

Antonio Camurri
Cristina Costa (Eds.)



78

Intelligent Technologies for Interactive Entertainment

4th International ICST Conference, INTETAIN 2011
Genova, Italy, May 2011
Revised Selected Papers



 Springer

Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering

78

Editorial Board

Ozgur Akan

Middle East Technical University, Ankara, Turkey

Paolo Bellavista

University of Bologna, Italy

Jiannong Cao

Hong Kong Polytechnic University, Hong Kong

Falko Dressler

University of Erlangen, Germany

Domenico Ferrari

Università Cattolica Piacenza, Italy

Mario Gerla

UCLA, USA

Hisashi Kobayashi

Princeton University, USA

Sergio Palazzo

University of Catania, Italy

Sartaj Sahni

University of Florida, USA

Xuemin (Sherman) Shen

University of Waterloo, Canada

Mircea Stan

University of Virginia, USA

Jia Xiaohua

City University of Hong Kong, Hong Kong

Albert Zomaya

University of Sydney, Australia

Geoffrey Coulson

Lancaster University, UK

Antonio Camurri Cristina Costa (Eds.)

Intelligent Technologies for Interactive Entertainment

4th International ICST Conference,
INTETAIN 2011
Genova, Italy, May 25–27, 2011
Revised Selected Papers

 Springer

Volume Editors

Antonio Camurri
Casa Paganini-InfoMus Research Centre
DIST, University of Genova
Piazza Santa Maria in Passione 34
16145 Genova, Italy
E-mail: antonio.camurri@unige.it

Cristina Costa
UBiNT, CREATE-NET
Via alla Cascata 56/D
Povo, 38123 Trento, Italy
E-mail: cristina.costa@create-net.org

ISSN 1867-8211
ISBN 978-3-642-30213-8
DOI 10.1007/978-3-642-30214-5

e-ISSN 1867-822X
e-ISBN 978-3-642-30214-5

Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2012937074

CR Subject Classification (1998): K.8, I.2.1, H.5, H.1.2, J.5

© ICST Institute for Computer Science, Social Informatics and Telecommunications Engineering 2012

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

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

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

Printed on acid-free paper

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

Preface

These are the proceedings of the 4th International ICST Conference on Intelligent Technologies for Interactive Entertainment. The conference, scheduled every 2 years, was initiated in 2005 in Madonna di Campiglio, Italy, targeting to provide a unique international forum for researchers in the field of interaction technologies, with a special focus on entertainment. For the fourth edition of the conference, Intetain returned to Italy, in Genoa, a city rich with historic heritage, monuments and natural attractions in the Mediterranean Riviera and inland.

The conference aims at enhancing the understanding of recent and anticipated advances in interactive technologies and their applications to entertainment, education, culture, and the arts. Interaction technologies are undergoing significant changes in the last few years, and will influence the way users consume and interact with the media and applications, both locally and over the Internet. The explosion of natural, multimodal, and touch-based interfaces, and their access to the general public, has made new interaction paradigms a reality.

The conference technical and demonstration sessions explored all these topics, bringing together researchers from academia and industry, practitioners, and students interested in future techniques for interaction, with the main aim of being a forum to present and discuss contributions from different domains, related to technology, business, the creative process and user-centered studies.

Technical sessions focused on the different aspects of interaction, and included the presentation of research works on virtual/mixed/augmented reality, hardware technologies for interaction and entertainment, devices, animation and virtual characters, nonverbal full-body interaction, storytelling, affective user interfaces, social interaction and children interaction (with a special “Children’s Corner” demo session).

It was our pleasure to also have two outstanding keynotes as part of the conference program: “Persuasive Systems for Small Groups in a Museum” by Oliviero Stock from FBK, Trento (Italy), and “Bebop Virtuosity” by Francois Pachet from Sony CSL, Paris (France). In addition to the technical program, demonstrations of interactive entertainment technology were included. Intetain participants had the opportunity to attend the half-day tutorial on EyesWeb, a widespread open platform enabling research and development of real-time multimodal systems and applications. Finally, the workshop on Social Behavior in Music (SBM) closed the conference and included the keynote by Alessandro Vinciarelli on “Social Signal Processing: Understanding Nonverbal Communication in Social Interactions.”

Intetain 2011 secured the in-cooperation status from the ACM Special Interest Group on Computer – Human Interaction.

The venue of the fourth edition of Intetain was the magnificent monumental building of Palazzo Ducale, in the heart of the city and the main site for cultural

activities. It is a prestigious site, former house of the “Doge” (king) of Genoa when the city was a republic in the Renaissance. The Casa Paganini-InfoMus research center of the University of Genoa hosted, in its recently restored monumental building of S. Maria delle Grazie, the EyesWeb Tutorial and part of the demonstrations.

Aiming at cross-fertilizing scientific and technological research with humanistic and artistic research and investigating new perspectives in user-centric media and future Internet, Casa Paganini-InfoMus activities support research on new paradigms of interaction and experience with a special focus on computational models and multimodal interfaces addressing nonverbal expressive gesture, emotions, and social signals.

We would like to thank our invited speakers Oliviero Stock, Francois Pachet, and Alessandro Vinciarelli for their outstanding keynotes.

Many thanks go to all the volunteers who shared their talent, dedication, and time for the conference organization and support as well as all our technical and financial sponsors. This conference would not be possible without their support.

We would like to thank our sponsors ICST, Create-Net, Casa Paganini-InfoMus, and Palazzo Ducale for their support. We would like to thank the University of Genoa and all the volunteers in the local organization, whose participation made this conference possible.

Special thanks to our local organizer, Barbara Mazzarino, for her precious work, the Demo Chair Donald Glowinski, the SBM Workshop Chair Giovanna Varni, and the EyesWeb Tutorial organizer Paolo Coletta. Another special thank you is for the Web Chair Michele Marchesoni for having effectively and timely supported the relevant tasks related to Web set-up and information updating.

A very special thanks also goes to the Technical Program Committee members for their support in the review process and program definition.

We also thank the members of the Intetain Steering Board, Imrich Chlamtac and Anton Nijholt, and all the staff members of Casa Paganini-InfoMus, and the conference co-ordinator Aza Swedin.

