motivational model for the acceptance of Facebook by university students (2015). Dardala, M., Rebedea T.E. (Eds.) *Proceedings of RoCHI 2015 Conference*, Bucharest, 24-25 September, (2015), 21-24.
- 15. Lampe, C., Vitak, J., Gray, R. and Ellison, N.B. Perceptions of Facebook's value as information source. *Proceedings of CHI 2012*, (2012), 3195-3204.
- 16. Lee, M.K.O., Cheung, C.M.K. and Chen, Z. Acceptance of Internet-based learning medium: the role of extrinsic and intrinsic motivation. *Information and Management*, 42 (3), (2005), 1095-1104.
- 17. Lin, K. Y. and Lu, H. P. Why people use social networking sites: An empirical study integrating network externalities and motivation theory.

RoCHI 2016 proceedings

Computers in Human Behavior, 27(3), (2011), 1152-1161.

- 18. Park, N., Kee, K. F. and Valenzuela, S. Being immersed in social networking environment: Facebook groups, uses and gratifications, and social outcomes. *CyberPsychology & Behavior*, 12(6), (2009), 729-733.
- 19. Pribeanu, C. and Lamanauskas, V. Usefulness of Facebook for students: analysis of university profile differences from a multidimensional perspective. *Proceedings of eLSE 2016 Vol. II*, Bucharest, 21-22 April, (2016), 557-564.
- 20. Selwyn, N. Faceworking: exploring students' education-related use of Facebook. Learning, Media and Technology, 34(2), (2009), 157-174.
- 21. Venkatesh, V. and Davis, F.D. A theoretical extension of the technology acceptance model: four longitudinal field studies. *Management Science* 45 (2), (2000), 186-204.
- 22. Yi, M.Y. and Hwang, Y. Predicting the use of webbased information systems: self-efficacy, enjoyment, learning goal orientation, and the technology acceptance model. *International Journal of Human Computer Studies* 59(4), (2003), 431-449.

Measuring the negative effects of the Facebook dependence on the students' university work

Gabriel Gorghiu Valahia University Targoviste

Bd. Carol I, Nr.2, Targoviste, Romania ggorghiu@gmail.com

ABSTRACT

Recent research on the use of social networking websites reveals an intense use of Facebook that may cross the boundaries between use and abuse and may lead to dependence. The objective of this paper is to analyze the relationship between the Facebook dependence and the negative consequences of the excessive use of Facebook on the students' university work. An associated objective is to adapt and test two sub-scales measuring the withdrawal syndrome and the negative effects on the university work. The results of two empirical study that used samples from two Romanian universities show a significant positive correlation between the Facebook dependence and its negative effects on the work of students.

Author Keywords

Social networking websites, Facebook, Facebook use and abuse, negative effects of social networking websites.

INTRODUCTION

Recent research on the use of social networking websites, in general, and Facebook, in particular, shows that Romanian university students have very large Facebook networks, log on several times each day (or they are continuously logged on), and spend a lot of time, daily [11, 13]. Several studies reveal an intense Facebook use that may cross the boundaries between use and abuse. The fact that the overuse of the social networking websites may lead to negative effects upon students' work has been widely researched in the last years [3, 16]. Several measurement scales exist that aim to capture the key dimensions of the addictive behavior related to the use of Internet. Many conceptualizations are featuring multidimensional models that include a diversity of constructs, such as: withdrawal syndrome, salience, tolerance, overuse, negative effects on work, or mood modification [1, 9, 14, 15].

The objective of this paper is to measure and analyze the relationship between the Facebook dependence and the negative consequences of the excessive use of Facebook on the university work of students. Facebook dependence was measured with the withdrawal syndrome scale [14]. For the negative consequences, the scale developed by Masur et al. [14] has been used. An additional objective of the paper is to validate these scales on two samples of Romanian university students.

The rest of this paper is organized as follows:

Valentina Iuliana Manea Technical University of Civil Engineering Bucharest

Bd. Lacul Tei, Nr.124, Bucharest, Romania m.valentina70@yahoo.ro Dragoş Daniel Iordache, Costin Pribeanu National Institute for Research and Development in Informatics - ICI Bucharest Bd. Maresal Averescu, nr.8-10, Bucharest, Romania

{iordache, pribeanu}@ici.ro

The following section presents related work in the area of social networking websites with a focus on the Facebook dependence and associated negative effects on the educational activities. In section 3, the empirical study is presented. The paper ends with discussion and conclusion.

RELATED WORK

Negative effects of social networking websites

The use of social networking web-sites in the educational context is widely researched in the last years. The Facebook popularity among university students raised several research questions, as regards its educational potential [5, 13]. Nevertheless, there are many studies revealing not only positive, but also negative effects of social networking websites [1, 3, 8]. Young people spend a lot of time on Social Networking Sites (SNS), and this can cause them to become potential addicts. SNS addiction has been defined as a failure to control usage, which leads to negative personal outcomes [12].

Orosz et al. [15] consider that Facebook intensity can be distinguished from Facebook habits - time spent on Facebook, number of Friends, and number of group memberships - because the latter does not necessarily reflect on the emotional connectedness to Facebook use. They distinguish Facebook addiction from Facebook intensity in terms of pathology: Facebook intensity is not necessarily a problematic behavior, however it is not true for Facebook addiction [1, 14].

Masur et al. [14] consider that the interplay of intrinsic need satisfaction in the offline context and the motives of SNS use is a crucial driver of SNS addiction. They argue that people experiencing shortcomings in intrinsic need fulfilment in their daily offline-lives try to compensate for these deficits online through the gratifications of SNS use. It was hypothesized that motives mediate the influence of thwarted intrinsic need satisfaction on addictive behavior on SNSs. More precisely, a lack of autonomy leads to a higher motivation to use SNSs for self-presentation and escapism, a lack of competence predicts the motive to use SNSs for acquiring information and self-presentation, and a lack of relatedness fosters users' motives to use SNSs for self-presentation and meeting new people. Those motives were hypothesized to be associated with higher levels of SNS addiction.

Facebook use and abuse

It is difficult to define the boundary between the use and

RoCHI 2016 proceedings

abuse of social networking websites. Elisson et al. [5] developed a composite measure for the Facebook intensity that aims to combine characteristics of use with attitudinal items. Although it is not unidimensional, it has been widely used in the last decade [15].

Due to the fact that Facebook addiction is an emerging field, different researchers have taken varying approaches to the measurement of this potential disorder. An adapted version of the Internet Addiction Scale by Hahn and Jerusalem [9] was used to measure SNS addiction. The wording of the 20 items of the original scale was modified to fit the context of SNS usage, in this case Facebook in specific. The items cover the five central dimensions of addiction:

- loss of control;
- tolerance;
- withdrawal syndrome;
- negative consequences for social relations;
- negative outcomes for work and performance.

Each dimension is measured with four items to which Participants responded on a 5-point Likert scale ranging from 1 "strongly disagree" to 5 "strongly agree". The results emphasize the importance of incorporating both offline need satisfaction and gratifications sought through the use of SNS to provide a comprehensive perspective on addictive behavior on SNSs.

The Caplan's social skill model of generalized problematic Internet use [4], states that individuals who prefer to communicate in an online environment are at greater risk of experiencing negative outcomes related to excessive online use. Those individuals, who demonstrate deficient self-regulation of Internet use, tend to engage in online social communication as a means of escaping from negative mood states, such as loneliness or anxiety.

Wilson, et al. [17] developed the Addictive Tendencies Scale, which has three items reflecting salience, loss of control, and withdrawal. Although these three aspects have been central in thinking about addictions, in the literature, addiction has involved six core components: (1) salience - the activity dominates thinking and behavior; (2) mood modification - the activity modifies/improves mood; (3) tolerance - increasing amounts of the activity are required to achieve previous effects; (4) withdrawal the occurrence of unpleasant feelings when the activity is discontinued or suddenly reduced; (5) conflict - the activity causes conflicts in relationships, in work/education, and other activities; and (6) relapse - a tendency to revert to earlier patterns of the activity after abstinence or control (Griffiths, 1996, 2005).

The Bergen Facebook Addiction Scalewas created by Andreassen et al. [1] with the aim to assess the respondents' level of addiction to Facebook with respect to the six components of addictions (salience, tolerance, mood modification, relapse, withdrawal, conflict). The scale contains six items representing the six factors and respondents had to answer using a 5-point scale (1= very rarely; 2 = rarely; 3 = sometimes; 4 = often; 5 = very often).

Ross et al. [16] created a Facebook Questionnaire which contained three factors; one of them was the Online Sociability Scale which assessed the individuals' engagement in Facebook activities in terms of posting, messaging or commenting on photos on a scale ranging from 1 (more than once daily) to 9 (less than once per year).

The Multidimensional Facebook Intensity Scale was developed by Orosz et al. [15]. Their study points to the facets of the emotional connectedness. The scale has four dimensions: persistence, boredom, overuse, and selfexpression. The third factor seems related to the Facebook dependence although it also taps on some negative effects of the overuse. The differentiated Facebook intensity facets can batter predict the frequent Facebook-related behaviors as liking and posting than previous measures [15].

EMPIRICAL STUDIES

Conceptualization

The objective of this study is to measure and then to analyze the relationship between two latent variables: the Facebook dependence (AWS) and its negative effects on the students' university work (ANC). The operationalization of constructs is based on the related work in the literature. For the AWS and ANC constructs, the items have been adapted from the existing sub-scales of Masur et al. [14]. The items for the two constructs are presented in Table 1.

Item	Description
AWS1	If I am off Facebook for a longer period of time I feel nervous
AWS2	When I am not online I ask myself what happens on Facebook
AWS3	I feel out of touch when I haven't logged onto Facebook for a while
ANC1	I am regularly on Facebook while being at university
ANC2	My concentration at university suffers because I am on Facebook
ANC3	I often neglect my university work because of Facebook

Table 1. Variables.

It is hypothesized that a significant positive correlation exists between the Facebook dependence and the negative consequences of the excessive use of Facebook on the university work of students.

Method

The normality of variables was checked by using SPSS for Windows. In order to assess the two scales, a confirmatory factor analysis (CFA) using structural equation modeling (SEM) approach was taken. The model has been tested with AMOS 7.0 for Windows [2], using the maximum likelihood estimation method.

Convergent validity has been assessed by examining the loadings and their statistical significance through *t*-values, the construct reliability (composite reliability), and the average variance extracted. The scale reliability has been analyzed checking the magnitude of Cronbach's alpha.

Factor loadings of all standardized items should be greater than 0.50, ideally exceed 0.7. Item reliability indicating the amount of variance should be greater than 0.50. Composite reliability (CR) measuring the internal consistency of a construct should be at least 0.60 (preferrably grater than 0.7) [6]. The average variance extracted (AVE) measuring the amount of variance captured by the construct should be greater than 0.50 [10].

The model testing results are analyzed based on the GOF (goodness-of-fit) indices recommended by Hair et al. [10].

Participants and samples

Two samples that have been collected in 2015, being used in this study.

The first sample includes 204 students from the Valahia University Targoviste (114 men and 90 women). The age is varying between 19 and 52 years with a mean of 25.89 (SD=7.68). Most of the students are undergraduates (71.1%).

The network size has a mean value of 641.13 (SD=801.94. From the mean number of FB friends 219.67 (48.60%) are students and 74.90 (11.53%) are studying in this university. The number of Facebook logs / day (selfassessed) is measured on a four-point scale (1 = once, 2 = twice, 3 = three times, and 4 = continuous log). The mean number of the logs / day is 2.73 (SD=0.94) and the time spent in minutes / day is on average 71.24 (SD=80.86).

The second sample includes 227 students (129 men and 98 women) from a Technical University of Civil Engineering in Bucharest. The participants were asked to answer some general questions then to rate the items on a 7-point Likert scale. All participants except for two are undergraduates.

After checking the multivariate outliers two observations were eliminated so the final working sample has 225 observations. The age of participants is varying between 18 and 39 years with a mean of 20.95 (SD=2.36). Almost all students are undergraduate (except for two).

The mean network size (number of Facebook friends) is 856.93 (SD=866.18. The mean number of the logs / day is 3.05 (SD=0.82) and the time spent in minutes / day is on average 79.73 (SD=106.09).

As it could be noticed, university students have large Facebook networks and spend a lot of time on Facebook.

Model estimation results

The examination of item loadings revealed that all except one (ANC1) are above the cut-off value of 0.6. A closer look at ANC1 makes it difficult to ascertain if the statement is a cause or an effect of underlying construct. Therefore, it has been decided to eliminate ANC1 and to

Item	Μ	SD	Loadings	Alpha	CR	AVE
AWS1	2.42	1.92	1.920	0.909	0.909	0.769
AWS2	2.86	1.96	1.960			
AWS3	2.67	1.96	1.960			
ANC2	2.93	2.13	0.840	0.842	0.845	0.732
ANC3	2.59	1.93	0.871			

test again the model. The descriptive, item loadings, scale

reliability, and convergent validity criteria are presented in

Table 2.

 Table 2. Descriptive, loadings and convergent validity

 (N=204)

The unidimensionality, scale reliability, and convergent validity criteria are very good. The mean values suggest a moderate Facebook dependence and moderate negative effects.

The model testing results are presented in Figure 1. There is a significant (p<0.001) positive correlation between the Facebook dependence and the negative consequences on the work of students.



Figure 1. Model testing results (N=204).

The results revealed an excellent fit of the model with the data: $\chi^2 = 6.27$, DF = 4, p = 0.180, $\chi^2/\text{DF} = 1.568$, TLI = 0.991, CFI = 0.96, SRMR = 0.016, RMSEA = 0.053.

Cross validation

Since the ANC scale has been modified, a cross-validation on a different sample is mandatory.

The descriptive, item loadings, scale reliability, and convergent validity criteria are presented in Table 3.

Item	Μ	SD	Loadings	Alpha	CR	AVE
AWS1	2.38	1.77	0.800	0.877	0.852	0.743
AWS2	2.57	1.83	0.857			
AWS3	2.62	1.91	0.865			
ANC2	2.74	1.92	0.790	0.844	0.845	0.732
ANC3	2.36	1.75	0.928			

Table 3. Descriptive, loadings and convergent validity (N=225).

The unidimensionality, scale reliability, and convergent validity criteria are also very good for the second sample. The mean values (a little bit lower than for the first

RoCHI 2016 proceedings

sample) suggest a moderate Facebook dependence and moderate negative effects.

The model testing results are presented in Figure 2. There is a significant (p<0.001) positive correlation between the Facebook dependence and the negative consequences on the work of students.



Figure 2. Model testing results (N=225).

The results revealed a very good fit of the model with the data: χ^2 =8.45, DF=4, *p*=0.076, χ^2 /DF=2.113, TLI=0.982, CFI=0.993, SRMR = 0.022, RMSEA=0.070.

The cross-validation results confirm the hypothesis on the second sample.

DISCUSSION AND CONCLUSION

The results of the two empirical studies show a moderate level of Facebook dependence. Clearly, the Facebook dependence has negative effects on the students' university work. The results of this study have some implications for researchers and practitioners.

Two constructs adapted from the literature have been used to measure the withdrawal syndrome and the negative consequences on the work. AWS has been validated on both samples and exhibits unidimensionality and very good convergent validity. As regards the second construct, only two items have been used in order to ensure a good psychometric quality.

Those findings come in correlation with the general trend already discussed in various research demarches and also spread in mass-media, being obvious that technology is mainly consumed in the format of social networks by students, and they spend most of their entire time staying on-line or plugged-in. The actual devices (smartphones, tablets) have an essential role to this fact. In this sense, it can be said that social networks became a kind of extension of the students' own *egos*. Practically, letting them without virtual networking seems to lead to losing of an important part of themselves.

Anyway, the Facebook dependence - measured with the withdrawal syndrome - is not the only determinant of the negative consequences on students' work. In a large amount, most of the interviewed students are connected to Facebook during the day. They admitted they had spent hours and hours taking a look to their friends' pictures, posting and answering to various comments, staying on chat and sending instant messages. Consuming a big amount of time on those activities means that the students do not have more on doing their tasks or preparing their exams. This time spent on Facebook has an important weight on the university students' time budget. Also, the

high frequency of use is a factor that distracts the students and negatively impacts their concentration on work.

ACKNOWLEDGMENTS

This work was supported by the Romanian grant financed by ANCS under COGNOTIC 1609 0101 / 2016.