May 2011

Antonio Camurri
Cristina Costa
Gualtiero Volpe

Organization

The 4th International ICST Conference on Intelligent Technologies for Interactive Entertainment was jointly organized in Genoa, Italy, by CREATE-NET and the Casa Paganini-InfoMus research center of the University of Genoa.

Steering Committee

Imrich Chlamtac	CREATE-NET, Trento, Italy
Anton Nijholt	HMI, University of Twente, The Netherlands

Executive Committee

Conference General Chair

Antonio Camurri	University of Genoa, Italy
-----------------	----------------------------

Technical Program Co-chairs

Cristina Costa	CREATE-NET, Trento, Italy
Gualtiero Volpe	University of Genoa, Italy

Demo Chair

Donald Glowinski	University of Genoa, Italy
------------------	----------------------------

SBM2011 Organization

Giovanna Varni	University of Genoa, Italy
Gualtiero Volpe	University of Genoa, Italy
Antonio Camurri	University of Genoa, Italy

Local Arrangements Chairs

Barbara Mazzarino	University of Genoa, Italy
-------------------	----------------------------

Web Chair

Michele Marchesoni	CREATE-NET, Trento, Italy
--------------------	---------------------------

Technical Program Committee

Anton Nijholt	University of Twente, The Netherlands
Arjan Egges	University of Utrecht, The Netherlands
Athanasios V. Vasilakos	University of Western Macedonia, Greece
Ben Falchuk	Telcordia Technologies, Piscataway, USA
Bill Swartout	University of Southern California, USA
David Tacconi	Futur3, Italy
Dzmitry Tsetserukou	EIRIS, Toyohashi University of Technology, Japan
Catherine Pelachaud	CNRS, Telecom ParisTech, France
Dirk Heylen	University of Twente, The Netherlands
Eelke Folmer	University of Nevada, Reno, USA
Fabrizio Granelli	University of Trento, Italy
Federico Avanzini	University of Padova, Italy
Florian Echtler	Munich University of Applied Sciences, Germany
Florian ‘Floyd’ Mueller	University of Melbourne, Australia
Francesco De Pellegrini	CREATE-NET, Italy
Frank Kresin	Waag Society, Amsterdam, The Netherlands
Frederic Bevilacqua	IRCAM, France
Ginevra Castellano	Queen Mary University of London, UK
Iacopo Carreras	CREATE-NET, Italy
Isaac Rudomin	Monterrey Institute of Technology, Mexico
Ivan Volosyak	University of Bremen, Germany
Jaime del Val Higuera	Reverso, Spain
John-Jules Meijer	University of Utrecht, The Netherlands
Kieth Cheverst	University of Lancaster, UK
Manuela Filippa	University of Paris X Nanterre, France
Marc Cavazza	University of Teesside, UK
Margarita Anastassova	CEA, France
Mark Maybury	Information Technoloy Center (ITC), MITR, USA
Mariet Theune	University of Twente, The Netherlands
Massimo Zancanaro	FBK-IRST, Italy
Mats Küssner	King’s College London, UK
Maurizio Mancini	University of Genoa, Italy
Mel Slater	University College London, UK
Nadia Berthouze	University College London, UK
Nicholas Gillian	Queen’s University Belfast, UK
Oscar Mayora	CREATE-NET, Italy
Oswald Lanz	FBK-IRST, Italy
Paolo Coletta	University of Genoa, Italy
Paolo Petta	Medical University of Vienna, Austria

Petra Sundström	The Swedish Institute of Computer Science (SICS), Sweden
Pieter-Jan Maes	IPEM - Ghent University, Belgium
Radoslaw Niewiadomski	Telecom Paristech, France
Renaud Chabrier	Freelance, France
Silvia Gabrielli	CREATE-NET, Italy
Sebastian Boring	University of Calgary, Canada
Sergio Canazza	University of Padova, Italy
Thierry Dutoit	UMONS, Belgium
Tom Cochrane	University of Geneva, Switzerland
Tommaso Bianco	Ircam, Paris
Tsvi Kuflik	The University of Haifa, Israel
Woontack Woo	Gwangju Institute of Science and Technology (GIST), Korea

SBM2011 Program Committee

Ginevra Castellano	Queen Mary University of London, UK
Beatrice de Gelder	Tilburg University, The Netherlands
Luciano Fadiga	Italian Institute of Technology, Italy
Didier Grandjean	Swiss Center for Affective Sciences, Switzerland
Peter Keller	Max Planck Institute for Human Cognitive and Brain Sciences, Germany
R. Benjamin Knapp	Queen's University Belfast, UK
Esteban Maestre	Pompeu Fabra University, Spain
Barbara Mazzarino	University of Genoa, Italy
Anton Nijholt	University of Twente, The Netherlands
Petri Toiviainen	University of Jyväskylä, Finland
Alessandro Vinciarelli	University of Glasgow, UK

Table of Contents

Virtual/Mixed/Augmented Reality

User Interface for Browsing Geotagged Data – Design and Evaluation	1
<i>Erika Reponen, Jaakko Keränen, and Viljakaisa Aaltonen</i>	
Towards Multimodal, Multi-party, and Social Brain-Computer Interfacing	12
<i>Anton Nijholt</i>	
Brain-Computer Interfaces: Proposal of a Paradigm to Increase Output Commands	18
<i>Ricardo Ron-Angevin, Francisco Velasco-Álvarez, and Salvador Sancha-Ros</i>	

Hardware Technologies for Interaction and Entertainment

Steady State Visual Evoked Potential Based Computer Gaming – The Maze	28
<i>Nikolay Chumerin, Nikolay V. Manyakov, Adrien Combaz, Arne Robben, Marijn van Vliet, and Marc M. Van Hulle</i>	
Single Value Devices	38
<i>Angelika Mader, Edwin Dertien, and Dennis Reidsma</i>	
A Kinect-Based Natural Interface for Quadrotor Control	48
<i>Andrea Sanna, Fabrizio Lamberti, Gianluca Paravati, Eduardo Andres Henao Ramirez, and Federico Manuri</i>	
Smart Material Interfaces: A Vision	57
<i>Andrea Minuto, Dhaval Vyas, Wim Poelman, and Anton Nijholt</i>	
User-Centered Evaluation of the Virtual Binocular Interface	63
<i>Donald Glowinski, Maurizio Mancini, Paolo Coletta, Simone Ghisio, Carlo Chiorri, Antonio Camurri, and Gualtiero Volpe</i>	

Displays and Devices

Does Movement Recognition Precision Affect the Player Experience in Exertion Games?	73
<i>Jasmir Nijhar, Nadia Bianchi-Berthouze, and Gemma Boguslawski</i>	

Animation and Virtual Characters

- Elckerlyc in Practice – On the Integration of a BML Realizer in Real Applications 83
Dennis Reidsma and Herwin van Welbergen

Non Verbal Full Body Interaction

- Evaluation of the Mobile Orchestra Explorer Paradigm 93
Donald Glowinski, Maurizio Mancini, and Alberto Massari
- As Wave Impels a Wave* Active Experience of Cultural Heritage and Artistic Content 103
Francesca Cavallero, Antonio Camurri, Corrado Canepa, Nicola Ferrari, Barbara Mazzarino, and Gualtiero Volpe

Storytelling

- An Intelligent Instructional Tool for Puppeteering in Virtual Shadow Puppet Play 113
Sirot Piman and Abdullah Zawawi Talib
- A Tabletop Board Game Interface for Multi-user Interaction with a Storytelling System 123
Thijs Alofs, Mariët Theune, and Ivo Swartjes

Children Interaction

- Design of an Interactive Playground Based on Traditional Children’s Play 129
Daniel Tetteroo, Dennis Reidsma, Betsy van Dijk, and Anton Nijholt
- Designing a Museum Multi-touch Table for Children 139
Betsy van Dijk, Frans van der Sluis, and Anton Nijholt

Affective User Interfaces

- Automatic Recognition of Affective Body Movement in a Video Game Scenario 149
Nikolaos Savva and Nadia Bianchi-Berthouze
- Towards Mimicry Recognition during Human Interactions: Automatic Feature Selection and Representation 160
Xiaofan Sun, Anton Nijholt, and Maja Pantic

Social Interaction

- A Playable Evolutionary Interface for Performance and Social Engagement 170
Insook Choi and Robin Bargar
- Social Interaction in a Cooperative Brain-Computer Interface Game 183
Michel Obbink, Hayrettin Gürkök, Danny Plass-Oude Bos, Gido Hakvoort, Mannes Poel, and Anton Nijholt

Posters

- LUCIA: An Open Source 3D Expressive Avatar for Multimodal h.m.i. 193
Giuseppe Riccardo Leone, Giulio Paci, and Piero Cosi
- The AnimaTricks System: Animating Intelligent Agents from High-Level Goal Declarations 203
Vincenzo Lombardo, Fabrizio Nunnari, and Rossana Damiano
- A Framework for Designing 3D Virtual Environments 209
Salvatore Catanese, Emilio Ferrara, Giacomo Fiumara, and Francesco Pagano

Demos

- The Mobile Orchestra Explorer 219
Donald Glowinski, Maurizio Mancini, and Alberto Massari
- Realtime Expressive Movement Detection Using the EyesWeb XMI Platform 221
Maurizio Mancini, Donald Glowinski, and Alberto Massari
- i-Theatre: Tangible Interactive Storytelling 223
Jesús Muñoz, Michele Marchesoni, and Cristina Costa
- An Invisible Line: Remote Communication Using Expressive Behavior 229
Andrea Cera, Andrew Gerzso, Corrado Canepa, Maurizio Mancini, Donald Glowinski, Simone Ghisio, Paolo Coletta, and Antonio Camurri
- Teaching by Means of a Technologically Augmented Environment: The Stanza Logo-Motoria 231
Serena Zanolla, Antonio Rodà, Filippo Romano, Francesco Scattolin, Gian Luca Foresti, Sergio Canazza, Corrado Canepa, Paolo Coletta, and Gualtiero Volpe

INSIDE: Intuitive Sonic Interaction Design for Education and Entertainment	236
<i>Alain Crevoisier and Cécile Picard-Limpens</i>	
My Presenting Avatar	240
<i>Laurent Ach, Laurent Durieu, Benoit Morel, Karine Chevreau, Hugues de Mazancourt, Bernard Normier, Catherine Pelachaud, and André-Marie Pez</i>	
Interacting with Emotional Virtual Agents	243
<i>Elisabetta Bevacqua, Florian Eyben, Dirk Heylen, Mark ter Maat, Sathish Pammi, Catherine Pelachaud, Marc Schröder, Björn Schuller, Etienne de Sevin, and Martin Wöllmer</i>	
Traditional Shadow Puppet Play – The Virtual Way	246
<i>Abdullah Zawawi Talib, Mohd Azam Osman, Kian Lam Tan, and Sirot Piman</i>	
The Attentive Machine: Be Different!	249
<i>Julien Leroy, Nicolas Riche, François Zajega, Matei Mancas, Joelle Tilmanne, Bernard Gosselin, and Thierry Dutoit</i>	
SBM2011 - Workshop on Social Behavior in Music	
Towards a Dynamic Approach to the Study of Emotions Expressed by Music	252
<i>Kim Torres-Eliard, Carolina Labbé, and Didier Grandjean</i>	
Mutual Engagement in Social Music Making	260
<i>Nick Bryan-Kinns</i>	
Measuring Ensemble Synchrony through Violin Performance Parameters: A Preliminary Progress Report	267
<i>Panagiotis Papiotis, Marco Marchini, Esteban Maestre, and Alfonso Perez</i>	
Communication in Orchestra Playing as Measured with Granger Causality	273
<i>Alessandro D’Ausilio, Leonardo Badino, Yi Li, Sera Tokay, Laila Craighero, Rosario Canto, Yiannis Aloimonos, and Luciano Fadiga</i>	
Author Index	277

User Interface for Browsing Geotagged Data – Design and Evaluation

Erika Reponen, Jaakko Keränen, and Viljakaisa Aaltonen

Nokia Research Center,
Visiokatu 1, 33720 Tampere, Finland

{erika.reponen, jaakko.keranen, viljakaisa.aaltonen}@nokia.com

Abstract. The surface of the Earth is getting covered with geotagged data. We describe a mobile application and UI that combines embodied interaction and a dynamic GUI for browsing geotagged data. We present the design process and analyze results from a user study. UI is based on a dynamic grid visualization that shows geotagged content from the places around the world where the user is pointing. It shows a continuous and interactive flow of items, including real-time content such as live videos. The application is aimed for entertaining and serendipitous use. The study results show that usability and intuitiveness were improved by providing an additional, familiar view and controls; showing transitions between view modes; and enhancing the unfamiliar views. Also, the content grid UI was found to be a good way to browse geotagged data.