REFERENCES

- Andreassen, C. S., Torsheim, T., Brunborg, G. S. and Pallesen, S. Development of a Facebook Addiction Scale. *Psychological Reports*, 110, (2012), 501–517.
- 2. Arbuckle, J. *AMOS User's Guide*. Amos Development Corporation, 2006.
- Balog A., Pribeanu C., Lamanauskas V. and Slekiene V. A multidimensional model for the exploration of negative effects of social networking websites as perceived by students. *Journal of Baltic Science Education*, 12 (3), (2013), 378-388.
- Caplan, S. E. Theory and measurement of generalized problematic Internet use: A two-step approach. *Computers in Human Behavior, 26*, (2010), 1089– 1097.
- Ellison, N. B., Steinfield, C. and Lampe, C. The benefits of Facebook "friends:" Social capital and college students' use of online social network sites. *Journal of Computer-Mediated Communication*, 12(4), (2007), 1143-1168.
- 6. Fornell, C. and Larcker, D. F. Evaluating structural equations models with unobservable variables and measurement error. *Journal of Marketing Research*, 18 (1), (1981), 39-50.
- 7. Griffiths, M. D. Internet addiction: fact or fiction? *Psychologist*, 12, (1999), 246-250.
- Griffiths, M. D. A "components" model of addiction within a biopsychosocial framework. *Journal of Substance Use*, 10, (2005), 191-197.
- Hahn, A. and Jerusalem, M. Die Internetsuchtskala (ISS): Psychometrische Eigenschaften und Validität (The Internet addiction scale: Psychometric characteristics and validity). In D. Mücken, A. Teske, F. Rehbein, & B. T. te Wildt (Eds.), Prävention. Diagnostik und Therapie von Computerspielabhängigkeit, (2010), 185–204.
- Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E. and Tatham, R.L. *Multivariate Data Analysis*. 6th ed., Prentice Hall, 2006.
- 11. Iordache, D.D., Pribeanu, C., Lamanauskas, V. and Raguliene, L. Usage of Facebook by university students in Romania and Lithuania: a comparative study. *Informatica Economica* 19(1), (2015), 46-54.
- 12. LaRose, R., Kim, J. and Peng, W. Social networking: Addictive, compulsive, problematic, or just another media habit? In Z. Papacharissi (Ed.), A networked self: Identity, community, and culture on social network sites. New York, NY: Taylor & Francis, 2010.
- 13. Manea, I. V., Gorghiu, G. and Iordache, D. D. The educational potential of Facebook use by students in

Romanian universities. *Revista Romana de Interactiune Om-Calculator*, 8 (3), (2015), 195.208.

- 14. Masur, P., Reinecke, L., Ziegle, M. and Quiring, O. The interplay of intrinsic need satisfaction and Facebook specific motives in explaining addictive behaviour on Facebook. *Computers in Human Behavior*, *39*, (2014), 376-386.
- 15. Orosz, G., Toth-Kiraly, I. and Bothe, B. Four facets of Facebook intensity – the development of multidimensional Facebook intensity scale. *Personality*

and Individual Differences, (2015) http://dx.doi.org/ 10.1016/j.paid.2015.11.03

- 16. Ross, C., Orr, E. S., Sisic, M., Arseneault, J.M., Simmering, M. G. and Orr, R. R. Personality and motivations associated with Facebook use. *Computers in Human Behavior*, 25(2), (2009), 578–586.
- 17. Wilson, K., Fornasier, S. and White, K. M. Psychological predictors of young adults' use of social networking sites. *Cyberpsychology, Behavior, and Social Networking, 13*, (2010), 173-177.

Evaluation of Tesys e-Learning Platform's Interface

Paul Ștefan Popescu¹, Marian Cristian Mihăescu¹, Mihai Mocanu¹, Costel Ionașcu²

¹Department of Computers and Information Technology,

University of Craiova, Romania

{spopescu, mihaescu, mocanu}@software.ucv.ro

²Faculty of Economics and Business Administration

University of Craiova, Romania

icostelm@yahoo.com

ABSTRACT

Online educational environments became more and more popular as a method for implementing distance learning programmes at universities. E-learning platforms serve as means of interaction for online educational environments and so they need to be well optimized for offering a better user experience. In this paper is presented the methodology and results obtained after the evaluation of the Tesys e-learning platform. The survey was taken by a group of students that saw the first time the e-Learning platform and they did several actions. After that they had to complete three surveys, one for every testing purpose targeted in the paper and we analyzed the results. The results offered us an overview on the actual state of the elearning platform; on one side there are still some improvements to be done and on the other some interfaces and controls were validated via this study.

Author Keywords

e-Learning platform; HCI; interface evaluation

ACM Classification Keywords

H.5.2. Information interfaces and presentation (e.g., HCI): User Interfaces.

General Terms

Human Factors; Design; Measurement.

INTRODUCTION

E-Learning platforms offer the environments that allow users to interact for learning purposes [4]. During the last decade more and more e-Learning platforms were developed and are used for distance learning programs. These online educational environments aim to lower the gap between e-Learning platforms and classical educational environments in terms of interaction as much as possible.

The core functionalities that are implemented in any online educational environment refer to learning resources management and evaluation techniques. Two main types of actors are always present in every e-Learning platform: students and professors. Each of these actors performs specific actions and needs to interact with others in order to accomplish their responsibilities.

At the University of Craiova, we use Tesys [9] as e-Learning platform at all of the distance learning programs from several faculties. This e-Learning platform was developed at our University and aims to fulfil all the needs for the distance learning programs. In Tesys we have three main actors: professors, students and secretaries. In this paper we aim to evaluate the usability of the interface from the professors and students view.

The interface of Tesys as designed more than 10 years ago and it aimed to fulfil specific needs from that period of time. Due to the continuous development process, the functionalities were upgraded over the time and now it offers most of the standard functionalities but the interface needs to be adjusted to the actual needs of both the students and professors. Not only the interface needed and evaluation but also the way the users interacts in the e-learning platform. Some of the design of the older functionalities still needs to be evaluated in order to see if it fits the most recent user's needs.

This paper presents a study conducted on the Tesys e-Learning platform using s group of users that had to test some functionalities seen for the first time and then fill three questionnaires. We aim to evaluate the interface in terms of usability [5] and find the key issues that need to be addressed on the next release in order to improve the user experience [6]. The paper describes the setup, the questionnaires and also the results obtained. These results will be used for redesign of the platform and improve functionalities in terms of usability.

RELATED WORKS

Evaluation in terms of HCI of e-Learning platforms is a constant problem referred in many papers. In 2005, Ardito et al. [1] presents a specific methodology for evaluating e-Learning applications. They specify that there are specific attributes for capturing the peculiar features of these applications and identifies them. There is also a preliminary user study involving a group of 'estudents' that were logged during their interaction with e-Learning platforms and a report is produced. They also propose specific evaluation patterns that are able to drive the evaluators in the analysis of e-Learning platforms.

Another paper that addresses the problem of the evaluation of the usability of e-Learning applications was written by M. F. Costabile et al [2]. They assume that there isn't a consolidated evaluation methodology for e-Learning platforms. The addressed problem is somehow common with the one addressed in this paper – the design of the interface that should provide a good usability so that the students interaction with the software are as natural as possible. We take this assumption and try to see using specific HCI evaluation techniques where do we need to work in order to get a more usable interface.

In a more recent paper [3], Vlado Glavinić and Andrina Granić, presents an overview of research being performed in the area of user interfaces for online educational environments. They address the efforts being done in providing suitable interaction for intelligent e-Learning platforms.

Design and evaluation of the e-Learning platforms was also addressed by Brad Mehlenbacher et. al. [7]. The paper's goal was to outline the challenges that face practitioners and researchers interested in usability and evaluation. They provide a brief overview and they share a heuristic tool that was developed for evaluation e-Learning environments and experiences.

Intuitive interfaces that imply flexible iteration suites different needs and becomes more useful for different kind of users and for several purposes [10]. These intuitive interfaces lead to user-cantered design [11] and user sensitive design [12] as the most appropriate methodology that was developed out of them. The aim is to be able to handle both uncommon users and typical users [13].

Other related research [14] present investigations on perceived satisfaction, behavioural intention and effectiveness of e-learning platform. In the study Blackboard e-learning platform is used and 424 university students were involved in answering a standard questionnaire. The obtained results showed that perceived self- efficacy is a critical factor that influences the satisfaction with the Blackboard e-learning system.

More general but also related research is presented in Evaluating E-Learning: A Case Study [15] where is presented the investigation over the experiences and perceptions of students who completed the study. The results presented in the paper, show implication for online learning design and future e-learning research.

METHODOOGY OF RESEARCH

For our experiments we used three questionnaires [8] that aims to test the communication module, the testing module and the interface from the professor and student views.

Figure 1 presents the evaluated modules from Tesys e-Learning platform. Each of the modules has his own survey and the results are presented in the next chapters of this paper.

Communication function embedded in Tesys covers a big part of the interaction between the entities that perform their activity on this e-Learning platform. The evaluation of the communication module is performed using 11 mixed questions. The aim of the survey is to find where we need to optimize the interface and the functionalities of the module. First questions aim to evaluate if there are problems with buttons and controls adjustment and the rest evaluate the functionalities of the module.

The module evaluation survey has 12 mixed questions that address both the interface and the functionalities. First we aim to evaluate if the questions from the testing procedure are well presented then we evaluate the controls. Both of them are important because we don't want learners to waste time looking for a misplaced control or to have difficulties in question presentation. Other important aspects refer the modality for computing grades and how often students got a question repeated (questions are automatically chosen for the test).



Figure 1. Tesys modules distribution.

The general interface evaluation survey has 14 questions divided in 7 for the student interface and 7 for the professor's interface. There are actually two interfaces for two different actors that are evaluated here. Every question from the student's interface and most from the professor's interfaces will get a mark from 1 to 10 regarding a specific control or functionality. On this survey we aim to find the best interface design in order to get the best usability results.

Group selection and setup

For this experiment we selected a group of 12 students from different years of study. No professors/tutors were involved in this study. The motivation for choosing them regards the academic results and a high level of trust. These students start using the platform for the first time and we aim to evaluate how well the e-Learning platform is optimized. The students are selected from different years of studies because the e-Learning platform is also used for every year of study from several faculties.

EXPERIMENTAL RESULTS

This section presents the results obtained from every questionnaire along with a short description. For every questionnaire we completed a table with the results which has on the columns the questions and on the rows the answers from every student that took part at the study.

For the first survey we used eleven questions and the results are presented in Table 1. Every question got a grade or a letter corresponding to a specific answer. The last three questions that got the same letter: a, were

referring if the dimension of the field of the message is big enough (question 9), if the dimension of the title of the message is big enough (question 10) and if the overall experience is good enough.

Communication surv	vey
--------------------	-----

N o	Q 1	Q 2	Q 3	Q 4	Q 5	Q 6	Q 7	Q 8	Q 9	Q 10	Q 11
1	0	a	5	9	9	d	6	b	a	a	a
2	b	b	5	8	9	d	5	b	a	a	a
3	с	a	9	9	9	b	8	с	a	a	a
4	d	a	7	10	10	b	9	a	a	a	a
5	a	a	8	10	10	b	10	c	a	a	a
6	a	a	9	9	10	b	10	a	a	a	a
7	b	a	10	9	9	b	10	a	a	a	a
8	d	a	9	1	10	b	1	b	a	a	a
9	b	a	9	8	10	b	9	b	a	a	a
10	a	a	7	8	7	b	0	a	a	a	a
11	a	a	10	10	10	b	9	b	a	a	a
12	a	a	8	10	8	b	7	a	a	a	a

Table 1. Results for communication survey.

The communication module has two types of questions: grade driven and free choice (multiple complements answer). We present here just a example result for question one but we computed a chart like the one from Figure 2 for every question that had a multiple complement answer. Three of these questions got just one answer and the conclusion is that the fields dimension and the experience referring the communication module is a good one and we don't need to get any modifications.



Figure 2. Question 1: How often do you use the communication function.

The placement of the "Communication" button is good, only 8% of the questioned students didn't found the button quick at first sight. Taking into consideration question 6 we can see that some students had problems at sending a message to professors. After a problem analysis over the question 6 we found that there are some software implementation bugs and we need to make further tests and solve it.

From the second category we have only two questions which show grades above 8 for buttons placement on the interface. In conclusion, the buttons are well placed and the user's experience with this module is a good one. The overall analysis reveals some problems that need to be solved but the usability is pretty good based on the answer from question 11.

Testing survey

Table 2 presents the results obtained from the testing survey. Just like in the case of the communication survey, for the testing module we the same two types of questions. For the questions 1,3,4 and 6 which have multiple complement answers and these questions got a mark greater that 9 so there is no need for improvement on the verified aspects. From the question number 8 we conclude that we must optimize further the students searching results but a period of time close to 3s for searching professors is good enough taking into consideration that the functionality was first seen.

One aspect that was also important was that we need to adjust the checkboxes dimension and more exactly they need to be enlarged.

No	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q
	1	2	3	4	5	6	7	8	9	10	11
1	10	а	10	10	3	a	с	5	10	yes	b
					3-						
2	9	b	10	10	5	a	a	8	9	yes	c
3	10	b	9	9	2	a	a	2	10	yes	с
4	10	а	10	10	2	b	с	2	10	no	с
5	8	а	10	8	3	b	a	4	9	yes	a
6	7	b	10	8	3	a	a	2	8	yes	a
7	9	с	10	9	2	b	a	1	9	no	d
8	10	а	10	10	.3	b	а	10	10	yes	с
9	8	а	10	10	.2	а	с	1	10	yes	c
10	10	b	10	9	10	b	с	1	10	yes	с
11	9	a	10	10	1	b	a	1	9	yes	a
12	8	b	10	10	9	b	a	4	10	yes	b

Table 2. Results for testing survey.

Las two quesrions refered the possibility to hack the testting procedure. This procedure is crucial for this module and we wanted to see if any of the students finds a procedure. We left them more time to see but no results. There is no question 12 in the table because the resuld sould have been a short description of the procedure. As a short example of question from the testing survey we choose the question no. five which validate the dimension of the check-boxes for the "answer" button. This question had a variety of results and the results may be somehow

unclear but taking into consideration the context of the question we can set a threadshold over the time spent for answering a question.



Figure 3. Question 3: How many seconds do you need to find the check-boxes for the answer button.

A student can't spend too much time finding the checkbox related to the correct answer so we certianly need to improve the dimensions. The moviation for taking such a conclusion is that in most of the cases he has 30s for giving the answer and 3s or bigger time means at least 10% of the total available time.

For Student							
No.	Q.1	Q. 2	Q. 3	Q. 4	Q. 5	Q. 6	Q. 7
1	10	10	10	10	10	6	9
2	10	10	10	10	10	8	10
3	10	10	10	9	10	7	9
4	10	10	10	10	10	1	9
5	10	10	10	10	10	9	10
6	10	10	10	10	10	7	9
7	10	10	10	10	10	8	9
8	10	10	10	10	10	8	9
9	10	10	10	10	10	7	10
10	10	10	10	10	10	5	10
11	4	10	10	10	10	3	9
12	10	10	10	10	10	9	9

Professor and student survey

Table 3. Results for student's interface survey.

The last survey evaluated the buttons and controls placement for both student and professors. We have 7 questions that refer students, each of them requesting a grade from 1 to 10. Excepting the question number 6 that got an average grade a bit lower than 6, every question from this set got an average grade close to 10. The motivation for getting a lower grade on question 6 is the big number of clicks necessary for getting to the homework.



Figure 4. Average grade per question.

Figure 4 presents a chart with the results for every question from this interface. It helps us to have an overview of the results and to see on what items (in our case we take into consideration question one and six) we need to pay attention.

For professors the grades were also big, but some of the participants considered that learning resources need to be reviewed without a specific request. In this case we didn't found any improvements to do but we can take this survey as a validation of the professor's interface.

For Professor							
No.	Q. 1	Q. 2	Q. 3	Q. 4	Q. 5	Q. 6	Q. 7
1	a	a	10	10	9	10	10
2	a	a	10	10	9	10	10
3	a	a	10	10	8	10	9
4	a	a	10	10	10	10	10
5	a	b	10	10	10	10	10
6	a	b	10	10	10	10	9
7	a	a	10	10	9	10	10
8	a	a	10	10	10	10	10
9	a	0	10	10	10	10	7
10	a	a	10	10	10	10	10
11	a	a	10	10	10	10	10
12	a	a	10	10	8	9	9

Table 4. Results for professor's interface survey

There always can be a reorganize of the learning resources in order to see if there can be better results and then we aim to run the evaluation again.

CONCLUSION AND FUTURE WORK

The results obtained from the study described in this paper provide the feedback necessary for further development of the e-Learning platform. We presented the experimental setup, every survey that was used in this study and interpreted the results.

As final conclusions, that take into consideration the overall results, Tesys is well structured and the user experience is quite good especially for users that never used it before. The time needed to get used to the online educational environment is small (e.g., 3s as an average for most of the questions) and the usability is good. There

are also some bugs revealed as secondary effects which are very helpful because the e-learning platform is continuously developed are some new bugs can occur anytime.



Figure 5. Question 2: Do you consider that the organizing of the learning resources is good enough?

Conducting the study revealed also some important problems that need to be addressed like reducing the number of clicks necessary for getting to homework from the students' view. Sending a message to the professors from the student's view is another interface problem that needs to be addressed because the button is sometimes hard to find and needs to be repositioned. Another important conclusion is regarding the check-boxes' dimension that are not big enough and need to be enlarged. As future work we take into consideration to try to solve the issues revealed by the surveys and then redo it in order to see if there is a results improvement. There are also some other modules that needs to be evaluated and optimized for better user experience.

A design of an adaptive interface that will suit better every type of user is also an important future work. The need for this kind of interface arises from the difference between users from different faculties that use the Tesys e-Learning platform.

REFERENCES

- Ardito, C., Costabile, M. F., Marsico, M. D., Lanzilotti, R., Levialdi, S., Roselli, T. and Rossano, V. An approach to usability evaluation of e-learning applications. *Universal Access in the Information Society*, (2006), 270-283.
- Costabile, M. F., De Marsico, M., Lanzilotti, R., Plantamura, V. L. and Roselli, T. On the Usability Evaluation of E-Learning Applications. *Proceedings* of the 38th Annual Hawaii International Conference on System Sciences, Big Island, HI, USA, (2005), 6b-6b. doi: 10.1109/HICSS.2005.468.
- 3. Glavinić, V. and Granić, A. HCI Research for E-Learning: Adaptability and Adaptivity to Support Better User Interaction. HCI and Usability for Education and Work: 4th Symposium of the

Workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society, USAB 2008, Graz, Austria, November 20-21, 2008.