Keywords: Geotagged data, Embodied interaction, User interface, Augmented reality.

1 Introduction

It has become common to tag data to locations, either manually or automatically based on metadata. This kind of *geotagged data* combines virtual information with physical locations in the real world. Also, mobile devices often are connected to the Internet, so sharing geotagged data is possible. The surface of the Earth is getting covered with geotagged data. The sensors in mobile devices (such as accelerometer, compass, and GPS) enable using the device for pointing based interaction. The device knows its geographic location and can determine its orientation in relation to the Earth's gravity and magnetic field, and this can be used in an application's user interface (UI) to determine which physical places the user is pointing the device at. Specialized graphics processing hardware of modern mobile phones allow the creation of graphically sophisticated user interfaces where techniques like real-time 3D rendering can be used. One challenge in presenting geotagged data in user interfaces is the huge amount of data available from different sources. In user interfaces that are based on a 2D map (e.g., Google Maps), geotagged data items are overlaid on the map as small badges (icons, thumbnails or text labels) [see also 14], many items shown at the same time. High density of geotagged data (e.g., tourist photos) may cause information overload. The badges also hinder the visibility of the

geographic view itself. Another example is to have a row of thumbnails above or below a map or AR view (e.g. Flickr's World Map view). To bring an alternative UI for browsing geospatial data, we developed a novel user interface for browsing geotagged data using a mobile device and created an application called MAA (Finnish word meaning the Earth) with which you can browse geotagged data all over the world by pointing to different directions around [11, 12]. In this paper we present and evaluate the full-featured user interface of MAA. We comment on what kind of an effect the changes applied to the preliminary version had. We first cover the relevant related work, including a brief description of the concept. Then, we describe our user interface design process and the user interface of MAA, followed by a description of the user study and an analysis of its results. In the end, there is some further discussion and conclusion, and the list of references.

2 Related Work

We consider our work to be part of reality-based interaction (RBI) [5], positioned between augmented reality (AR) and map-based interaction. Our work combines a graphical user interface seen on the device screen with pointing based interaction. There are two kinds of common mobile applications that use pointing based user interfaces: 1) AR applications that augment the camera viewfinder with information about nearby geotagged data, and 2) applications that determine the pointing direction of the device and show a view into a virtual world accordingly. Example of the former is Nokia Point and Find [4] which is a system that enables getting information about pointed objects through an AR view and the latter astronomy application Sky Walk [13]. A lot of research exists on augmented reality. Rekimoto et al. [9] have talked about an augmented interaction style that focuses on human-real world interaction and not just human-computer interaction. Also philosophers like John Baudrillard [2] and media artists such as Myron W. Kryeger [7] have been presenting various ideas related to mixing physical and virtual reality. Jorge et al. have published research about whole-body orientation in virtual reality interaction [6]. Content that is created and/or consumed in real-time is becoming more common with social media applications such as Twitter. Users are presented with a continuously updated flow of information. Real-time video connections between places out of viewing distance has been researched first by Kit & Sherrie who arranged a media art experiment "Hole in Space" [3] which is a live video connection between NY and LA, and found surprised and excited reactions. Newer research on the real-time video communication tells that also mobile phone videos can be used in real-time communication [10].

2.1 MAA Concept

MAA is an application that utilizes an embodied interaction method for browsing any geotagged data in an entertaining and captivating way. It is an AR application in the sense that it presents the user with an augmented, first person view of the surrounding physical world. However, while traditional AR deals with things that are physically

visible to the user, MAA is about seeing geotagged data that is hundreds or thousands of kilometers away. Users own location in the application is same as his/hers current location in physical world, and right on the surface of the Earth, not below or above it. By pointing downwards, one can view locations on the other side of the planet, as seen through the Earth (Figure 1). The MAA user interface is based on a 3D model of the planet drawn on the display of a mobile device so that it matches with the physical reality around the user. By pointing with the device to any direction around him or her, the user is able to see geotagged data and places in the pointed direction. Differently to traditional 2D map: the user sees the surface from underground, through the Earth, and due to the curvature of the surface areas near the user appear to be “squished” to the horizon.

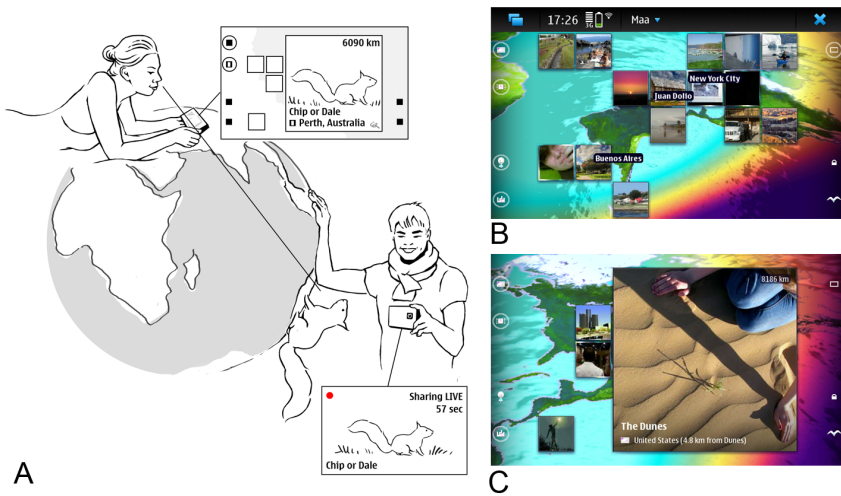


Fig. 1. A: The MAA concept: seeing geotagged data through the Earth, as if the mobile device was a window through the planet, B: MAA user interface, C: Photo preview opened

3 Design Process

Our goal was to make use of an embodied interaction method for browsing geotagged data in a novel and playful way, to allow the user to see the world and geotagged data directly through the Earth, reminding of living on a spherical planet. We were thinking about augmented reality in playful way. We wanted to find a solution that would be able to cope with any amount of content and give easy and direct access to it. The user interface should be simple and elegant, attractive, and easy to use. However, there should always be something new for the user to view. From the beginning, the UI combined embodied interaction with a touch-screen GUI. This combination allowed an interaction style where the user points in the real world and a mobile device acts as a “lens” or “window”, showing a view through the Earth.

The first step was UI sketches on paper, but early prototyping was important because from the beginning we wanted to run the UI on real hardware to get a feel of

the interaction. Development work was carried out on a mobile device with a 3.5", 800 x 480 pixel touch screen. The device had accelerometer and compass sensors. We quickly settled on visualization where geotagged content is presented on a 3D globe. We augment reality in such a way that it would allow the user to see the inner surface of a hollow Earth. To keep the design clean and simple we used as few icons and other visual UI controls as possible. The focus should be on the view of the Earth and the geotagged content itself, not generic icons and badges representing the content. One of the biggest challenges in the visualization was to make the Earth look familiar enough to be recognizable, but still avoid the impression that the user's viewpoint is like in maps somewhere above the surface of the planet. As the UI was intended for mobile touch-screen devices and finger-based use, certain restrictions needed to be applied. Thumbnails, text labels, and other content that the user clicks on were designed to be large enough but user interface not get too cluttered with content. Toolbar icons were designed so that the icons are, although small, so far away from each other that they are easy to hit with a finger.

To validate design ideas we first conducted a study on working prototype of the central aspects of the application with the basic features and the preliminary user interface showing placeholder content: a view of the Earth, some city labels, and a set of dummy photographs tagged to arbitrary locations on the planet. The first evaluation [11] focused on the user experience of seeing through the Earth. Results of that study were used to develop a full-featured version of the user interface. Updated prototype was validated in a user study, reported in this paper.

4 User Interface

MAA UI was designed for an enjoyable user experience, showing content as a dynamic flow, which retains the user's interest while preventing information overflow. The user interface (Figure 1) has three layers: 1) The 3D Earth, 2) Content and 3) UI Controls.

The 3D Earth layer fills the whole screen background and shows a view of the planet. It is presented as seen in first person perspective, through the ground on which the user stands. Oceans and continents as well as polar regions are shown in different colors. Water is considered transparent, so that the sky on the other side of the planet is seen through the oceans and lakes. The day and night sides of the planet are visualized using different color schemes. The approximate direction of sunlight is calculated at the time of day according to the clock of the device. For regions on the day side, the colors are light green and turquoise, and on the night side they are dark violet. Day/night boundary is visualized with a colorful gradient that emulates the colors of a sunrise and sunset in a stylized fashion. When viewing the planet in bird view from above the surface (i.e., from space), the color scheme is gray (Figure 2). The intent is to create a contrast between the "inside" and "outside" views.

The Content layer shows geotagged data from places visible on the screen. Videos and photographs are presented as thumbnails on a grid of 7x5 square cells. There is a subtle effect for separating photos and videos: an animated, glittering frame is drawn

around video thumbnails. The grid does not cover the entire screen to leave room for other UI elements. The size of the grid cells is large enough to allow the user to make sense of the thumbnails. Thumbnails appear and disappear constantly with an animation where the thumbnail flies to/from the screen from/to the tagged locations on the Earth. Consequently, only one thumbnail is visible from the area covered by a grid cell. Zooming in is required if the user wishes to see several items from a small region at the same time. Content is animated but never overlapping, and a maximum of 15 grid cells are filled at once (always leaving 20 cells empty while a preview is not shown), preventing a situation where the screen gets covered in thumbnails and no space is left for seeing the Earth model. This way, large amount of data can be viewed on small screen. When the user clicks on a thumbnail, a preview of it is opened. The clicked cell expands with an animation to cover the area of 5x5 cells. The locations of the world's largest cities (e.g., New York, Sidney) are shown as black text labels, located so that the city is in the center of the label. The number of labels shown simultaneously was restricted to three, to prevent the whole display to be filled up with them. As with thumbnails, each label is only shown for a period of time, after which it is hidden and another label is shown elsewhere on the screen.

On the *UI Controls layer* there is a row of toggle buttons on the left and right sides of the screen. The left side buttons are used for selecting the visible content types: photos, videos, friends, and cities. The right side buttons control the view: a fullscreen toggle, bird view and lock mode. Activated buttons are circled. All the icons are mostly white so that it would be easy to see them above the colorful 3D Earth layer and to make sure they are visually distinct from the items on the content layer.

5 User Study

We arranged a user study to evaluate MAA with the completed user interface and see what kind of impact the visualization and user interface improvements had. The study focused on general usability issues and in how easy the “see through” concept was to understand. For the study, the MAA prototype was configured to show content from a number of existing services on the internet. We used photos from Panoramio, videos from Qik, cities from Geonames, and manually placed the information of five people around the world to act as the user's friends. We recruited 12 participants for the study (6 men, 6 women), between 16 – 55 years, mostly in the age group 31-35. They had a mixed background: various professionals such as user experience designers, managers and engineers working in a large technology company, and one participant was a high school student. A test session lasted 45 minutes. The sessions were conducted inside, in a first floor room with a window. We chose this venue instead of a closed-off usability lab so that the participants could see outside through the window. The sessions were organized as semi-structured interviews while the participants were interacting with the device and performing a set of tasks. During the session the participants tried out the application for 30 minutes. We gave them very little information about the application beforehand to get their initial reactions to the concept. As the session progressed we gave them more information and described the

idea of the application. Questions such as: “describe what kind of thoughts this application raises?”, “what can you do with this application?”, “what kind of content you would like to have in this application?” and “could you show me where London is?” were asked. At the end of the session the participants filled in a questionnaire with free-form answers to 11 questions, including ones such as: “my primary feelings after using the application are ___” and “the application made me think that the Earth is ___”. The users also filled in an AttrakDiff survey [1]. AttrakDiff measures practical quality, hedonic quality identity, attractiveness, and hedonic quality stimulation: it was translated into Finnish for this study (the native language of the participants) but was also available in English. The general structure of the test sessions was similar to the one we had used when evaluating the preliminary version, enabling easy comparison the results. Some additional questions were added about the bird view (seeing from space), viewing near-by content, content grid, and recognizing the type of different content items. Afterwards, all results were combined and analyzed.