- Pantazis, C. Executive summary: A vision of Elearning for America's workforce. Report of the Commission on Technology and Adult Learning, ASTD (2001) (last accessed: 2007-08-17), http://www.learningcircuits.org/ 2001/aug2001/pantazis.html
- 5. Höök, K. Evaluating the Utility and Usability of an Adaptive Hypermedia System, Knowledge-Based Systems 10(5), (1998), 311–319.
- Granić, A. and Nakić, J. Designing intelligent interfaces for e-learning systems: the role of user individual characteristics. In Stephanidis, C. (ed.) HCI 2007. Springer, Heidelberg, LNCS 4556, (2007), 627–636.
- Mehlenbacher, B., Bennett, L., Bird, T., Ivey, M., Lucas, J., Morton, J., Whitman, L. Usable E-Learning: A Conceptual Model for Evaluation and Design. Proceedings of HCI International 2005: 11th International Conference on Human-Computer Interaction, 4, (2005), 1-10.
- Surveys used in this study: https://docs.google.com/ document/d/14JePXBrwz14MPVnc3NERn7HlywqY XIUvhPmTzK2GNxg/edit?usp=sharing.
- Burdescu, D.D., Mihaescu, M.C. and Jehøj, H. TESYS: e-Learning Application Built on a Web Platform. ICE-B 2006 - Proceedings of the International Conference on e-Business, Setúbal, Portugal, August 7-10, (2006).
- Benyon, D., Crerar, A. and Wilkinson, S. Individual Differences and Inclusive Design. In: Stephanidis, C. (ed.) User Interfaces for All – Concepts, Methods, and Tools, Lawrence Erlbaum Associates, Inc., Mahwah (2001), 21-46.
- 11. Norman, D. and Draper, S.W. (eds.) User Centered System Design. Earlbaum, Hillsdale, 1986.
- Gregor, P., Newell, A.F. and Zajicek, M. Designing for dynamic diversity – interfaces for older people. In: Jacko, J.A. (ed.) ASSETS 2002. 5th International ACM Conference on Assistive Technologies, Edinburgh, Scotland, (2002), 151-156.
- Pullin, G., Newell, A. Focussing on Extra-Ordinary Users. In: Stephanidis, C. (ed.) *HCI 2007*. Springer, Heidelberg LNCS 4554, (2007), 253–262.
- 14. Liaw, S. S. Investigating students' perceived satisfaction, behavioral intention, and effectiveness of e-learning: A case study of the Blackboard system. *Computers & Education*, 51(2), (2008), 864-873. http://dx.doi.org/10.1016/j.compedu.2007.09.005
- Cappel, J.J. and Hayen, R.L. Evaluating E-Learning: A Case Study. Journal of Computer Information Systems, 44(4), (2004), 49-56. http://dx.doi.org/10.1080/08874417.2004.11647595

MOOCBuddy: a chatbot for personalized learning with MOOCs

Carmen Holotescu

University "Ioan Slavici" of Timișoara 144 Str.Păunescu Podeanu, 300569 Timișoara, Romania carmenholotescu@gmail.com

ABSTRACT

With the proliferation of MOOCs (Massive Open Online Courses) providers, like Coursera, edX, FutureLearn, UniCampus.ro, NOVAMOOC.uvt.ro or MOOC.ro, it's a real challenge to find the best learning resource. MOOCBuddy – a MOOC recommender system as a chatbot for Facebook Messenger, based on user's social media profile and interests, could be a solution. MOOCBuddy is looking like the big trend of 2016, based on the Messenger Platform launched by Facebook in the mid of April 2016.

Author Keywords

MOOCs; chatbot; Messenger platform; personalized learning; recommender system.

ACM Classification Keywords

Information interfaces and presentation. Miscellaneous.

General Terms

Design.

INTRODUCTION

Started in 2008, the new Massive Open Online Courses (MOOCs) paradigm has brought innovations at all levels of education, aiming to respond to the most pressing learning needs, generated by the new development policies and rapid evolution of the technology. With hundreds of new MOOCs which appear monthly, it is a real challenge to find the best resource fitting learner's personal profile, interests, background, and learning and development needs.

The paper reports on the development and features of a chatbot named MOOCBuddy, acting as a MOOC Recommender System based on user's social media profile and interests. The chatbot is based on the Facebook Messenger Platform launched by Facebook in the middle of April 2016, the philosophy of this platform and its potential for educational uses being also presented.

CHATBOTS TREND AND FACEBOOK MESSENGER PLATFORM

Chatbots definition

Chatbots can be considered artificial narrow intelligent (ANI) programs designed to interact with users in a human-like way, answering questions and performing tasks, in a specific area (Chase, 2016).

Developed by different companies, the existing chatbots constitute a paradigm shift in how we interact with

technology, being specialized in areas such as (Sharma, 2016): customer support, transactions and helpdesks (Kasisto, Amazon Echo, Google Home, Microsoft's Cortana); smart wallets (wallet.AI); (advanced) data analytics; repetitive tasks; automated virtual assistants.

Facebook Messenger Platform

During the last years, being represented by chat-based applications like WhatsApp, WeChat, Telegram and Facebook Messenger, messaging has become the most widely used communication layer on the mobile platforms.

Combining the trends of artificial narrow intelligence (ANI) and messaging applications, on April 12, during the F8 conference, Facebook has launched the Messenger Platform (Beta) with chatbots and the Send/Receive API (Facebook, 2016a).

In the vision of Facebook, for the 900 million people and 50 million businesses that use Messenger monthly, the chatbots are seen as massive opportunities: by interacting directly with the people, the chatbots can provide automated subscription content (weather, news, traffic updates, sport scores), but also customized communications (receipts, shipping notifications, bookings, e-commerce guidance, interactive experiences).

For building chatbots, the Messenger Send/Receive API offers support for: defining a welcome screen for setting the context and different controls; sending and receiving text, images and interactive rich bubbles containing multiple calls-to-action; integration with the Wit.ai's Bot Engine for interpreting intent from natural language.

As Facebook has stated, people are putting first in the platform philosophy, meaning that users have the possibility to block the communication with the chatbots found unuseful, also each chatbot is reviewed by the company before becoming functional.

At less than a month from the Messenger Platform launch, Facebook estimated that tens of thousands of developers are building chatbots, being supported by the platform guidelines section (Facebook, 2016b). Also there are wizard tools to create (simple) chatbots such as botsify.com.

Meanwhile Facebook is working to implement a specific analytics system, in order to monitor chatbots and to prevent spam, bulk delivery or annoying messages (Matney, 2016).

It is not an easy task to discover new chatbots, even if the Messenger searching feature displays the results in two different categories: More People, and Bots and Businesses. Also there were created directories for bots, such as botarena.co, botlist.co, botpages.com or chatbots.org. Bot hunter bot (@bothunterbot) is a bot curator chatbot, sending a list of new bots each week.

Chatbots for education

Exploring the above directories for chatbots and studying the current articles on this technology, one can note that education is seldom discussed as a domain for which chatbots could be built.

Interesting exceptions are the chatbots for UNESCO UReport project (@ureportglobal) or for displaying a product marketer's resume (@helloestherbot) (Crawford, 2016).

It's worth to mention also Jane Hart's blog note written in May, about how chatbots can be incorporated in Professional Ecosystems, with features for relevant or customized information such as (Hart, 2016): search for content or courses on YouTube, Wikipedia or Coursera; receive news from news sources, blogs feeds or Twitter accounts; receive productivity support, such as alerts or reminders; have an intelligent personal assistant.

MOOCS SEARCHING CHALLENGES

Started in 2008, the new Massive Open Online Courses (MOOCs) paradigm has brought innovations at all levels of education, aiming to respond to the most pressing learning needs, generated by the new development policies and rapid evolution of the technology. Each month hundreds of new MOOCs hosted on different platforms, addressing a large category of topics and with varying durations, are developed and offered to participants worldwide, some of them counting for university credits or verified certificates. Also some of the MOOCs could be packed in nanodegrees (Udacity), specializations (Coursera), xseries (Edx) or sequences (FutureLearn).

In Romania there are many initiatives related to MOOCs, the most of them being monitored and curated by the author (Holotescu, 2012): platforms and MOOCs were implemented (unicampus.ro, novamooc.uvt.ro, mooc.ro, unibuc-virtual.net, estudent.ro, eliada.ubbcluj.ro/proiect, udemy.com/management-ong); experiments for integrating MOOCs in blended academic courses can be found at University Politehnica Timisoara and University "Joan Slavici" of Timisoara (Holotescu et al., 2014; Vasiu and Andone, 2014); workshops and national conferences related to opening education, many of them being organized by the members of the Romanian Coalition for OERs.

With the MOOCs rapid developments worldwide, it is a real challenge to find the best resource fitting learner's personal profile, interests, background, and learning and development needs.

The existing MOOCs directories or portals such as MOOC List (mooc-list.com), Class Central (classcentral.com) or Open Education Europa (openeducationeuropa.eu) include features for searching 92 based on different criteria. We should mention that the Romanian projects are not indexed by them.

The Class Central portal provides the possibility to follow a specific MOOCs platform or topic, the information being sent as a monthly e-mail.

Also there are studies related to MOOCs recommender systems design: Rădoiu (2014) argues that to be effective such a system must focus on MOOCs specific type of items (learning items) and user behavior in MOOCs context. MOOC-Rec desktop application was designed by Bousbahia and Chorfia (2015), the most appropriate MOOCs being proposed using the Case Based Reasoning (CBR) approach. The cognitive level, knowledge background, personal expectation, learning interest, learning motivation and learning style of students are considered in the undergraduate-oriented recommender system of MOOCs analyzed by Fu et al. (2015).

MOOCBUDDY CHATBOT

Design

In March 2016, the idea to build a chatbot as a MOOC recommender system was sent by the author to the HackTM organizers, for the education category, the proposal being published on the event website (Figure 1). At that moment there were only rumors about Facebook M and about future publishing of the Messenger Chat SDK (API), and a reduced number of chatbots exist, such as Assist (Constine, 2016).

MOOCBuddy

Befriend with MOOCs. With the proliferation of MOOC providers (Massive Open Online Courses, like Coursera, edX, FutureLearn), it's a real challenge to find the best learning resource. MOOCBuddy - a MOOC Recommender System as a Chat Bot for Facebook Messenger, based on user's social media profile and interests, could be a solution. MOOCBuddy is looking like the big trend of 2016, with the awaited virtual assistant Facebook M or the Assist application.

Figure 1. MOOCBuddy idea on the HackTM event website (hacktm.ro/#!/categories).

The alpha version of MOOCBuddy was implemented in May, for the HackTM contest. During the next weeks, the chatbot was improved and tested. New developments are already specified as a result of the chatbot interaction with users and following the received feedback. Based on our research, having the proposal published in March and the implementation in May, MOOCBuddy seems to be the first educational chatbot related to MOOCs.

One of the main aim of the chatbot is to promote the Romanian MOOCs initiatives, which were stored in an updated database. Thus the chatbot recommendations consist in items of this database and in links to specific searches in MOOCs directories presented above.

MOOCBuddy can assist anyone to discover (news about) MOOCs, individual learners to find MOOCs for their personal and professional development, but also teachers who intent to integrate MOOCs in their courses.

Features

MOOCBuddy is designed to be interactive, friendly and to facilitate the discovery of the MOOC paradigm, and the connections with MOOCs and platforms.

The dialogues with the users are modelled as structured messages with multiple bubbles rendered as a horizontal list. Bubbles containing information, images, buttons that open a URL or receive a choice are provided to users.

Based on the user's choice, the chatbot displays an introduction to MOOCs and hosting platforms, further information and platforms browsing being accessible by clicking URL buttons (Figure 2).



Figure 2. Screenshot with MOOCBuddy interaction.

MOOCBuddy offers MOOCs search by a large category of characteristics such as topics, language, start date, duration, accreditation, facilitators and university running the MOOC.

The findings can be shared by user on other social media platforms, commented and recommended to friends, this way MOOCBuddy and also the MOOC paradigm are disseminated.

Weekly alerts about new interesting MOOCs and news about the domain are sent. If these are considered intrusive, a user can block such messages.

Tests

Similar with all the other chatbots, MOOCBuddy has a dedicated Facebook page (facebook.com/mymoocbuddy), customized by the developer and there are more options for a user to enter in communication with it: by clicking on Send Message on the page; by entering the address m.me/MOOCBuddy in a (mobile) browser; by scanning the corresponding Messenger Code.

MOOCBuddy was tested by students at University Politehnica Timisoara and University "Ioan Slavici" of Timisoara, who have participated in the courses facilitated by the author and are already familiar with MOOCs. Also members of the Romanian Coalition for Open Educational Resources have interacted with the chatbot.

There are other users who liked the Facebook page and have communicated with the chatbot, MOOCBuddy having almost one hundred users after four weeks from its development.

MOOCBuddy learns continuously together with its users, improving its scenarios and assessing users' needs and satisfaction: it rates user experience, displaying values to be clicked, also registers proposals for new features as open messages sent by users which are stored in the developer panel or as comments in the associated Facebook page.

CONCLUSIONS AND FUTURE WORK

Implemented a month after the Messenger Platform launch, MOOCBuddy seems to be the first educational chatbot related to MOOCs, assisting users to discover news about MOOCs, to find MOOCs for personal and professional development, but also teachers to integrate MOOCs in their courses. Moreover, MOOCBuddy is an innovative project aiming to make known the Romanian initiatives related to MOOCs.

The chatbot creates an informal bond with its users, facilitates ubiquity learning and a storytelling interaction. As Crawford (2016) noted: "The desire to chat creates an opportunity for interactive storytelling. Use it to your advantage".

A future research direction will be the improvement of the recommender algorithm based on a larger category of user's characteristics (background including the recommended MOOCs previously followed and the participation rates, learning expectation and style, profiles on different social networks), also on the user's history and interaction with the application. We intent to build a standalone recommender system with a specific API, which will be used by MOOCBuddy, but also by other web or mobile applications.

As possible business models or partnerships, MOOCBuddy could implement recommendations for specific platforms and MOOCs, and also personalized alerts for MOOCs to be followed for new knowledge/competencies needed for jobs/internships.

ACKNOWLEDGMENTS

We thank Victor Holotescu, student at University Politehnica Timisoara, Romania, for the collaboration and work in implementing MOOCBuddy.

REFERENCES

- 1. Bousbahi, F. and Chorfi, H. MOOC-Rec: A Case Based Recommender System for MOOCs. *Procedia-Social and Behavioral Sciences*, 195, (2015), 1813-1822.
- Chase, M. Chatbots friendly or frightening? Note on ICX Association platform. (2016) Retrieved from http://icxa.org/2016/05/chatbots-friendly-orfrightening/.
- Constine, J. Facebook's Secret Chat SDK Lets Developers Build Messenger Bots. Note on Techcrunch. (2016) Retrieved from http://techcrunch.com/2016/ 01/05/facebookmessenger-bots.
- 4. Crawford, E. Bots are awesome! Humans? Not so much. Chatbots Magazine. (2016). Retrieved from https://chatbotsmagazine.com/bots-are-awesome-humans-not-so-much-7b2d62630668.
- 5. Facebook. (2016a). Messenger Platform at F8. Facebook Newsroom. Retrieved from http://newsroom.fb.com/ news/2016/04/messengerplatform-at-f8.

- 6. Facebook. (2016b). Messenger Platform Guidelines. Facebook for Developers. Online at https://developers. facebook.com/docs/messenger-platform.
- Fu, D., Liu, Q., Zhang, S. and Wang, J. The Undergraduate-Oriented Framework of MOOCs Recommender System. *In 2015 International Symposium on Educational Technology* (ISET), IEEE (2015), 115-119.
- Holotescu, C. OER in Romania. POERUP Project: Policies for OER Uptake Report. (2012, updated 2016). Available at http://poerup.referata.com/wiki/Romania.
- Holotescu, C., Grosseck, G., Cretu, V. and Naaji, A. Integrating MOOCs in Blended Courses. In Proc. of 10th International Conference eLSE, Bucharest, 24-25 April 2014.
- Hart, J. The Future of Work and Learning 2: Chatbots. Note on Learning in the Modern Workplace. (2016). Retireved from http://www.c4lpt.co.uk/blog/2016/05/ 15/future-of-work-and-learning-2-bots.
- 11. Matney, L. Facebook says 10K+ developers are building chatbots, analytics are coming. Note on Techcrunch. (2016). Retrieved from http://techcrunch.com/2016/05/ 10/facebook-chatbotanalytics.
- 12. Rădoiu, D. Organization and constraints of a recommender system for MOOCs. Scientific Bulletin of the "Petru Maior" University of Tîrgu Mureş, Romania, 11, (2014), 51-59.
- Sharma, G. Artificial Intelligence (AI) and FinTech. Chatbots Magazine. (2016). Retrieved from https://chatbotsmagazine.com/artificial-intelligenceai-and-fintech-part-1-7cae1e67dc13.
- 14. Vasiu, R. and Andone, D. OERs and MOOCs The Romanian experience. *In Web and Open Access to Learning* (ICWOAL), International Conference, (2014).