6 Study Results and Analysis

The study results and their analysis are divided into two main topics: *Earth navigation* and *the content grid UI*. We summarize the challenges we identified in earlier study [11] and then describe the changes applied to the prototype for current study and analyze their impact. Participants felt that the concept was novel and inspiring. However, they were unfamiliar with the mental model of seeing through the planet. This was not unexpected, as the users had had no prior exposure to the concept. As the AttrakDiff graph (Figure 2) shows, the completed user interface and real world content did not change the overall response and user experience significantly.

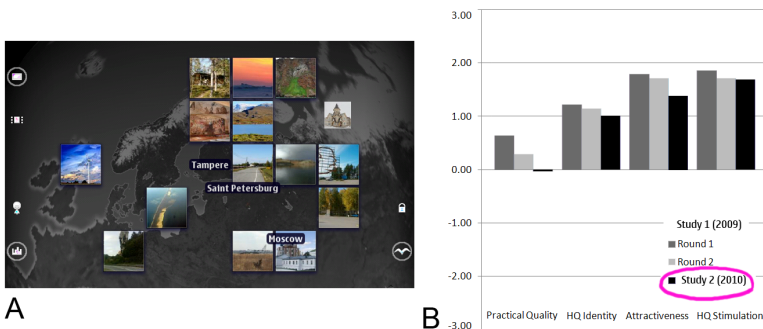


Fig. 2. A: MAA bird view, B: AttrakDiff results: Darkest ones being the ones from the current study, earlier ones from the 2-round study with preliminary UI

6.1 Earth Navigation

Using one’s own body in relation to the device and environment is a good way to browse geospatial content. However, users are expecting to also have more control over the UI in addition to using the body and its position. Embodied interaction style

makes exploring the Earth and content around it more captivating and personal, compared to existing solutions, but the visualization we were using in preliminary version was not intuitive enough. When encountering this kind of a view for the first time, the users were a bit perplexed as to what they were seeing on the screen; the participants had not thought this way about the Earth before. The idea of seeing directly through the Earth was new for them and they found the application novel and unprecedented. Difficulty understanding see through idea derives from the users' pre-existing mental models. To address this we must make it more evident that the view is through the Earth, not from an elevated vantage point or simply a mirrored 2D map. We found three ways to improve the user interface for navigating the Earth and understanding the mental model: A) *providing an additional, familiar view and controls*, B) *transitions between view modes* and C) *enhancing the see-through view*.

A) *Providing an additional, familiar view and controls*. In study with preliminary UI traditional map view and possibility of changing one's own location within the application were wished for. Another feature request was the ability to "zoom all the way to the street level", like in Google Street View. Yet another wish for the UI was the possibility to lock or freeze the view to the pointed direction.

To make MAA better, we then added the bird view mode where the Earth is shown as seen in third person, as if you were looking from space (Fig 2). To make the bird view clearly different than the see-through view, the Earth model was visualized in gray scale in the bird view, and water was made opaque in the bird view. We also enabled locking the view. Locking only affects the viewing direction, so the flow of displayed content is unaffected. We replaced the on-screen zoom buttons with hard keys on the device (volume keys) but we did not increase the maximum zoom level.

The bird view was well understood and got a positive response from users. Users intuitively applied the bird view mode for browsing content from nearby areas — a task that was considered difficult in the default see-through view. The challenge in visualizing areas near the viewpoint in the see-through view is that all regions inside a radius of about 1000 km is shown "flattened" near the horizon. Content grid cells at horizon level also cover a larger geographical area than cells below the horizon. The bird view is similar to traditional 2D/3D maps of the Earth, except in our UI the user's own location is always fixed to the center of the device screen — allowing the pointed directions to continue to match between the physical world and the application's virtual representation of it. The gray scale color scheme in the bird view was not appreciated. While users realized it was beneficial to differentiate it from the see-through view, they found it too dark, gray, and unrealistic. Most of the participants wanted to have both the see-through and bird views, as they suited different situations and needs. The lock mode was welcomed. It was found useful when showing the view to others or for keeping the target fixed when observing one location for a longer period of time. It was also considered to be a possible starting point for manually moving the user's own location, by panning with fingers, to see a little bit behind the Earth in the bird view, or to better see the horizon areas in the see-through view.

B) *Transitions between view modes* were understood better when used manually than when seeing the opening animation. Changing between see-through-mode and

lock-mode manually clearly helped users to understand the difference between the modes and also understand the concept better. Results on the opening animation reveal that users did not realize that they were seeing the inside surface of the Earth after seeing the animation. It was clear that it was showing moving closer to the planet but what happened when the animation ended was misunderstood: “We are moving closer to Finland, surface of the Earth. Ok, and now we are going to China. Wow, really nice.” Some commented that typically opening animations are there just to have something to see while the application is being loaded, and therefore they were not really paying attention to what was happening in the animation. Although the users did not quite understand the animation, they thought it was attractive. When seeing the opening animation for the second time, after having heard that the application allows them to see through the Earth, some said: “Right, ok, that’s what the text says, but I didn’t get it.”

C) Enhancing the see-through view. To make the see-through view more intuitive, we made adjustments to the overall visualization of the world. The most significant changes were a more saturated color scheme and animated, rippling water. These were intended to give the view a more real-time, dynamic appearance. A multi-color gradient was added to visualize the areas where the Sun was currently rising or setting. We were aiming for an increase in the visual impression and feeling of seeing the “imaginary inner surface” of the Earth. As a further link to the physical world, we added the Sun to the sky in the correct direction. The Sun was also visible through the oceans if it happened to reside on the other side of the planet. The day/night boundary was too colorful in some participants’ opinion. Also, its meaning was not clear to all. Seeing the difference between day and night areas was considered important, though, as it indicated which time of day it was in specific locations. On the whole, the color scheme of the see-through view was liked. Many users wanted to check if the Sun was in fact shown in the correct direction within the application and compared it to what they were seeing outside in the window. The animated water went largely uncommented but some misunderstandings were happening: it was either mistaken for real-time animated clouds from satellite pictures, ocean currents, or wind patterns. Nevertheless, we think that these comments hint at animation in the Earth model enhancing the feeling of liveliness.

6.2 Content Grid UI: Browsing Geotagged Data in Real Time

In the earlier study, featuring only placeholder content it was unclear to the users whether they could trust it was coming from the correct places. It was expected that tapping on a content thumbnail would allow them to see the real photograph or details about the location. Users also wished to see, for example, the locations of their friends and real-time videos from around the world.

In the current full featured UI the users were presented with a dynamic flow of real content from around the world. We also added a visualization for showing friend location on the Earth: pulsating towers of light protruding inwards from the Earth’s surface, accompanied by text labels showing the friend’s name. Due to a flaw in the implementation the friends were not visible in all test sessions.

Overall, the finished content grid UI was intuitive to the users. They liked watching the dynamic flow of content on top of the Earth model. They found it inspiring and nice that they serendipitously got interesting content from pointed locations without active searching or selecting. The animated transitions for appearing and disappearing real online content made it clear that the content was coming from actual locations on the Earth. Most of the participants liked being able to turn content types on and off by selecting and unselecting the content type icons and thus filtering the contents to be shown. They also liked that they could control the geographic coverage of the grid by zooming. When the preview was open (Figure 1), the UI also showed the geographic location of the content (e.g., country and the nearest city) and how distant the content was from the user in kilometers. This got a lot of attention from the users as it introduced a link to reality in an otherwise unfamiliar user interface. The users found it fun to see how far the locations are. Although mostly liked, a couple of users commented that the dynamic content grid was too busy. Some users commented that interesting content item disappeared before the user had time to tap on it. Some wanted to have a line or dot indicating where the content came from. While the location of friends was important information for the participants, their visualization was insufficient; they were expecting to see friends as photo thumbnails in the grid. Also, the participants had trouble seeing the difference between video and photo thumbnails. The glittering frame animation in the video thumbnails was not clear enough to separate it from photo thumbnails. Either the animation was not noticed at all (too subtle), or the meaning was unknown. Sometimes it was also unclear which of the thumbnails glittered and which did not, when multiple thumbnails were next to each other. Users were expecting to see playing video, at least in the preview if not already within the thumbnails. The participants suggested adding a familiar ‘play’ triangle symbol on the thumbnails, or using film strip borders. The toolbar icons were considered too small and not easy enough to recognize quickly. However, the users were able to find out what the toolbar icons meant after trying them out. For example, when asked to search for videos, they turned the other content types off and then knew that all they would be seeing was videos.

7 Discussion and Concluding Remarks

Compared to traditional map based user interfaces, we see MAA as a more entertaining and inspiring way to look at the world. It still has a strong basis in the physical reality and makes us pay more attention to the planet below our feet. The kind of user interface we have presented is best suited for exploring content in a serendipitous manner. Only a small number of content items are shown at any given point in time and the user is motivated to keep watching and exploring the content flow. No user actions are required for watching the content flow. MAA gives a user the ability to concentrate on and enjoy the content he or she wants to see, without searching or selecting from the huge amount of available data. However, a search feature would be required for pinpointing the location of specific items of interest. Content is not restricted to the types used in our application, as any kind of geotagged

data could be used with this kind of user interface. Also, the grid UI is suitable for both map based and augmented reality views, as demonstrated in our bird and see-through views. Allowing the user to change his or her location within the application would be interesting because in the see-through view, the Earth looks different depending on where you are watching it. We have presented the design process and analyzed the user study results of the MAA user interface, which is a combination of embodied interaction and a dynamic grid based GUI for browsing geotagged information. User interface shows geotagged data directly through the Earth, in the pointed direction. Additionally, a bird view is provided for focusing on content near the user's current location. Our study results show that providing familiar user interface views in addition to the unfamiliar see-through view, the new application became more usable and understandable. An animated transition between the familiar and unfamiliar views helped the users understand the unfamiliar view. Both see-through and bird views were liked. The see-through view was new, inspiring, and straightforward to use just by pointing, and the bird view was familiar from earlier experience and enabled to see the near-by areas in more detail. Combining old and new was found to be a good solution. UI that combines embodied interaction with a dynamic content grid was considered a good way to browse geotagged data. Together with an Earth model visualization that features day/night colors, the Sun, and animated water, it gives the impression of exploring the world in real-time. Interesting areas for future research include visualization techniques for new kinds of content, techniques for changing the user's location in the see-through view, and even better ways to assist the user understand the mental model of seeing through the Earth.

References

1. AttrakDiff, <http://www.attrakdiff.de/en/Home/>
2. Baudrillard, J.: *Simulacra and Simulation*. University of Michigan Press (1994)
3. Galloway, K., Rabinowitz, S.: *Hole in space* (1980), <http://www.ecafe.com/getty/HIS/index.html>
4. Gao, J., Spasojevic, M., Jacob, M., Setlur, V., Reponen, E., Pulkkinen, M., Schloter, P., Pulli, K.: *Intelligent Visual Matching for Providing Context-Aware Information to Mobile Users*. In: *Supplemental Proc. UbiComp 2007* (2007)
5. Jacob, R.J.K., Girouard, A., Hirshfield, L.M., Horn, M.S., Shaer, O., Solovey, E.T., Zigelbaum, J.: *Reality-Based Interaction: A Framework for Post-WIMP Interfaces*. In: *Proc. CHI 2008*, pp. 201–210. ACM Press (2008)
6. Jorge, V.A.M., Ibiapina, J.M.T., Silva, L.F.M.S., Maciel, A., Nedel, L.P.: *Using Whole-Body Orientation for Virtual Reality Interaction*. In: *Proc. SVR 2009*, pp. 268–272. Brazilian Computer Society (2009)
7. Krueger, M.W.: *Artificial Reality 2*. Addison-Wesley Professional (1991)
8. Rekimoto, J., Nagao, K.: *The World through the Computer: Computer Augmented Interaction with Real World Environments*. In: *Proc. UIST 1995*. ACM Press, USA (1995)
9. Reponen, E.: *Live @ Dublin – Mobile Phone Live Video Group Communication Experiment*. In: Tscheligi, M., Obrist, M., Lugmayr, A. (eds.) *EuroITV 2008*. LNCS, vol. 5066, pp. 133–142. Springer, Heidelberg (2008)