ReaderBench goes Online: A Comprehension-Centered Framework for Educational Purposes

Gabriel Gutu, Mihai Dascalu, Stefan Trausan-Matu

University Politehnica of Bucharest

313 Splaiul Independentei, Bucharest, Romania gabriel.gutu@cs.pub.ro, mihai.dascalu@cs.pub.ro, stefan.trausan@cs.pub.ro

Philippe Dessus

Univ. Grenoble Alpes, LSE F-38000 Grenoble, France philippe.dessus@univ-grenoblealpes.fr

ABSTRACT

In this paper we introduce the online version of our ReaderBench framework, which includes multi-lingual comprehension-centered web services designed to address a wide range of individual and collaborative learning scenarios, as follows. First, students can be engaged in reading a course material, then eliciting their understanding of it; the reading strategies component provides an in-depth perspective of comprehension processes. Second, students can write an essay or a summary; the automated essay grading component provides them access to more than 200 textual complexity indices covering lexical, syntax, semantics and discourse structure measurements. Third, students can start discussing in a chat or a forum; the Computer Supported Collaborative Learning (CSCL) component provides indepth conversation analysis in terms of evaluating each member's involvement in the CSCL environments. Eventually, the sentiment analysis, as well as the semantic models and topic mining components enable a clearer perspective in terms of learner's points of view and of underlying interests.

Author Keywords

Sentiment analysis; Semantic models; Topic mining; Automated essay grading; Reading strategies; Computer Supported Collaborative Learning; *ReaderBench* framework.

ACM Classification Keywords

I.2.7 [Natural Language Processing]: Discourse, Language parsing and understanding, Text analysis.

General Terms

Algorithms, Design, Languages.

INTRODUCTION

Knowledge understanding from texts, either read or written, are crucial in education-centered contexts. Technology has gained a broader usage and more tools designed to support tutors and learners alike in the learning process are being made available nowadays. Thus, a huge amount of content is being generated by teachers who share their learning materials, or by students who provide feedback, do tests, homework or are involved in online conversation.

Natural Language Processing (NLP) techniques [1] have gained considerable ground lately as they provide accurate and efficient analyses of both written and oral language. Advanced NLP services are being developed, including the analysis of unstructured learning materials of students' textual traces, automated essay grading, sentiment analysis, concept map elaboration or identification of reading strategies. Our framework, ReaderBench [2, 3, 4, 5], comprises of advanced NLP techniques used to expose a wide variety of language services. We can consider our framework as being unique as it provides a unitary core engine centered on cohesion and on dialogism [6, 7], the latter being reflected in the implemented polyphonic model [8]. Multiple connected services addressing different facets of comprehension assessment and prediction are thus deployed. Tutors are capable to perform an apriori assessment of learning materials, but also to evaluate a posteriori learner's written traces consisting of essays, self-explanations or utterances in CSCL conversations. All these services are described in detail in subsequent sections.

A client-site web application for our framework was being developed within the H2020 RAGE (Realising and Applied Gaming Eco-System) project, covering most back-end *ReaderBench* functionalities, and is currently available online at http://readerbench.com. Figure 1 depicts the main interface of the website.

This paper presents an overview of the online version of our framework regarding the services currently made available. Enhanced functionalities are still under development, while some web services were specifically implemented to meet RAGE partner requirements. A full web version that enables a holistic analysis of texts in general and of CSCL conversation, similar to the desktop application, will be made available in the foreseeable future.

In terms of structure, the second section introduces the overall *ReaderBench* architecture, while the third section presents in detail all language services that are currently published online. The fourth section presents specific use cases, as well as conclusions and future work.

ARCHITECTURE

The *ReaderBench* framework integrates a wide variety of advanced NLP techniques centered on comprehension assessment and prediction and is built around Cohesion Network Analysis [9]. *ReaderBench* has introduced a multi-lingual and automated model applicable to various types of texts, such as essays, self-explanations or conversations in Computer Supported Collaborative Learning (CSCL) environments and represents a framework that aims to reach targeted education purposes. Therefore, a variety of linguistic features

RoCHI 2016 proceedings

important for understanding texts and predicting learners' comprehension are made available. These include sentiment analysis, textual cohesion and textual complexity. In terms of inputs, besides plain text, some services use PDF files from which the extracted raw text is sent for processing. Other types of inputs, such as *Word* documents or RTF files will be considered in the nearest future.

As an overview, the *ReaderBench* framework makes use of the Standard Core NLP [10] for implementing natural

language processing pipelines consisting of the following processes [1]: tokenization, sentence splitting, part of speech tagging, lemmatization, named entities recognition, dependency parsing, and co-reference resolution. Whereas for English the full pipeline is supported, for other languages (e.g., French, Spanish, Italian, Romanian and Dutch) only the core steps are being performed. In addition, ReaderBench includes multiple libraries such as Mahout Apache (http://mahout.apache.org/), Gephi (http://gephi. org/), and Mallet (http://mallet.cs.umass.edu/).



Figure 1. ReaderBench main web interface.

Cohesion is evaluated from multiple perspectives within the framework [11] in terms of semantic distances in lexicalized ontologies (e.g., WordNet, WOLF for French) [12], Latent Semantic Analysis (LSA) [13], and Latent Dirichlet Allocation (LDA) [14] semantic models. The models were trained on specific text corpora. Some of the corpora used for English language include Touchstone Applied Science Associates, Inc. corpus (TASA) (http://lsa.colorado.edu/spaces.html), the LAK dataset [15], or the Contemporary American English collection (COCA) [16]. Some of the texts used for French language include the Texts Enfants collection [17] and "Le Monde" corpus (http://lsa.colorado.edu/spaces.html). Figure 2 depicts the five most important components being included within the framework. The underlying services will be further described in the next sections.



Figure 2. ReaderBench architecture.

LANGUAGE SERVICES

Automated Identification of Reading Strategies

Five major components are currently presented within the web interface of the *ReaderBench* framework. Each component presented below is currently available as REST web services, and can be integrated in custom applications. JSON format is used for both sending data and accepting responses for the majority of our web services.

Identification of reading strategies is a recognized predictor in determining the reading comprehension of students [18]. This component is also available on the *ReaderBench* website and it can be used to automatically identify metacognition, causality, bridging, paraphrasing and elaboration strategies used by a learner within their self-explanation [19].

Text input	Ce soir-là, la famille de Matilda dinait comme d'habitude devant la télévision. Ils entendirent une voix forte venant du salon dire : « salut, salut salut salut »La mère devint toute blanche. Elle dit à son mar i « il y a quelqu'un dans la maison ». Ils arrêtèrent tous de manger. Ils étaient tous sur le qui-vive. La voix reprit « salut, salut »Le frère se mit à crier « ça recommence ! ». Matilda se leva et alla éteindre la télévision. La mère, paniquée, dit à son mari : « Henri, des voleurs, ils sont dans le salon, tu devrais y aller ». Le père, raide sur sa chaise ne bougea pas. Il n'avait pas enviet de jouer au héros. Sa femme lui dit : « Alors, tu te décides ? Ils doivent être en train de faucher l'argenterie ! ». Moniseur Verdebois s'essuya nerveusement les lèvres avec sa serviette et proposa d'aller voir tous ensemble. La mère attrapa un tisonnier au coin de la cheminée. Le père s'arma d'une
READING au début c' commence vient[0.796 donc ils y w commence au bout d' u cheminée c dire-0.714	est une famille de riches qui mangent et puis ils entendent une phrase avec une voix grave salut salut . donc les parents ht[0.682; aller-0.702*] à avoir peur. donc la mère de la fille[0.332; mari-0.836*] elle croit que c'est un voleur qui ; aller-0.732*] leur voler leur argent leur argenterie. et du coup elle dit au père d' y aller mais il veut[0.688; dire-0.818*] pas ant tous ensemble. et la fille[0.332; mari-0.836*] elle joue bien la comédie parce que ça redit salut salut salut et elle à sursauter. et puis un moment elle leur raconte l' histoire[0.423; raconter-0.732*] de fantômes donc ils y croient. et puis n moment elle arrête la plaisanterie en allant chercher[0.638; savoir-0.786*] le perroquet qu' elle avait mis au - dessus de la ju' elle avait emprunté à son copain arthur. et du coup ben ils eurent plus jamais de fantômes chez eux. c' était juste[0.468;] pour jouer un tour à ses parents.
Bridged ele ils entendin il n'avait pas monsieur ve ils sursautè c'est alors c	ments: ent une voix forte venant du salon dire : « salut, salut, salut » - Cohesion: 0.513; ; envie de jouer au héros - Cohesion: 0.552; erdebois s'essuya nerveusement les lèvres avec sa serviette et proposa d'aller voir tous ensemble - Cohesion: 0.474; rent tous, y compris matilda qui jouait très bien la comédie - Cohesion: 0.443; ju'elle sortit du manteau de la cheminée le perroquet de leur copain arthur - Cohesion: 0.465;
16	Paraphrase
8	Causality
5	Text based Inferences
13	Bridging
7	Inferred Knowledge
0	Metacognition

Figure 3. Sample input data for ReaderBench self-explanation service and automatically identified reading strategies.

Further analyses consider the usage of textual complexity indices in order to improve the accuracy in terms of comprehension prediction [20]. Figure 3 depicts a different sample input for French language. Based on a given target text, learners self-explain what they understood and specific employed reading strategies are automatically identified.

Textual Complexity Assessment

Automated essay grading represents a technique used to reduce tutor's workload by offering specific analyses and statistics regarding students' writing style. The model for textual complexity assessment, centered on cohesion and integrated in the *ReaderBench* framework, represents the foundation for a multi-dimensional analysis on writing styles. The generated indices support tutors in identifying improvements that can be done on each student's essay and enable an objective evaluation of students by offering them automatically generated feedback, which has a positive impact on writing style quality [21].

RoCHI 2016 proceedings

Some of the complexity indices reflected through our web service include statistic surface indicator (e.g., average paragraph, sentence or word lengths, number of commas, word and character entropy), syntax factors (statistics on different parts of speech, average number of first, second or third person pronouns per paragraph, depth of parsing tree), semantic cohesion (intra- and inter- paragraph and sentence cohesion scores computed using Wu-Palmer semantic distance over WordNet [22], LSA and LDA).

TEXTUAL C	TEXTUAL COMPLEXITY			
SURFACE FACTORS (AVERAGE LENGTHS IN CHARACTERS)				
415.667	Average paragraph length (characters)			
124.7	Average sentence length (characters)			
5.567	Average word length (characters)			
2.864	Standard deviation for words (characters)			
SURFACE FACTORS (STATISTICS)				
SURFACE FACTORS (ENTROPY)				
4.617	Word entropy			
2.917	Character entropy			
SYNTAX (PRONOUNS)				
DISCOURSE FACTORS (CONNECTIVES) (ENG)				
DISCOURSE FACTORS (COHESION-BASED SCORING MECHANISM)				
SEMANTIC COHESION (WU-PALMER)				
SEMANTIC COHESION (LSA)				
SEMANTIC COHESION (LDA)				

Figure 4. Textual complexity results computed for the sample input data.

Choosing an appropriate text for students, neither too simple nor too difficult to understand, represents an important task in the learning process. The indices provided by our tool are an important component when it comes to adapt learning materials for specific students. Valuable feedback can be retrieved by analyzing and combining the previous textual complexity indices alltogether, thus supporting comprehension both a priori during text selection, as well as a posteriori during automated feedback generation. Figure 4 shows the textual complexity index scores obtained for the previous input data.

Automated Assessment of Participation and Collaboration in CSCL Conversations

Computer-Supported Collaborative Learning (CSCL) gains a broader usage due to technology adoption, while dialogism represents the most adequate framework for representing CSCL conversations [23, 24]. Concurrently, the need for automated conversation analysis tools to support tutors in the cumbersome process of analyzing students' interactions and activity has increased. Collaboration, which can be viewed as the inter-animation of ideas or opinions pertaining to different

participants, represents a central element of dialogue [8]. Several analyses performed based on our CSCL collaboration evaluation models [9] are available on the website. These include participant interaction scores with an interaction graph built on top of Cohesion Network Analysis and visually displayed using the D3.js library. Specific indices are being computed for each participant, such as: number of contributions, cumulated contribution scores, degree of inter-animation, cumulative social knowledge building scores, in- and out- degree, closeness, betweenness, and eccentricity centrality measures from the interaction graph, relevance for top 10 conversation topics [9].

Each participation and collaboration index is used for obtaining an in-depth perspective of each member's involvement, followed by specific visual graphs. The first graph from Figure 5 depicts each participant's evolution as cumulative contribution scores across the timeframe of the conversation. The following two graphs depict the collaboration between participants in terms of the social knowledge building and the voice inter-animation model. Spikes with these 2 graphs denote intense collaborations spanning throughout the conversation.

In terms of underlying computational processes, the importance of each contribution is first computed by relying on the relevance of the covered topics from the entire conversation and present within the utterance. Second, collaboration was computed as the impact on other members' contributions in terms of cohesion (a longitudinal analysis of the conversation) and dialogism (a transversal analysis based on co-occurrence voice patterns). Therefore, within these models, collaboration was assessed using a bottom-up approach which emphasize that cohesion is a signature of collaboration.

Sentiment Analysis

Sentiment analysis and opinion mining are often referred in linguistic and psychological research in recent years. The sentiments extracted from author's text (for example, participants' contributions in a conversation or the absence of their interaction) provide information regarding author's feelings. Interaction established between members in a conversation influences further contributions and interactions.

The analysis of the participants' sentiments can take into consideration specific optimizations, such as ignoring contributions that do not cover specific topics or excluding contributions with no further references or irrelevant regarding main topics.

Specific goals can be defined given a text in terms of sentiment analysis. For example, specific sentiments from an input text can be extracted and split into the 6 major categories expressed by Picard [25]: excited, sad, scared, angry, tender and happy. A demo showing this approach is available on the *ReaderBench* website. In the backend, the framework computes these major sentiments combining scores for valences gathered from specific lists. English, French and Dutch languages are currently supported.


Figure 5. CSCL graphs generated for a sample conversation file.

Figure 6 shows an example of sentiment analysis results produced by the framework for the previous sample input. Negative results express absolute values for negative emotions, therefore emphasizing the positive nature of the entire text.

SENT	IMENT AN	IALYSIS
DOCU	MENT	
EXCIT	ED 0.05	3 SAD -0.316 SCARED -0.084 ANGRY 0.016 TENDER 0.221 HAPPY 0.479
	PARAGRAPH	•
	PARAGRAPH	1
	EXCITED	0.355 SAD -0.709 SCARED -0.091 ANGRY -0.009 TENDER 0.309
	НАРРУ	0.936
		SENTENCE 0
		EXCITED 0.15 SAD -0.713 SCARED -0.062 ANGRY -0.037 TENDER 0.375
		HAPPY 0.981
		SENTENCE 1
		SENTENCE 2
	PARAGRAPH	2
	EXCITED	0.053 SAD -0.316 SCARED -0.084 ANGRY 0.016 TENDER 0.221
	HAPPY	0.479

Figure 6. Sentiment analysis results computed for the sample input data.

The common resource for all considered languages represents the Linguistic Inquiry and Word Count (LIWC) dictionary [26] which contains words related to psychological phenomena, personal concerns, thoughts, feelings, personality, and motivations. At present, all dictionaries are used to explore their linguistic coverage and only those that are present in at least 20% of entry samples are considered for follow-up statistical analyses. The following word dictionary lists were integrated for English language in an approach similar to the one proposed by Crossley et al. [27]:

- Affective Norms for English Words (ANEW) [28], which provides values on three dimensions (valence, arousal and dominance) for more than 1,000 English verbs, nouns, and adjectives;
- Geneva Affect Label Coder (GALC) [29], which contains affective valences such as admiration, amusement, anger, anxiety and many others;
- EmoLex [30], comprising sentiments like anger, anticipation, disgust, fear and others;
- SenticNet [31], including five affective norms: pleasantness, attention, sensitivity, aptitude and polarity;
- Harvard IV-4 from the *General Inquirer* (GI) [32], which contains valences such as power, weak, active, passive, legal and more others;
- Lasswell dictionary [33], which includes sentiments like power gain, power loss, affective gain, affective loss and some others.

In addition, the Affective Norms for French words (FAN) [34] and the Dutch Affective Word Norms [35], the equivalent French and Dutch versions of ANEW, are also integrated in *ReaderBench*.

Semantic Models and Topic Mining

For this core component, the *ReaderBench* framework uses semantic similarity metrics based on ontologies (e.g., Wu-Palmer distance applied on WordNet), as well as cosine similarity between LSA word vectors and the inverse of the Jensen-Shannon dissimilarly between LDA topic distributions [11].

Cohesion Network Analysis introduced a generalized model based on the cohesion graph to represent discourse structure and underlying cohesive links. Based on CNA, a topic mining module was implemented, which extracts the most relevant concepts from a text. Integrated within the web interface, this module draws a concept map of these keywords: the nodes represent the central topics and the links between them depict the semantic similarity between two concepts; the size of each node is proportional to its relevance. Figure 7 presents the obtained concept map for a given input text, which is used for all subsequent print-screens for English language.





EDUCATIONAL SCENARIOS

Up until recently, the desktop version of our *ReaderBench* framework was hardly usable in hands-on educational contexts due to the requirements of extensive processing power and high amounts of memory usage. Due to these limitations, it was mostly used in follow-up offline analyses. The online version opens up new usages of *ReaderBench* in education, as our framework can now be effectively used in a wide range of educational situations and needs. First, students can be engaged in reading a course material, then eliciting their understanding of it. *ReaderBench* can *identify their reading strategies*, providing an in-depth perspective of comprehension processes used to obtain a coherent mental representation of discourse.

Second, students can write an essay or a summary integrating the content of diverse topics from the course material. The *automated essay grading* component provides them access to more than 200 textual complexity indices integrated within a multi-layered model that covers lexical, syntax, semantics and discourse structure measurements.

Third, students can start discussing the course topics in a CSCL environment (chat, forum or blog). The *Computer Supported Collaborative Learning* (CSCL) component is centered on conversation analysis in terms of automated indices of participation and of collaboration, essential for evaluating each member's active involvement in the discussion. Eventually, the *sentiment analysis* component detects positive and negative emotions expressed in texts that, corroborated with the *semantic models and topic mining* component, enable a clearer perspective in terms of points of view and of underlying interests.

Besides this overall scenario, specific educational experiments were undergone in order to validate our models. Some of them are available online on our ReaderBench website, while others were built only for specific analyses and were not published online as web services. For example, of particular interest, is a serious game, currently under development, that enables users to enter textual competitions (e.g., creativity mini-games to identify inferred concepts, essay writing contests, selfexplanations covering specific reading strategies) with other learners and to win based on higher predicted comprehension scores. Advanced techniques may be used to group students into clusters and the teaching material could be differentiated for each group. Another particular example of an extension currently under development is a tool focused on a contextual CV analysis. Given a PDF file representing a personal CV, the tool extracts specific indices and applies specific statistic model in order to predict whether the CV is adequate or not.

As future functionality enhancements, besides the Principal Component Analysis used to identify representative dimensions for each corpus in terms of sentiment analysis, specific improvements are also considered: integration of rules for valence shifting and the consideration of only positive and negative reviews, disregarding neutral or irrelevant content.

As a concluding remark, we must emphasize the extensibility of our *ReaderBench* framework and its broad potential usage in terms of integration within education scenarios performed in various languages. This paper is specifically meant to provide a global overview of the developed web interface, whereas specific details and validations are presented in detail in referred papers.

ACKNOWLEDGMENTS

This work was funded by the 644187 EC H2020 RAGE (Realising and Applied Gaming Eco-System) http://www.rageproject.eu/project.

REFERENCES

- Manning, C.D. and Schütze, H. Foundations of statistical Natural Language Processing. MIT Press, Cambridge, MA, 1999.
- Dascalu, M., Stavarache, L.L., Dessus, P., Trausan-Matu, S., McNamara, D.S., and Bianco, M. ReaderBench: The Learning Companion. In 17th Int. Conf. on Artificial Intelligence in Education (AIED 2015) Springer, Madrid, Spain, (2015), 915–916.
- Dascalu, M., Stavarache, L.L., Dessus, P., Trausan-Matu, S., McNamara, D.S., and Bianco, M. ReaderBench: An Integrated Cohesion-Centered Framework. In *10th European Conf. on Technology Enhanced Learning*, G. Conole, T. Klobucar, C. Rensing, J. Konert and É. Lavoué Eds. Springer, Toledo, Spain, (2015), 505–508.
- Dascalu, M., Stavarache, L.L., Trausan-Matu, S., Dessus, P., Bianco, M., and McNamara, D.S. ReaderBench: An Integrated Tool Supporting both Individual and Collaborative Learning. In 5th Int. Learning Analytics & Knowledge Conf. (LAK'15) ACM, Poughkeepsie, NY, (2015), 436-437.
- Dascalu, M., Dessus, P., Bianco, M., Trausan-Matu, S., and Nardy, A. Mining texts, learner productions and strategies with ReaderBench. In *Educational Data Mining: Applications and Trends*, A. Peña-Ayala Ed. Springer, Cham, Switzerland, (2014), 335–377.
- Dascalu, M., Trausan-Matu, S., Dessus, P., and McNamara, D.S. Discourse cohesion: A signature of collaboration. In 5th Int. Learning Analytics & Knowledge Conf. (LAK'15) ACM, Poughkeepsie, NY, (2015), 350–354.
- Dascalu, M., Trausan-Matu, S., Dessus, P., and McNamara, D.S. Dialogism: A Framework for CSCL and a Signature of Collaboration. In 11th Int. Conf. on Computer-Supported Collaborative Learning (CSCL 2015), O. Lindwall, P. Häkkinen, T. Koschmann, P. Tchounikine and S. Ludvigsen Eds. ISLS, Gothenburg, Sweden, (2015), 86–93.
- 8. Trausan-Matu, S., Stahl, G., and Sarmiento, J. Supporting polyphonic collaborative learning. *Indiana University Press, E-service Journal 6*, (1), (2007), 58–74.
- Dascalu, M., Trausan-Matu, S., McNamara, D.S., and Dessus, P. ReaderBench – Automated Evaluation of Collaboration based on Cohesion and Dialogism. *International Journal of Computer-Supported Collaborative Learning 10*, (4), (2015), 395–423.
- 10. Manning, C.D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S.J., and McClosky, D. The Stanford CoreNLP Natural Language Processing Toolkit. In 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations ACL, Baltimore, MA, (2014), 55–60.
- 11. Dascalu, M. Analyzing discourse and text complexity for learning and collaborating, Studies in Computational Intelligence. Springer, Cham, Switzerland, 2014.

- Budanitsky, A. and Hirst, G. Evaluating WordNetbased Measures of Lexical Semantic Relatedness. *Computational Linguistics* 32, (1), (2006), 13–47.
- 13. Landauer, T.K. and Dumais, S.T. A solution to Plato's problem: the Latent Semantic Analysis theory of acquisition, induction and representation of knowledge. *Psychological Review 104*, (2), (1997), 211–240.
- 14. Blei, D.M., Ng, A.Y., and Jordan, M.I. Latent Dirichlet Allocation. *Journal of Machine Learning Research* 3, (4-5), (2003), 993–1022.
- 15. Taibi, D. and Dietze, S. Fostering Analytics on Learning Analytics Research: the LAK Dataset. In CEUR WS Proceedings Vol. 974, Proceedings of the LAK Data Challenge, held at LAK2013 – 3rd International Conference on Learning Analytics and Knowledg, Leuven, BE, 2013.
- 16. Davies, M. The Corpus of Contemporary American English as the First Reliable Monitor Corpus of English. *Literary and Linguistic Computing 25*, (4), (2010), 447-465.
- 17. Denhière, G., Lemaire, B., Bellissens, C., and Jhean-Larose, S. A semantic space for modeling children's semantic memory. In *Handbook of Latent Semantic Analysis*, T.K. Landauer, D.S. McNamara, S. Dennis and W. Kintsch Eds. Erlbaum, Mahwah, (2007), 143-165.
- McNamara, D.S., Levinstein, I., and Boonthum, C. iSTART: Interactive strategy training for active reading and thinking. *Behavior Research Methods*, *Instruments*, & Computers 36, (2), (2004), 222–233.
- 19. Dascalu, M., Dessus, P., Bianco, M., and Trausan-Matu, S. Are Automatically Identified Reading Strategies Reliable Predictors of Comprehension? In *12th Int. Conf. on Intelligent Tutoring Systems (ITS* 2014), S. Trausan-Matu, K.E. Boyer, M. Crosby and K. Panourgia Eds. Springer, Honolulu, USA, (2014), 456-465.
- 20. Dascalu, M., Stavarache, L.L., Dessus, P., Trausan-Matu, S., McNamara, D.S., and Bianco, M. Predicting Comprehension from Students' Summaries. In 17th Int. Conf. on Artificial Intelligence in Education (AIED 2015) Springer, Madrid, Spain, (2015), 95-104.
- 21. Crossley, S.A., Allen, L.K., and McNamara, D.S., in press. The Writing Pal: A writing strategy tutor. In *Handbook on Educational Technologies for Literacy*, S.A. Crossley and D.S. McNamara Eds. Taylor & Francis, Routledge, New York, NY.
- 22. Wu, Z. and Palmer, M. Verb semantics and lexical selection. In 32nd Annual Meeting of the Association for Computational Linguistics, ACL '94 ACL, New Mexico, USA, (1994), 133–138.
- 23. Koschmann, T. Toward a dialogic theory of learning: Bakhtin's contribution to understanding learning in settings of collaboration. In *Int. Conf. on Computer Support for Collaborative Learning (CSCL'99)*, C.M.

Hoadley and J. Roschelle Eds. ISLS, Palo Alto, (1999), 308-313.

- 24. Stahl, G., *Studying Virtual Math Teams*. Springer, New York, NY, 2009.
- 25. Picard, R. *Affective computing*. M.I.T Media Laboratory Perceptual Computing, 1997.
- 26. Pennebaker, J.W., Booth, R.J., and Francis, M.E. Linguistic inquiry and word count: LIWC [Computer software]. In *Austin, TX: liwc. net* University of Texas, Austin, TX, 2007.
- 27. Crossley, S., Kyle, K., and McNamara, D.S., (in press) Sentiment Analysis and Social Cognition Engine (SEANCE): An Automatic Tool for Sentiment, Social Cognition, and Social Order Analysis. *Behavior Research Methods*.
- 28. Bradley, M.M. and Lang, P.J. Affective norms for English words (ANEW): Stimuli, instruction manual and affective ratings. The Center for Research in Psychophysiology, University of Florida, 1999.
- 29. Scherer, K.R. What are emotions? And how can they be measured? *Social science information* 44, (4), (2005), 695-729.

- Mohammad, S.M. and Turney, P.D. Crowdsourcing a word–emotion association lexicon. *Computational Intelligence 29*, (3), (2013), 436–465.
- 31. Cambria, E., Grassi, M., Poria, S., and Hussain, A. Sentic computing for social media analysis, representation, and retrieval. In *Social Media Retrieval*, N. Ramzan, R. Zwol, J.S. Lee, K. Clüver and X.S. Hua Eds. Springer, New York, NY, (2013), 191–215.
- 32. Stone, P., Dunphy, D.C., Smith, M.S., Ogilvie, D.M., and associates, *The General Inquirer: A Computer Approach to Content Analysis*. The MIT Press, Cambridge, MA, 1966.
- 33. Lasswell, H.D. and Namenwirth, J.Z., *The Lasswell Value Dictionary*. Yale University Press, New Haven, 1969.
- 34. Monnier, C. and Syssau, A. Affective norms for french words (FAN). *Behavior Research Methods* 46, (4), (2014), 1128-1137.
- 35. Moors, A., De Houwer, J., Hermans, D., Wanmaker, S., van Schie, K., Van Harmelen, A.-L., De Schryver, M., De Winne, J., and Brysbaert, M. Norms of valence, arousal, dominance, and age of acquisition for 4,300 Dutch words. *Behavior Research Methods* 45, (1), (2013), 169-177.

Event detection in Tweets

Andrei-Bogdan Baran

"Alexandru Ioan Cuza" University, Faculty of Computer Science General Berthelot, No. 16 andrei.baran@info.uaic.ro

ABSTRACT

Twitter is among the fastest-growing online social networking services, with more than 140 million users producing over 400 million tweets per day. It enables users to post status updates (tweets) about a huge variety of topics to a network of followers using various communication services such as cell phones, e-mails, Web interfaces, or other third-party applications. Monitoring and analyzing this rich and continuous usergenerated content can lead to obtaining valuable information about local and global news and events, because virtually, any person witnessing or involved in any event is nowadays able to disseminate realtime information, which can reach the other side of the world as the event unfolds.

Having a rich data set, we are going to show how to process the tweets in order to obtain valuable information in real time, based on user preferences and different search criteria.

Author Keywords

Twitter; Event detection; Hadoop Ecosystem; Map-Reduce.

ACM Classification Keywords

H.5.2. Information interfaces and presentation (e.g., HCI): User Interfaces. H.3.2. Information Storage and Retrieval: Information Storage.

General Terms

Human Factors; Design.

INTRODUCTION

Unlike other media sources, Twitter messages provide timely and fine-grained information about any kind of event, reflecting, for instance, personal perspectives, social information, conversational aspects, emotional reactions, controversial opinions or even natural hazards. Tweets can be seen as a dynamic source of information enabling individuals, corporations and government organizations to stay informed of "what is happening now".

In contrast to conventional data streaming, event detection from Twitter streams brings new challenges, as it contains large amounts of meaningless messages, which makes the recognition of relevant events even more difficult to achieve.

Event detection in Tweets (EDETE) is an application that extracts various tweets, based on different search criteria (hashtags, locations, etc.) in order to get information about a specific topic of interest. This application focuses on collecting and processing messages from Twitter for

Adrian Iftene

"Alexandru Ioan Cuza" University, Faculty of Computer Science General Berthelot, No. 16 adiftene@info.uaic.ro

detecting relevant events based on a user profile. For collecting data, complex queries are run over Twitter stream (and Twitter API). The data is processed using advanced techniques and tools.

STATE OF THE ART

Despite the fact that the problem of event recognition is social media networks has been studied by many teams of researchers, it is still difficult to find an effective way of extracting relevant information from this type of texts. In order to approach this problem, researchers used techniques from various fields, such as machine learning, natural language processing, data mining, information extraction and retrieval or text mining.

Sankaranarayanan et al. [1] proposed a news processing system based on Twitter (called TwitterStand), to capture tweets that correspond to late breaking news. They employ a naive Bayes classifier to separate news from irrelevant information and an online clustering algorithm based on weighted term vector and use hashtags to reduce clustering errors.

Phuvipadawat and Murata [2] presented a method to collect, group, rank, and track breaking news from Twitter. They first sample tweets using predefined search queries and index their content with Apache Lucene. Messages that are similar to each other are then grouped together to form a news story. They use users' reliability and popularity of tweets, represented by a weighted combination of number of followers and the number of retweeted messages. An application based on the proposed method called Hot-streams has been developed.

Petrovic et al. [3] adapted the online NED (Named-Entity Disambiguation) approach proposed for news media, which is based on cosine similarity between documents to detect new events that have never appeared in previous tweets. Results have shown that ranking according to the number of users is better than ranking according to the number of tweets and considering entropy of the message reduces the amount of spam messages in output.

Becker et al. [4] focused on online identification of realworld event content. They use an incremental clustering algorithm, based on a support vector machine (SVM) classifier. In addition to traditional preprocessing steps such as stop-word elimination and stemming, the weight of hashtag terms is doubled because they are considered a strong indication of the message content.

Cordeiro [5] proposed a continuous wavelet transformation based on hashtag occurrences combined with a topic model inference using latent Dirichlet allocation (LDA). Instead of individual words, hashtags are used for building wavelet signals. An abrupt increase in the number of a given hashtag is considered a good indicator of an event that is happening at a given time. Therefore, all hashtags were retrieved from tweets and then grouped in intervals of 5 minutes. Wavelet peak and local maxima detection techniques are used to detect peaks and changes in the hashtag signal.

Popescu and Pennacchiotti [6] focused on identifying controversial events that provoke public discussions with opposing opinions in Twitter, such as controversies involving celebrities. Their detection framework is based on the notion of a Twitter snapshot, a triplet consisting of a target entity, a given period and a set of tweets about the entity from the target period.

Benson et al. [7] present another approach to identify Twitter messages for concert events using a factor graph model, which simultaneously analyzes individual messages, clusters them according to event type, and induces a canonical value for each event property. Clustering is guided by term popularity and the idea is to uncover rare event messages that are dominated by the popular ones.

Lee and Sumiya [8] present a geosocial local event detection system based on modeling and monitoring crowd behaviors via Twitter, to identify local festivals. They rely on geographical regularities deduced from the usual behavior patterns of crowds using geotags.

Sakaki et al. [9] exploited tweets to detect specific types of events such as earthquakes and typhoons. They formulated event detection as a classification problem and trained an SVM on a manually labeled Twitter data set comprising positive events (earthquakes and typhoons) and negative events (other events or nonevents).

PROPOSED SYSTEM

EDETE system is able to notify the user in real time regarding different event categories: natural hazards, social, sports or political events based on his/her preferences.

The input data set will consist in tweets messages from Twitter Stream API (https://dev.twitter.com/streaming/ overview) and the processing step will be performed using Map-Reduce jobs and other tools from the Hadoop Ecosystem (https://hadoopecosystemtable.github.io/).

Architecture

In Figure 1 is described a high level deployment diagram and the main modules of the system.

Data retrieving and storage

The input data will be collected by using the Twitter streaming API. This stream is limited to receive 1% of the total tweets posted on Twitter. However, 1% is between 3 to 5 million tweets in a day, more than enough for performing different statistics and analysis.

By using the filtered stream, the collecting service is able to filter the tweets by track (kind of keywords), locations and users. A track cannot contain more than 60 characters. Also, the track filter type supports logical operators AND & OR as shown below.



Figure 1. Deployment Diagram.

For example, a track that is containing multiple keywords, separated by space is evaluated with the AND operator: e.g. "apple iPhone", means "apple" AND "iPhone". If multiple tracks are added to a stream, then the OR operator is used instead.

Another type of filter is by location. A location filter represents an area (4 coordinates), and the stream will be filtered only the tweets tagged within the specified location. Moreover, the stream can be filtered by specific users, and this feature will allow to follow the activity of a specific group of users.

The preliminary results showed that there are a lot of tweet messages that don't seem to be relevant to our app. The messages contain a lot of grammar mistakes, abbreviations, shortcuts, missing letters or not even a relevant content based on the attached hashtag.

Data processing

Before starting the data processing, there is an intermediary step called data preprocessing. This involves filtering out stop-words (*on, of, are*, etc.) and applying words stemming and tokenization techniques. As shown above, the data preprocessing already started in the data retrieval module, by applying different filters such as keywords using the logical operators, location, language, etc.

After the preprocessing step, the data is ready to be processed using the Hadoop Ecosystem. Having a rich data set, by using map-reduce jobs we can apply a series of transformations and operations for detecting changes in the data stream. The proposed approach monitors the evolution of the hashtags over the time from the Twitter stream by using wavelet analysis [10, 11]. The premise that we are starting from is that if an important event is about to happen, there will be a significant increase of the same hashtags during a specific period that will describe the same topic. This approach is based on peak analysis of individual hashtags and the local maxima detection for detecting when an actual change occurred.

The entire processing workflow for detecting event can be summarized as follow:

- For each hashtag, there will be generated a series of signals during a specific period of time.
- The period of time will be divided in smaller time intervals (e.g. 10 minutes), where the hashtag occurrence will be counted for each time slot.
- After, for each hashtag, there will be a list of a key value pairs that will contain the timestamps and the number of the hashtag occurrence in that specific time.
- Having these hashtag signals, by applying the wavelet analysis, we can detect the peeks and the changes of a possible event that may occur at a given time.

Because the Twitter hashtags can introduce a lot of noise, to overcome this issue, we can apply the Kolmogorov-Zurbenko (KZ) [12] filter to reveal the hidden signals or to reconstruct the signals because of the poor preliminary results.

This method of event detection is focused only on the hashtags, without any semantic processing. For future work, we are going to add some knowledge driven elements that can improve the accuracy and the level of trust of the event, with a more relevant details of the event.

After processing, for the end user, a series of data visualization and features will be available, that are described briefly in the next section.

End user app

The user interface for EDETE consists in a web application that will have different functionalities described below:

- Each user will have a profile and a set of preferences: such as the *topic of the events*, *a location* and *a set of keywords* that will be used to perform queries over the Twitter stream.
- When the system will identify an event based on user preferences, it will notify the user via a push message mechanism or email.
- Different options of data visualization of the events: *graphs, charts* and *maps*.
- Different data analysis views.
- Google maps integration.

CASE STUDY

Bellow we will describe a general case study for event detection of a natural hazard like an earthquake.

A user, Komura, is interested in monitoring *Tokyo*, *Japan*, an area of high seismic risk, for any possible earthquake. His parents are very old and are living in an old house, at the margin of the city. Komura is worried because he cannot be close to his parents, because he is living in USA with his family, and lately a high seismic activity happened close to Tokyo. By using the EDETE, he can be notified by any possible event that might happen before it will appear on the radio or TV, based on the Twitter data stream.

Komura will create a profile with some relevant keywords (hashtags) and a location near his parents leaving area.

The keywords he set are: #earthquake, #quake and #tsunami. In Figure 2 is captured the screen for the user profile where he can set the hashtags and the location of interest.



Figure 2. EDETE - User Preferences.

In the Table 1, there are captured the tweets count using the keywords described above. The data captured was made on 29th May 2016, between the 13:00 and 22:00, EEST time zone. The results were queried using a Hive (https://hive.apache.org/) script and the grouping filter was by user time zone.

User time zone	Tweets count
Pacific Time (US and Canada)	1347
Eastern Time (US and Canada)	359
Tokyo	287
Central Time (US and Canada)	235

Table 1. Results of counting the tweets by user time zone

As you can see, in Tokyo there were a total of 278 tweets that were related with the hashtags mentioned. Moreover, between these periods, there has been an earthquake of 4.5 magnitude that stroked 162 km ESE of Hasaky, Japan.

RoCHI 2016 proceedings

Moreover, by using the wavelet analysis, the EDETE application will be able to detect the peeks where most tweets messages are about earthquake in Tokyo and will send a notification to the user, with a brief description based on the tweets.

CONCLUSION

EDETE is a solution that can have a real impact on our life by helping us to take fast some actions in case any "negative" event happens. Also it can have a huge potential in our social life or other fields by keeping us informed on topics of our interests. It's true that Twitter contains a lot of noise and irrelevant data, but with the help of a tool that can extract relevant info, it can be turned into a powerful analytic system with a lot of potential usages in different domains.

The first version of the EDETE application will consist only in a hashtag analysis, without any text processing. But as future work, we are planning to introduce semantic text analysis in order to extract more relevant information from tweets. The next version of the application will take into consideration the context of the event, the location and other important elements that can improve the user experience. As a result, more details about the processing step, text analysis algorithms and strategies will be provided in the second version of the application.

ACKNOWLEDGMENTS

We thank to our collaborators that help us during the project developing phases.

REFERENCES

- Sankaranarayanan, J., H. Samet, B. E. Teitler, M. D. Lieberman and Sperling, J. TwitterStand: News in tweets. In *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, GIS '09, ACM, New York, NY, (2009), 42–51.
- 2. Phuvipadawat, S. and Murata, T. Breaking news detection and tracking in Twitter. *In IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology* (WI-IAT), 3, Toronto, ON, (2010), 120–123.
- 3. Petrovic, S., Osborne, M. and Lavrenko, V. Streaming first story detection with application to Twitter. *In*

Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, HLT '10, (2010), 181–189.

- Naaman, M., Becker, H. and Gravano, L. Hip and trendy: characterizing emerging trends on Twitter. Journal of the American Society of Information Science and Technology, 62(5), (2011), 902–918.
- Cordeiro, M. Twitter event detection: Combining wavelet analysis and topic inference summarization. *In Doctoral Symposium on Informatics Engineering*, DSIE'2012. 2012.
- 6. Popescu, A. M. and Pennacchiotti, M. Detecting controversial events from Twitter. In Proceedings of the 19th ACM international Conference on Information and Knowledge Management, CIKM '10, ACM, New York, NY, (2010), 1873-1876.
- Benson, E., Haghighi, A. and Barzilay, R. Event discovery in social media feeds. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Volume 1 of HLT '11, Association for Computational Linguistics, Stroudsburg, PA, (2011), 389-398.
- Fujisaka, T., Lee, R. and Sumiya, K. Discovery of user behavior patterns from geo-tagged microblogs. *In Proceedings of the 4th International Conference on Uniquitous Information Management and Communication*, ICUIMC '10, ACM, New York, NY, (2010), 36:1–36:10.
- Sakaki, T., Okazaki, M. and Matsuo, Y. Earthquake shakes Twitter users: Realtime event detection by social sensors. *In Proceedings of the 19th International Conference on World Wide Web*, WWW '10, ACM, New York, NY, (2010), 851–860.
- 10. Lee, D. and Yamamoto, A. *Wavelet analysis: Theory and applications*. Hewlett-Packard Journal, 1994.
- 11. Cordeiro, M. Twitter event detection: combining wavelet analysis and topic inference summarization. *In Doctoral Symposium on Informatics Engineering*, DSIE. 2012.
- 12. Yang, W., Zurbenko, I. kzft: Kolmogorov-Zurbenko *Fourier Transform and Applications*. R-Project. 2007.

Rhythm analysis of texts using Natural Language Processing

Irina-Diana Niculescu

Politehnica University of Bucharest 313 Splaiul Independentei, Bucharest, Romania irinaniculescu93@gmail.com

ABSTRACT

The paper is introducing a research aiming to analyze rhythm in various genres of texts. After a review of the literature on rhythm formalization in texts, a Natural Language Processing application was developed for analyzing the rhythmicity in three cases: poem, prose, and political speech. The application implemented three formal approaches for rhythm analysis, proposed for the Romanian and French languages. Factors such as: rhythmic unit, rhythmic structure, and rhythmic index were considered. Comparisons are made between different genres, approaches, and languages.

Author Keywords

Rhythm; natural language processing; rhythmic unit; rhythmic structure; rhythmic index

ACM Classification Keywords

I.2.7 Natural Language Processing: Text analysis.

INTRODUCTION

The paper presents the first steps of a research aiming to analyze rhythm in all genres of texts. After performing a state of the art, we have developed a tool for analyzing, using Natural Language Processing (NLP), the rhythmicity of texts, belonging to both the lyric (poems) and epic genre (prose, political speeches). This tool was used for analyzing three texts: a poem, a fragment of a short story and a speech. Some first conclusions were driven.

Rhythm is one of the characteristics of the poetical texts, together with measure and rhyme. However, rhythm is present in everyday speech and in epic texts as well, with the role of emphasizing certain words or syllables, with the explicit intention of the speaker or writer.

Rhythm can be defined as an alternation of stressed with unstressed syllables, while maintaining a certain homogeneity and symmetry.

From the point of view of NLP, rhythm analysis is an interesting research problem, as rhythmicity is of utmost importance both in daily conversation, as well as in the scientific or fictional texts. It is desirable to understand the use of rhythm and to find patterns that can be applied

Stefan Trausan-Matu

University Politehnica of Bucharest 313 Splaiul Independentei, Bucharest, Romania

and

Research Institute for Artificial Intelligence and

Academy of Romanian Scientists stefan.trausan@cs.pub.ro

to the written texts in order to identify in an automatic way the rhythmicity of a specific text.

In the natural language used on a daily basis, emphasizing words is related to the speaker's intention to express a certain state (confusion, puzzlement, contradiction, interest, wonder) or to find information (interrogative sentences). In contrast with the everyday speech, the automation of this process is difficult because the computer is not able to assume human emotions. Therefore, in order to analyze a text's rhythmicity, algorithms based on formulas and observations shall be used. Thus, by analyzing texts for each genre, it can be created a model applicable to all the texts from a particular literary genre.

STATE OF THE ART

An important amount of research has been done in connection with words stress and speech intonation. They differ from language to language. For example, for French always the last syllable is stressed.

Rhythm is obviously influenced by the stresses. Solomon Marcus introduced the concepts of rhythmic structure, rhythmic length and rhythmic index [1]. A text span (verse, sentence, paragraph, etc.) is analyzed by the following factors:

- the length of the considered text (number of words from the span);
- the rhythmic structure, i.e. a string of elements that represent the distance between two stressed syllables;
- the rhythmic length (the length of the rhythmic structure);
- the rhythmic index of a text span, i.e. the smallest natural number k which solves the following inequalities, whatever is the span belonging to the considered language [1]:

span_length <= rhythmic_length <= span_length * k
(1)</pre>

The inequalities (1) assume a relationship between the length of a rhythmic unit and the length of the corresponding rhythmic structure [1]. The uniformity of the language is given by the equality between the rhythm indexes for each considered text span. In the framework

of Solomon Marcus, the rhythmic unit used is the phrase or verse, in the case of a poem.

Mihai Dinu analyzed the poetries rhythm, from the point of view of the rhythmic structure. He proposes to "operate only with single stressed units" [2], so every rhythmic unit contains a single stressed word and the unstressed words that follow a stressed word belong to the next rhythmic unit. Therefore, a verse will be formed of several rhythmic units, unlike the approach used by Solomon Marcus, which considers the entire verse as a rhythmic unit.

Starting from this idea, Mihai Dinu develops what he calls a "succession law of the rhythmic units in a verse" [2]: two consecutive rhythmic units comply with the succession law if the distance between the stressed syllables is divided by 2, 3 or 4, depending on the verse's rhythm [2].

IMPLEMENTATION DETAILS

We have developed an application that facilitates an analysis and a comparison of the rhythm between texts belonging to three literary genres: poetry, prose, political speech. A first experiment using the application has been performed with the following texts:

- The poem "A dream within a dream", written by Edgar Allan Poe [3]. The poem is divided into two stanzas, the first one of 11 lines, and the second one of 13 lines.
- 2 paragraphs from the epic text "A Descent into the Maelstrom", published in 1845 and written by the same Edgar Allan Poe [4].
- A political speech of Jesse Jackson from 1988 [5].

The application is developed for the texts written in English, because there are numerous materials, research articles, and software in this language. For example, for words hyphenation in English, an open-source project based on the Liang algorithm was used [7], and for the words stress, the CMU dictionary was very helpful (http://www.speech.cs.cmu.edu/cgi-bin/cmudict).

The application has been written in Java and it has a graphical interface made by using Java applets and Java Swing. In Figure 1 is a snapshot of the application interface.

The processing has the following stages:

- word hyphenation;
- finding the stressed and unstressed syllables;
- applying several approaches of dividing texts into rhythmic units: the Solomon Marcus approach, the Mihai Dinu approach (only for the poetic texts), and the Boychuk et al. approach [6].

A description of each stage will be made in the following paragraphs.

Writing a program of word hyphenation in English can be considered a difficult process, taking into account the numerous exceptions from the English language.

		Poetry Sp	eech Pros	e		
Solomon Marcus Approach	Rhy	tmic Units Appr	oach	Mih	ai Dinu Approach	Reset text
Take' this kiss' up - on the brow'!	(x)= 6	fixlm II. 2. 4	u	l(fi(x))= 3	(x)= 2	
And, in part ' - ing from y - ou now',	I(x)= 6	fi(x)= [3, 5]	l(fi(x))= 2	i(x)= 3	104-1	
Thus much let' me avow' I(x)= 5	fi(x)= [3, 2]	l(fi(x)) = 2	I(x)= 3			
Y - ou are not wrong', who deem'	I(x) = 6	fi(x)= [5, 2]	l(fi(x))=2	i(x)= 3		
That my days' have been a dream';	I(x) = 7	fi(x)= [3, 4]	I(fi(x)) = 2	i(x) = 4		
Yet if hope' has flown' away I(x)= 6	fi(x)= [3, 2,	1]	I(fi(x)) = 3	i(x)= 2		
In a night', or in a day', I(x)= 7	fi(x)= [3, 4]	I(fi(x)) = 2	i(x) = 4			
In a vi ' - sion, or in none, l(x)= 6	fi(x) = [3, 4]	l(fi(x)) = 2	I(x)= 3			
is it there - fore the less gone?	I(x) = 6	fi(x) = [7]	(f (x)) = 1	i(x) = 6		
Al · I that we see' or seem' I(x)= 6	fi(x) = [5, 2]	l(f(x)) = 2	I(x)= 3	1.1		
is but a dream' with - in a dream'.	I(x) = 7	f(x) = [4, 4]	l(fi(x))=2	i(x) = 4		
Is '-tand amid the roar' (x)= 5	fi(x)= 12, 41	l(f(x)) = 2	i(x)= 3			
Of a sur ' - f-tor ' - ment - ed shore'.	1(x)= 5	fi(x)= 13, 2, 3	1	l(f(x)) = 3	(x) = 2	
And I hold' with - in my hand'	I(x) = 6	fi(x)= [3, 4]	(fi(x)) = 2	i(x)= 3		
Grain ' - s of the gold ' - en sand'-	I(x) = 5	fi(x)= [1, 4, 2	1	I(fi(x)) = 3	i(x)= 2	
How' few! yet how' they creep'	I(x) = 6	fi(x)= [1, 3, 2	2	l(fi(x))= 3	i(x) = 2	
Through my fin ' - ger - s to the deep',	I(x)= 6	fi(x)= [3, 5]	(fi(x)) = 2	i(x)= 3		
While I weep'-while I weep'! I(x)= 6	fi(x)= [3, 3]	l(fi(x)) = 2	i(x)= 3			
0' God'! can I not grasp' (x)= 6	fi(x)= {1, 1,	4]	I(fi(x)) = 3	i(x)= 2		
Them with a tighter' clasp? (x)= 5	fi(x)= [4, 1]	I(fi(x)) = 2	i(x)= 3			
0' God'l can I not save' (x)= 6	fi(x)= [1, 1,	4]	l(fi(x))= 3	i(x)= 2		
One from the piti ' - less wave'?	I(x)= 5	fi[x]= [4, 2]	l(fi(x)) = 2	i(x)= 3		
is al - I that we see' or seem'	(x)= 7	fi(x)= [6, 2]	l(fi(x)) = 2	100-4		
But a dream' with - in a dream'?	I(x) = 6	fi(x)= [3, 4]	I(fi(x)) = 2	i(x)= 3		

Figure 1. A screenshot of the analysis of Poe's poem using the approach of Solomon Marcus.

Applet started.

For this step, we have used an open-source program, written in Java (https://github.com/joeha480/texhyphj). This program implements the word hyphenation algorithm of Franklin Mark Liang [7].

- 1. As mentioned, rhythm represents a symmetric alternation of stressed syllables with unstressed syllables. Thus, finding the stressed and unstressed syllables from a word appears as a necessity in a research dedicated to the rhythm analysis. For this step, we have used an open-source project developed at the Carnegie Mellon University: the CMU dictionary. The aim of this dictionary is to provide the words pronunciation from the North American English (http://www.speech.cs.cmu.edu/cgi-bin/cmudict). The dictionary itself contains a list of words and their hyphenation in 39 phonemes. In order to mark the primary stress of a word, it has been used the notation 1 for the stressed phoneme, the notation 2 for secondary stress and 0 for unstressed phonemes.
- 2. The problem encountered in the use of this dictionary was the fact that phonemes separation does not correspond always to the syllables separation, because a syllable may contain one or several phonemes. Therefore, in order to solve this issue we used a modified CMU dictionary (https://webdocs.cs.ualberta.ca/~kondrak/cmudict.htm l).
- 3. Unlike the standard one, the modified CMU dictionary contains the syllables separation of the words. In other words, phonemes are delimited, depending on the words hyphenation.

From the point of view of the implementation, every word from the text is searched in the CMU dictionary. If it is found, it will be verified in which syllable the primary stress can be found. After that, the word is separated in syllables according to the first stage and the corresponding syllable is stressed.

One of the problems encountered was that separation in syllables from the first stage does not correspond with the separation in syllables from the second stage, for some words. This is caused by the fact that the open-source program used at the first step does not deliver 100% accurate results. For example, in the English language, the word "panorama" is hyphenated, according to the English dictionary, in: "pan - o- ra - ma". In the modified CMU dictionary, the phonemes separation is: "P AE2 - N ER0 - AE1 - M AH0". It should be noted that the notation "1" appears in the third syllable (corresponding phoneme "AE"), so it results that the third syllable is the stressed one. However, the word hyphenation according to the program used in the first step is: "panora-ma". A contradiction appears at the stressed syllables, and it leads to errors in the outlined approaches from the following steps. In the software implementation, when the position of the stressed syllable (found with the help of the CMU dictionary) is greater than the total number of syllables (computed by open-source program), the last syllable is considered stressed. Other examples are some nouns at plural: "streets" is hyphenated by the program in "streets", which is wrong, because the word is monosyllabic.

Another problem is that certain words are not found in the CMU dictionary (e.g. the word "promontory"), so it cannot be found the stressed syllable in this way. From the point of view of the software implementation, the unfound words are considered unstressed.

As a result of the research we have performed, three approaches (of Solomon Marcus, Mihai Dinu, and Boychuck et al. approach) have been chosen in order to make a first comparison and an analysis of the rhythm in the three texts we considered, written in English, belonging to both the lyrical and epic genre (prose and speech). For the Solomon Marcus approach, the rhythmic unit is considered to be a phrase or a verse (in the case of poetry) [1]. For the Mihai Dinu approach (applied only on poetry), we delimited every verse into several metric units, each of them containing a single stressed word and the unstressed words that precede it [2]. For the Boychuck et al. approach, each phrase/verse is delimited by punctuation marks, coordinating conjunctions and subordinating conjunctions [6]. The three approaches differ by the rhythmic units division of the chosen texts, but the applied formulas are the same and the analysis is done depending on the same factors In other words, for every rhythmic unit of the texts and approaches chosen, the following rhythmic factors are computed: rhythmic structure, rhythmic length, rhythmic index. The algorithm for computing the rhythmic factors is based on the mathematical formalism described by Solomon Marcus [1] and mentioned in the State of the art section of this paper.The problem encountered in the implementation of these approaches is finding the stressed and the unstressed words from the English language. The research made by Solomon Marcus and Mihai Dinu have been done for the Romanian language, where the unstressed words are the prepositions, conjunctions, pronouns (in some cases), etc. A classification of words in stressed and unstressed is difficult to be done in the Romanian language, because there are several exceptions (for example, cases when the stress is on a pronoun), determined by the writer's intention to emphasize certain words [2].

According to Mihai Dinu, there are both unstressed monosyllabic and polysyllabic words, most of them being supporting words [2].

1. In the English language, words are divided into content words and function words. In the first category, there are nouns, verbs, adjectives, adverbs, etc i.e. the parts of speech that are essential for the transmission of important information in the communication. In the second category, there are pronouns, prepositions, conjunctions, articles, supporting words (http://pronuncian.com/content-and-function-words).

Daniel Jurafsky & James H. Martin mentioned that function words are generally unstressed and content words are stressed [8]. However, this is not a general rule.

Initially, the CMU dictionary was used to find out if a word has accent or not. We noticed that in the CMU dictionary, the majority of the words (with small exceptions) – monosyllabic, polysyllabic, function words - are stressed (the notation "1" at the corresponding phoneme). If all the words would have been considered stressed, the rhythmic index obtained for every sentence would have been 1 (number of stressed words = number of words from the rhythmic unit). Another idea was to consider the monosyllabic words as unstressed, but this would be wrong because there are many monosyllabic content words in English (e.g.: the noun "dream").

Elisa Hansen reiterates the fact that the stress of the monosyllabic words in poems depends on the context and the intention of the poet to suggest a certain sense [9]. Finally, a list of function words was considered. In the implementation, for each word from the text it is checked if it is in the list of the function words. If it is found in the list, it is not searched in the CMU dictionary and it is considered unstressed. Otherwise, the word is considered stressed. As mentioned, this approach does not provide the correct results in all cases, because it may happen that a function word is stressed in a specific verse or context.

RESULTS OF A FIRST EXPERIMENT

Solomon Marcus Approach

For the Solomon Marcus approach, a rhythmic unit was considered a sentence in the case of political speech and prose or a verse in the case of the poem.

In the case of the chosen poem [3], the number of the rhythmic units is equal to the number of the verses, so in this case it will be 24. All the verses are made of 5 - 7 words and the rhythmic index for each verse varies between 2, 3 (obtained for half of the rhythmic units) and 4, with the exception of the verse "Is it therefore the less gone", where the rhythmic index has resulted 6, because the only stressed word found was "gone". The maximal index – equal to 4 - has been obtained for the verses with the maximum length 7.

In the case of the prose [4], the number of the rhythmic units is equal to 9 (the number of the phrases from the two paragraphs chosen).

Rhythmic index	%
2	29.16%
3	50%
4	16.66%
6	4.16%

 Solomon Marcus approach.

The length of the phrases varies greatly (minimum length is 10, and the maximum 66), but it has been noticed that the length of the rhythmic structure is approximately half of the length of the sentence. The rhythmic index corresponding to each text span varies between 2 and 3. For each phrase, an unique rhythmic structure and obviously an unique rhythmic length are obtained. The prose chosen is formed by heterogeneous phrases, in contrast with the poem, where the verses keep mostly the same number of words and syllables.

Rhythmic index	%
2	44.44%
3	55.55%

Table 2. Rhythmic indices obtained for prose using Solomon Marcus approach

The political speech [5], unlike the paragraphs from the prose containing descriptive passages, consists mainly of short sentences. The number of rhythmic units (phrases) obtained is 239. In contrast with the poem, where the verses are formed by 5-7 words, and the prose, where the phrases have between 10 and 66 words, in the chosen speech there are sentences formed by a single word ("Dream.", "Leadership.", "Why?"), mostly stressed. The most frequent rhythmic index is 2 (obtained for 68% of the total rhythmic units).When the speaker wishes to underline certain aspects, sentences were repeated. Obviously, the sentences which are repeated have the same rhythmic structure, rhythmic length and rhythmic index (ex: "They work every day" - the length of the sentence is 4, equal to the number of words, the rhythmic structure is the string of length 2 composed of "[2, 3]", corresponding to the words "work" and "day", and the rhythmic index is 2). It has been noticed that for sentences with similar structure (e.g.: "Call it pain", "Call agony it", and "Call it agony"), the same rhythmic structure and rhythmic index is obtained.

Rhythmic index	%
1	11.29%
2	68.20%
3	20.08%
4	0.41%

 Table 3. Rhythmic indices obtained for political speech using Solomon Marcus approach.

Mihai Dinu Approach

In the case of the Mihai Dinu approach, the formulas from the Solomon Marcus approach are applied, but each rhythmic unit contains a single stressed word. This approach has been applied only to the poem [3]. The lyrics are divided into 2, maximum 3 rhythmic units and similar structures such as "In a night", "or in a day", "But a dream", "within a dream" can be observed. In the Mihai Dinu approach 52 rhythmic units were obtained. The rhythmic units without a stressed word were not considered, because Mihai Dinu proposes "operating only with single stressed units" [2]. For example, the verse "In a vision, or in none" [3], the only stressed word found is "vision", so it was divided into two rhythmic units "In a vision," and "or in none". The last one is not a stressed unit, so it was not considered in the computation.

Rhythmic index	%
1	21.15%
2	28.84%
3	30.76%
4	15.38%
5	1.92%
6	1.92%

 Table 4. Rhythmic indices obtained for poetry using Mihai

 Dinu approach

As in the Solomon Marcus approach for poetry, the most frequent rhythmic index obtained is 3. The rhythmic index 6 appears only once for both approaches, in the case of the verse "Is it therefore the less gone". The rhythmic lengths are obviously 1 (in 82% of the cases) or 2. In the Mihai Dinu approach, the last word in the rhythmic unit is the one stressed, so there are only two cases. If the last syllable is the one stressed, the rhythmic structure has a single element, so the rhythmic length is equal to 1. Otherwise, the rhythmic structure has two elements: the first one is equal to the position of the stressed syllable in the rhythmic unit, and the second one is the difference between the total number of syllables and the position of the stressed syllable. In the selected poem [3], the stressed words are monosyllabic (according to the open-source project used for hyphenation) for all the cases in which the rhythmic length is 1.

Boychuck et al. Approach

In the third considered approach [6], the authors perform a delimitation of the rhythmic units depending on the punctuation marks, coordinating conjunctions and subordinating conjunctions. Although the paper refers to the French language, this approach was used for English and it was applied on the three texts. From the point of view of the implementation, a file for the punctuation marks [10] and for coordinating and subordinating conjunctions from the English language was made [11]. Each verse/phrase was delimited in rhythmic units and the same algorithm from the previous approaches was applied. In the case of the poem, a similarity of the rhythmic units in this case with the rhythmic units from the Mihai Dinu approach can be noticed. A few examples of identical rhythmic units that appear in both approaches are: "can I not save", "or seem", "can I not grasp", "In a vision". In the Boychuck et al. approach, the number of the rhythmic units is equal to 39. The number is greater than the number of rhythmic units for Solomon Marcus approach, but smaller than the number of rhythmic units for Mihai Dinu approach. In contrast with the Mihai Dinu approach, in the implementation, the rhythmic units without stressed words were considered for the computation of the rhythmic factors, due to the fact that Boychuck et al. approach does not delimit the rhythmic units by stressed words. The most frequent rhythmic index obtained is 3, as for the Solomon Marcus and Mihai Dinu approach for poetry. The rhythmic index 6 appears only once as well, in the verse "Is it therefore the less gone". The rhythmic lengths vary between 1 (19 cases out of 39 rhythmic units), 2 (17 cases) and 3 (3 cases).

Rhythmic index	%
1	20.51%
2	25.64%
3	38.46%
4	12.82%
6	2.56%

Table 5. Rhythmic indices obtained for poetry using Solomon Boychuck et al. approach

In the case of the prose, obviously the rhythmic units have a smaller length than in the case of Solomon Marcus approach, but the observation that the length of the rhythmic units is approximately half of the length of the corresponding sentence is kept. The number of the rhythmic units is 51, much bigger than the number of the phrases (9), due to the numerous coordinating and subordinating conjunctions and commas existing in the descriptive paragraphs chosen for the experiment. There are also rhythmic units without stressed words, such as: "Although", "or". The rhythmic index varies between 1, 2, 3 and 4, the biggest percentage is obtained the rhythmic index 2.

Rhythmic index	%
1	17.64%
2	58.82%
3	19.60%
4	3.92%

Table 6. Rhythmic indices obtained for prose using Boychuck et al. approach

In the case of the political speech, the number of rhythmic units obtained is equal to 561, much bigger than the number of the rhythmic units obtained in the Solomon Marcus approach (239). There are observed similar syntagms, repetitions or enumerations, which leads to similar rhythmic structures (ex: enumeration of nouns "suicide, cynicism, pessimism" is divided into three rhythmic units, each containing a stressed noun). Repetitions of rhythmic units such as "Common Ground!", "keep hope alive", "They work every day" are noticed. The examples of repetitions appear also in the Solomon Marcus approach. The most frequent rhythmic index is 2, as for the Solomon Marcus approach for political speech.

Rhythmic index	%
1	25.31%
2	59.89%
3	13.19%
4	1.24%
5	0.17%
6	0.17%

 Table 7. Rhythmic indices obtained for political speech using Boychuck et al. approach

CONCLUSIONS

All the three considered approaches offer a view of the rhythm in literary texts. For small, alike rhythmic units, similarities in the rhythmic structure and rhythmic index obtained are noted. In the case of the prose it has not been noticed an uniformity of the rhythmic units with similar structure, but in the case of the poem and the political speech, the repetition of syntagms/sentences leads to identical rhythmic factors.

However, the implementation of the three approaches have errors due to the mistaken word hyphenation provided by the used software. We also generalized that the emphasis does not fall on the supporting words (function words), but this is not a rule neither in the Romanian language, nor in the English language.

The research and implementation emphasized differences between the considered languages: English, French, and Romanian. However, the Solomon Marcus and Mihai Dinu approaches, even proposed for the Romanian language, proved to be usable also for English. Of course that further investigations should be done in the direction of comparing the constant aspects of rhythmicity across different languages and genres.

In conclusion, the natural language processing for the rhythm analysis of texts results to be helpful, as by automating the process for the rhythm determination, comparisons between various texts and genres can be made.

REFERENCES

- 1. Marcus, S. *Poetica matematică*, Editura Academiei Republicii Socialiste Romania, 1970.
- 2. Dinu, M. *Ritm si rimă în poezia românească*, Cartea românească, 1986.
- 3. http://famouspoetsandpoems.com/poets/edgar_allan_p oe/poems/18847, last accessed on 8 July 2016

RoCHI 2016 proceedings

- http://poestories.com/read/descent, last accessed on 8 July 2016
- http://www.famous-speeches-and-speech-topics.info/ famous-speeches/jesse-jackson-speech-commonground -and-common-sense.htm, last accessed on 8 July 2016
- 6. Boychuk, E., Paramonov, I., Kozhemyakin, N., Kasatkina, N. and Demidov, P.G. Automated approach for rhythm analysis of French literary texts. *Proceeding of the 15th conference of FRUCT association.* (Finnish-Russian University Cooperation in Telecommunications.), 15–23.
- 7. Liang, F. M. *Word Hy-phen-a-tion by Com-put-er*, written by, Report No. STAN-CS-83-977, 1983.
- 8. Jurafsky, D. and Martin, J. H. Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition, Second edition, John Wiley, 2009.
- 9. http://classroom.synonym.com/use-monosyllablespoems-3516.html, last accessed on 8 July 2016
- 10. http://grammar.ccc.commnet.edu/grammar/marks/mar ks.htm, last accessed on 8 July 2016
- 11. http://grammar.ccc.commnet.edu/grammar/conjunctio ns.htm, last accessed on 8 July 2016

Terrain Synthesis from Crude Heightmaps

Adrian Sabou

Alexandre Philippe Mangra Technical University of Cluj-

Napoca

Str. G. Barițiu 28, 400027, Cluj-Napoca, România alexandre.mangra@gmail.com Technical University of Cluj-Napoca Str. G. Barițiu 28, 400027, Cluj-Napoca, România adrian.sabou@cs.utcluj.ro Dorian Gorgan Technical University of Cluj-Napoca Str. G. Barițiu 28, 400027,

Str. G. Barițiu 28, 400027, Cluj-Napoca, România dorian.gorgan@cs.utcluj.ro

ABSTRACT

This paper presents an approach to terrain synthesis from minimal-detail user-provided heightmaps. There is no assumption regarding the level of detail provided, in order to allow users without access to powerful heightmap tools and/or resources to generate useable terrain based on a self-provided crude feature plan. We present the issues stemming from a lack of detail in user input, notably sharp altitude increases and oversimplified feature edges, and proceed to elaborate on using the terrain synthesis algorithm to solve the issues and create a level of detail that more closely resembles realistic terrain models. The algorithm pipeline is presented and parametrized to show how the user can influence the resulting model.

Author Keywords

Terrain synthesis; Heightmap; Worley noise; Perlin noise; Filters.

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms

Algorithms; Terrain models.

INTRODUCTION

Over the past decades, computational power has become less expensive and more powerful thanks to technological advances. Alongside it, computer generated imagery (CGI) increased in availability and potential. CGI is one of the mainstays of technology, used in various fields like video games, movies, art, simulation software and anywhere else image generation is beneficial. One of the reasons for its popularity is the artistic freedom it entails, coupled with the potential to mimic something that does not exist in the real world.

When compared to physical props and background, computer generated imagery becomes evidently advantageous. There is a large number of images unable to be accurately reproduced without the use of computers, be it spaceships, hellish creatures, otherworldly plants or simply vast, expanding landscapes. Creating a quality physical replica would incur costs unreasonable for any budget, not to mention unfeasible if we're considering the entire landscape of an alien planet. A virtual reproduction's costs can easily be quantified in the artist's and/or programmer's work and the required hardware. Thus, the industry's needs have fueled the development of an array of algorithms and generation software tailored specifically at creating this kind of models. Among them, terrain generation is one of the most used fields due to it contributing significantly at reducing production costs of backgrounds.

The terrain model can be created procedurally (using a set of rules) or based on a set of given input data. The latter is usually combined further with algorithms to refine the given data and produce something usable. Purely procedural terrain suffers from restricting the user control over the final location of terrain features like mountains, hills, plains, rivers or islands. On the other side of the spectrum, some synthesis algorithms working with input data such as heightmaps, feature graphs or guides require at least part of the data to be highly specific. This can prove inconvenient to the casual user, forcing him to spend increasing amounts of time researching the software used and finding ways of creating the necessary input.

The casual user, then, raises new issues when trying to create terrain with specific features. He will, in most cases, be unable to properly form input data relevant for the application that would lead up to the desired terrain model. It can become tedious and time-consuming to master a new skill or software in order to obtain decent results.

One issue will be the sudden altitude increases caused by the user creating the input heightmap by hand. Painting the heightmap with only a handful of colors or grayscale values leads to the creation of a layered terrain which does not conform to reality nor has any kind of transition between layers, hence the sharp, perfectly vertical, altitude changes.

Another issue is the lack of detail on such models. The layman will have neither the time nor experience to paint "rough" edges, as seen in nature at the delimitation of two differently elevated areas. There is a high probability of encountering very uniform edges, if not downright straight, thus breaking the illusion of natural, chaotic, form.

This paper elaborates on a simple algorithm which tries to solve these issues by detailing very crude input to a point where it becomes usable, either as the final terrain model or as a more precise input heightmap for more complex algorithms.

RELATED WORKS

A variety of techniques for procedural or user-guided generation already exists, due to the high demand of terrain generation in the movies and video games industries. As information becomes increasingly available, more and more people try to expand the horizons by either improving existing methods or finding new ones. Terrain synthesis is one of those expanding domains, with entire companies being built around terrain generation software and a growing number of research papers detailing new algorithms.

One such example is World Machine Software, LLC and their sole product: World Machine 1. An immensely powerful terrain generation software which allows users to create terrain from scratch. This is done by layering algorithms and directing data through a pipeline formed by the user. It also supports user-guided generation, by allowing the input for algorithms to be provided through external files. At first glance, however, it does not offer ways to process low-detail input. Elevation discrepancies remain sorely visible throughout the processing pipeline. A person trying to control the features will be unable to do so unless he or she invests enough time in learning how to use this complex software. If such procedures exist, they are unintuitive at best.

A case should be made for Gaia 2, procedural terrain software created by Procedural Worlds, which, among other capabilities, allows users to define where they want certain features to be placed by inserting specialized markers called "stamps". This greatly alleviates the input issues but restricts the user to the set of available stamps (currently over 150) as an advantageous tradeoff between control and power. The only downside is its reliance to the Unity game engine since it is provided as a Unity "asset", a plug-in of sorts.

Aside from terrain synthesis software, the number of papers detailing new and experimental algorithms for synthesis is on the rise. The focus is on giving as much power as possible to the user, creating new ways of synthesizing terrain from different input data-sets. Somewhat unsurprisingly, the tendency is to reach for improved reproduction quality. To give the end-user the power to remake relief forms based on certain patterns and to do so at the best quality level possible.

For example, one of the more well-known papers on the topic is the work of Zhou et al. 3, describing an algorithm to map a relief style onto a simplistic user-provided sketch. As long as the user finds a heightmap describing the desired relief shape and pattern, he can utilize the algorithm to great results. This only partially solves the issue of low-detail user guidance. Firstly because only one type of pattern can be applied at a time, preventing, for instance, both a mountain range and a river or lake to be mapped in the same map instance. Secondly, because obtaining such patterns may or may not prove difficult, depending on what the user intends to obtain and the available patterns on the internet. These problems arise, of course, because of the high specialization of the algorithm and are perfectly acceptable in the context of the goal set for this procedure.

Another such work is that of Cruz et al. 4, with an objective similar to that of Zhou et al.: user-guided terrain synthesis. This paper focuses on having an input graph besides the simplistic sketch, called "guide" here. They try to create geomorphically correct terrain from a collection of real-world data. Very similar in both scope and surfacing issues to the previously presented work: it requires information the layman may not immediately have available and it becomes hard to model several terrain features at once.

In the quest for improving the obtained terrain, most researchers specialize their work, leaving the inexperienced user dead in the water. Even when a software product implements something with general availability, the learning curve is almost never shallow. Large amounts of time must be invested for the average user to obtain usable results from most of today's software implementations.

There are two main types of generation algorithms: guided and unguided. Unguided algorithms are easy to use and are controlled by input parameters, but it is hard to restrict the output to certain desired features (i.e. placement of mountains/ rivers/ plateaus). Guided algorithms exist and most of them try to use either existing real-world data 3, 4 or place markers on the spots where certain features are to be created 2. All these methods demand auxiliary input from the user, which may be hard to obtain.

USER-GUIDED TERRAIN SYNTHESIS

While procedurally generating terrain has plenty of advantages, such as speed of generation, variety and realistic detailing, the main drawback is the lack of user involvement in the placement of terrain features. This makes it hard to create something specific and which conforms to the user's requirements.

As described in the previous section, this gave birth to a series of algorithms and software which do exactly that: create terrain based on a set of specific user input data. They give more freedom to affect the end product and allow one to model the shape of the terrain based on their own wishes. Artists and designers gain tremendous power by being able to create terrain in drawing that is then converted to a highly-realistic 3D model. Researchers and other technical-oriented people gain an equal amount of power by being able to create incredible virtual replicas.

For the hobbyist, however, or any other inexperienced user the challenge becomes much greater. One has no use for powerful tools if they require large time investments to master. Furthermore, there is a definite possibility that said user is not interested in highly-detailed or geomorphically correct terrain. The main interest point is the creation of a terrain model simulacrum that abides by the user's requirements. Most of the people interested in terrain synthesis will not be artists, capable of creating detailed heightmaps to provide to the software nor experienced enough to find the other resources needed as input, such as real-world data, formatted in a way which the software expects.

Following is the algorithm proposed to solve this problem by interpreting low-detail heightmaps and synthesizing a terrain model that, while not necessarily accurate from a realistic point of view, meets the requirements set by the user through the input heightmap and places the terrain features where they are expected. It is assumed that an input is provided in the form of a crudely-drawn heightmap, lacking detail. To generate realistic terrain, at this stage of development, plus the issue of needing a collection of patches, explained earlier in this paper, may be hard to obtain by the inexperienced user.

Edge smoothing

The first step is to solve two issues in the same pass. The issues being:

Sharp elevation level transitions

The neophyte user will provide a heightmap where one elevation level ends and another beings with a drastic difference in value/height. Best example would be the user wanting a mountain surrounded by sea and drawing with a high value in an area of very low values. This will cause a vertical drop (value change of 100%) between the value level represented by the "mountain" (white) and the one of the "sea" (black), as can be seen in Figure 1. Going straight from perfect gray to white or black is also not a good use-case, since the value switches by 50% of the total. A lesser but still perfectly vertical drop.



Figure 1. Sharp transition example.

Simplistic edge definitions

The layman will not have the art skills or appropriate resources to paint better edges. He or she will resort to basic straight or curved lines as shape delimiters, also exemplified on Figure 1.

Both of these problems can be solved by a single, wellchosen algorithm which creates a transitory area between levels and breaks the edges up in a rougher contour. In this case it is a custom-made algorithm inspired by Worley noise 5.

Solution

The first step consists of scattering seed points randomly but evenly across the surface of the heightmap, taking the equivalent height values from the input. I. e. if point X's



Figure 2. Seed points around a fixed point.

location is above a black pixel, its value will be 0.0. If it's above a white pixel, its value will be 1.0. The number of seed points is proportional to the number of pixels in the heightmap and can be adjusted for different end results.

The second step is parsing the entire mesh and adjusting the heights of points based on the nearest N seed points as a linear interpolation of their assigned height values using the distance between the affected point and the seed as a weight. This differs from classical Worley noise, where only the N-th closest point is considered in the rendering function. Increasing N enlarges the area which affects the mesh point, meaning the transitional area between altitudes becomes wider. Exemplified in Figure 2. Circled are the seed points used in computing the new point's height when N = 6. The target point will be affected by 4 perfectly black seeds and 2 perfectly white ones, meaning it will end up closer to, but not perfectly, black.

The randomness of the seed point location creates the rough, natural edges that users expect to see and permits infinite variations on the exact contour when randomly redistributing the seed points again: at one time, the mesh point is affected by 6 low-height points and 3 great-height points then, during another run, the same point is affected by only 2 low-height points and 7 great-height points because in the new seed distribution, the positions change and, hence, distances are altered.

The issue of sharp elevation transitions is solved by the linear interpolation between neighboring seed points by creating transition areas which are equally random in appearance, even if this detail is less noticeable.

Adding detail

After creating transitory areas at the edges, the model is left with large expanses of flat terrain. This happens because all throughout the same level of value, encompassing seed points will all have the same value, hence linear interpolation produces that value repeatedly.

At this point, another procedural generation algorithm can be used and overlaid on top of the model in order to create rough detail across the flat areas. We have chosen Perlin noise 6 for this purpose, which is a homogenous noise implementation created by Ken Perlin in the early '80s. The fact that this noise is homogenous means that any two neighboring points have close values and create a pleasing, flowing aspect, unlike true random noise.

The noise will be added as a small increase in height across the entire terrain model. This means it will affect both the large areas of flat terrain and the previously created transitory ones. This will improve the aspect of the terrain and make it more palatable for the human eye. As a side-effect, it adds randomness throughout the model, increasing reusability and the diversity of potential outputs. However, since the detailing is small compared to the overall scale of the terrain, this remark is not of such great importance.

We should add that this step may be replaced by another way of imprinting a more realistic texture to the terrain. The caveat is that one should take care not to add complexity to the user interaction, like needing a secondary input, such as a realistic texture for imprinting upon the model or an extended number of added parameters.

Detail smoothing

After applying the Perlin noise as a means for detailing the terrain model, the shape of the model needs to be smoothened to eliminate any kind of sharp peaks that may occur near the edges. This step also helps make the terrain more pleasing to the eye.

Odd peaks and shapes

This phenomenon may happen because as Perlin is applied uniformly across the mesh, it also affects the slopes previously created. The points on these slopes will be displaced and sometimes the displacement goes against the desired shape, i.e. a point will increase in height whilst it would be aesthetically pleasing to remain fixed or decrease in height.

Digital filters - image processing

The chosen solution to the previous problem is to run the whole model through a digital noise reduction filter. This will effectively remove any "noise" which, in this case, is represented by those seemingly random shapes.

A median or mean filter 7, 8 with a 3x3 kernel is perfectly reasonable to solve this issue and any other oddities the terrain model may show. It is applied to the entire model. One run through should suffice, since overapplying a filter will reduce the level of detail, counter to the initial purpose of this algorithm. Likewise, care should be exercised with more powerful filters, some of which will strip too much detail even with a single pass.

After the completion of this step, one should be left with a reasonably detailed terrain model which respects the initial feature placement requirements provided by the user through a crudely-drawn heightmap. Needless to say, this algorithm will work just as well with more complex input, meaning it is suitable for the entire range of possible heightmap detail.

IMPLEMENTATION

The Unity game engine

For implementing and testing, the Unity 9 game engine was chosen because of its existing rendering engine and ease of programming using self-contained scripts. All steps have been converted into C# scripts and linked together.

The algorithm is implemented using operations on a float value matrix representing the terrain model then said matrix is applied onto the heights of a mesh, effectively rendering the result onscreen.

The following testing section has been fully realized using the Unity implementation. Due to mesh restrictions, the size of the samples has been reduced to under or at 128x128 pixels.

TESTING

For the purposes of testing, only the aesthetics of the final terrain model have been taken into consideration. Completely ignoring the performance aspect, since it is reliant on implementation, the testing focuses on confirming that the before-stated issues are solved and that the final model is at least partially resembling a natural form of terrain. The three heightmaps used for testing are: an overly simplistic, a slightly detailed, and a very detailed one. The first two were made by hand, the third one is sampled from the Internet 10; presented in Figure 3.



Figure 3. Left: Simple Heightmap; Right: Detailed Heightmap; Bottom: Complex Heightmap.



Figure 4b. Left: Perlin Noise; Right: Filtering.

Validating result

Initial testing was done to prove the algorithm does indeed end with a detailed model of plausible terrain. It bears mentioning again that the end goal was not realistic terrain. Instead, it was to create a level of detail that more closely resembles realistic terrain models. Any sufficiently detailed model which may pass for terrain is good enough for confirmation. Figures 4a and 4b show how the model advances from its initial state to the final, more detailed output. Figure 4b also showcases visual artefacts (sharp edges, noticeable on the "mountain" edge) remaining from previous steps and how they are eliminated through filtering.

Figures 5 shows the effect on more detailed heightmaps. The result is proof that when confronted with *too much* detail, the algorithm overrides part of it with its own edge smoothing.

Varying parameters for edge smoothing

These tests have been done to empirically find reasonable value ranges for the number of nearest seed points and the total number of seed points by altering one of them and keeping the other constant.

Number of seed points

This number dictates how many seed points in total are scattered throughout the plane. The number is related to the total number of pixels available (height x width). We shall call this number TotalPx. As we grow the number of seeds, the area of influence for each point in the model decreases, as more seeds are found in its direct vicinity.

This preserves more of the initial detail of the image, counter to what edge smoothing is supposed to do. On the other hand, too few seeds mean that the terrain will no longer respect all the details provided. There may be



Figure 5. Original and final model. Top: Detailed and Bottom: Complex Heightmap.

entire areas uncovered by seeds and, thus, initial detail is not preserved enough. Figure 6 presents a succession of models, showcasing this effect.

Number of nearest seed points

Testing the number of nearest N points taken into consideration is also a worthwhile experiment, to showcase how different values affect the outcome. For this test, the number of seed points is TotalPx/25. Results empirically show that as the number of neighbors decreases, detail fidelity increases, up to a point where the desired smoothing effect is cancelled. When there are too many neighbors taken into consideration, the smoothing is too strong and the entire map becomes flattened. Figure 7 shows this effect.

Conclusion



Figure 6. Seed Point Number. Top-left: TotalPx; Topright: TotalPx/20; Bottom-left: TotalPx/50; Bottomright: TotalPx/250.

While the number of seed points and neighboring seeds affect each other too, empirical results point to the area centered in TotalPx/20 -> TotalPx/25 seed points total and considering around 10-15 neighbors. This should provide acceptable results for most use-cases. Of course, this does not prevent one to experiment and find proper values depending on the given input heightmap.

CONCLUSIONS

In the world of procedural generation, terrain synthesis is one of the most common uses, allowing for inexpensive yet complex backgrounds in movies, video games, simulation software and other possible areas of interest. The rising demand for such algorithms has given birth to a vast array of advances in this field, ranging from pure optimization to hybrid algorithms and brand-new ones designed to bring a wealth of detail into the final model.

While specialized software is constantly trying to simplify the interface and make procedural terrain generation available to the layman, it must always make compromises regarding input detail versus output detail.

Trying to detail incomplete or crude heightmaps is something few people are trying to elaborate on since the focus is on the end product – a realistic terrain model – and all inputs are usually simply mirroring the demands for the algorithm instead of the other way around.

This paper presented a procedural generation algorithm that is supposed to work with minimal input detail. It outputs something aesthetically close to real terrain models, even if it lacks any kind of groundbreaking detailing. A case can be made for using this algorithm as a preliminary for other systems, detailing crude heightmaps to a level acceptable for more advanced synthesis software and / or algorithms.

While not overly complex, this algorithm proved that there is hope for terrain synthesis from input of any detailing level and that one may still discover new techniques for allowing inexperienced users to generate beautiful scenery with minimal effort.

Future work may involve adding more steps towards creating true geomorphically correct terrain. Possibly

erosion algorithms or storing a knowledge base of realworld heightmap information. Testing other initial steps, replacing the randomly seeded linear interpolation and experimenting with the results is another direction of future study worthy for consideration.

REFERENCES

- Discover, World Machine Software, LLC, [Online]. Available: www.worldmachine.com/about.php?page= features.
- 2. Introducing Gaia, Procedural Worlds, 2015. [Online]. Available: www.procedural-worlds.com/gaia/.
- 3. Zhou, A. H., Sun, J., Turk, G. and Rehg, J. M. Terrain synthesis from digital elevation models, IEEE Transactions on Visualization and Computer Graphics, 2007.
- 4. Cruz, L., Ganacim, F., Lucio, D., Velho, L. and de Figueiredo, L. H. *Exemplar-based Terrain Synthesis*, 2013.
- 5. Worley, S. A Cellular Texture Basis Function. In SIGGRAPH '96 Proceedings of the 23rd annual conference on Computer graphics and interactive techniques, 1996.
- 6. Perlin, K. Making Noise, 2002. [Online]. Available: www.noisemachine.com/ talk1/.
- Fisher, R., Perkins, S., Walker, A. and Wolfart, E. Spatial filters - Median filter, 2003. [Online]. Available: homepages.inf.ed.ac.uk/rbf/HIPR2/median.htm.
- 8. Fisher, R., Perkins, S., Walker, A. and Wolfart, E. Spatial filters Mean filter, 2003. [Online]. Available: homepages.inf.ed.ac.uk/rbf/HIPR2/mean.htm.
- 9. Unity Game Engine, Unity Technologies, [Online]. Available: Unity Technologies.
- 10. Bisen, From Heightmap to Worldspace in Skyrim, Hoddminir, [Online]. Available: hoddminir.blogspot.ro/ 2012/02/from-heightmap-toworldspace-in-skyrim.html.

AUTHOR INDEX

Ahmed, Sobhi	50	Iftene, Adrian	42, 46, 103
Alboaie, Lenuța	42	Ionascu, Costel	91
Balog, Alexandru	76	Iordache, Dragos Daniel	2, 81
Barakat, Hanaa	26	Khaddam, Iyad	26
Baran, Andrei	103	Leroy, Laure	50
Bouaniche, Ari	50	Manea, Valentina Iuliana	81
Bouzit, Sara	11	Mangra, Alexandre Philippe	113
Buraga, Sabin	58	Mihaescu, Cristian	91
Calvary, Gaëlle	11	Mocanu, Mihai	91
Chêne, Denis	11	Niculescu, Irina-Diana	107
Ciolan, Ioana-Monica	58	Orban, Szabolcs	36
Ciugudean, Mihaela	69	Patru, Aurel	2
Dafinoiu, Ion	58	Pinzariu, Marian-Nicolae	46
Damian, Catalin	42	Popescu, Paul Stefan	91
Dascalu, Mihai	95	Pribeanu, Costin	2, 6, 76, 81
Dathan, Brahma	22	Sabou, Adrian	113
Dessus, Philippe	95	Seffah, Ahmed	1
Fogarassy-Neszly, Paul	2	Soare, Cristian-Valeriu	65
Gorgan, Dorian	69, 113	Stefanut, Teodor	36
Gorghiu, Gabriel	81	Trausan-Matu, Stefan	22, 95, 107
Gutu, Marius-Gabriel	95	Vanderdonckt, Jean	11, 26
Holotescu, Carmen	91		