10. Reponen, E., Keränen, J.: Mobile Interaction with Real-Time Geospatial Data by Pointing Through Transparent Earth. In: Proc. NordiCHI 2010. ACM Press (2010)
11. Reponen, E., Keränen, J., Korhonen, H.: World-Wide Access to Geospatial Data by Pointing Through The Earth. In: Extended Abstracts CHI 2010, pp. 3895–3900 (2010)
12. Sky Walk, <http://vitotechnology.com/star-walk.html>
13. Uusitalo, S., Eskolin, P., Belimpasakis, P.: A solution for navigating user-generated content. In: Proc. ISMAR 2009. ACM Press (2009)

Towards Multimodal, Multi-party, and Social Brain-Computer Interfacing

Anton Nijholt

University of Twente, Human Media Interaction
P.O. Box 217, 7500 AE Enschede, The Netherlands
anijholt@cs.utwente.nl

Abstract. In this paper we identify developments that have led to the current interest from computer scientists in Brain-Computer Interfacing (BCI). Non-disabled users have become a target group for BCI applications. Non-disabled users can not be treated as patients. They are free to move and use their hands during the interaction with an application. Therefore BCI should be integrated in a multimodal approach. Games are an important research area since shortcomings of BCI can be translated into challenges in multimodal cooperative, competitive, social and casual games.

Keywords: Brain-Computer Interfacing, Human-Computer Interaction, Multimodal interaction, Games.

1 Introduction

Until recently Brain-Computer Interfacing (BCI) was only considered to be useful for applications that were intended for disabled users. BCI has been used for restoring the communication and mobility of disabled people through applications such as spellers, browsers and wheelchair controls. BCI research aimed almost solely at improving the life of ALS (Amyotrophic Lateral Sclerosis) patients, Parkinson patients, or other kinds of patients in need of interaction with their environment. When there is no alternative, people accept that training is necessary, that they need help to get connected to a computing device, that performance is far from perfect, that they need to concentrate fully on the task at hand, and that any disturbance of this concentration will lead to a breakdown in control. BCI was not part of Computer Science (CS) and Human-Computer Interaction (HCI), let alone a modality that could be used in combination with other modalities to control applications for non-disabled users.

2 Towards BCI for Various Target Populations

In this paper we survey and summarize developments in BCI and HCI that made it possible that BCI research has become an accepted topic in HCI research.

Ambient Intelligence. Ambient Intelligence (AmI), pervasive computing, ubiquitous computing and 'disappearing computers' are names that have been introduced to describe the research domains in which we assume that we live or will live in sensor-equipped environments, that the sensors will be embedded, that they will have local intelligence, and that the information they collect and process can be distributed to other intelligent sensors and computing devices. Obviously, there are already sensor-equipped environments, but, as long as their design is tuned to rather specialized applications, they will certainly not achieve their full potential. In AmI environments, sensors can be used to detect and interpret human behavior and activities, to anticipate certain activities or desires in order to provide real-time support, and to allow explicit control of the environment by its inhabitants by providing feedback and appropriate actions on commands of the inhabitants. These views have led to an increase in attention for sensors in general, including sensors that allow us to issue commands, for example for games and domestic applications, through BCI devices and systems.

HCI and Physiological Measurements. There has always been interest in using (neuro-) physiological measurements to learn about the cognitive load associated with performing certain tasks using a particular interface. Hence, in this case we are not talking about real-time support of the user. We are talking about providing the designer of the interface with information about how users use the interface and about how users experience the interface. Measuring cognitive load is of course the standard example of what interface designers are interested in. This kind of information is meant to re-think, to re-design, and to re-implement the interface in order that it should perform better for a particular user or group of users. However, in recent years many more methods have become available to measure experience. Computer vision, speech analysis, and eye tracking are among them, and this has led to a boost of interest, methods and devices, including BCI devices, that not only measure user experience for redesign and performing tasks more efficiently, but also look at 'tasks' that do not necessarily require efficiency but rather aim at providing positive experiences such as fun, game experience, relaxation, and edutainment. And, moreover, use the information that is sensed in real time to adapt the interface, the task (e.g., the game level) and the interaction modalities to user and context.

BCI Paradigms. The possibility of having sensors (cameras, microphones, accelerators, pressure sensors, proximity sensors, physiological sensors, etc.) that gather knowledge about user characteristics and user activities and behavior is, from an HCI point of view, very useful. It allows us to provide better support to the user of a smart environment. Looking at the possibility of gathering as much information from a user as can be sensed has become a research aim. Brain activity can be sensed. So, why not use it? Knowledge has become available about activity and its appearance in specific regions of the brain. Therefore, various BCI paradigms could be introduced. Activity can be evoked by presenting external stimuli (visual, auditory, tactile), which means that choices can be presented to a user. Brain activity can be measured while a user performs a certain task and the results can be used to adapt task and interface to the user. Certain brain activity is related to a user making errors, a user noticing something irregular, or a

user noticing an event that was anticipated. But there can also be internally evoked activity: elicited by imagining a movement or consciously performing a mental activity. These paradigms allow the mapping of brain activity to implicit or explicit control activity for particular applications.

Towards Unobtrusive Sensing. Measuring brain activity, whether it is caused by internally evoked mental activity or by external activity, requires sensors. For brain research it is no problem to ask a person to perform a certain physical or mental task in a situation that is not necessarily a 'daily life' situation. Rather complicated and expensive devices, e.g. an fMRI scanner, can be used. Fortunately, there are other ways to record brain activity, but unfortunately, they are less precise and only allow a limited number of applications. However, activity related to the BCI paradigms mentioned above can be measured, not yet unobtrusive, but slowly getting there.

The standard way of measuring brain activity in BCI applications is to use a BCI electrodes cap that is on the head of the user and is connected to a computer. The computer can be embedded in an application device, for example a wheelchair, but it can also be a standard PC. This EEG (ElectroEncephaloGraphy) method has up to 256 electrodes integrated in a cap and placed directly on the head. Their positions on the head make it possible to gain information about the electrical activity and, importantly, the function of this activity. But there is not always a need for an EEG cap with lots of electrodes when we look at applications. Apart from being expensive, it has the disadvantage that users are physically connected to the computer, and time is needed to position the electrodes, apply conductive gel and clean-up afterwards.

However, in recent years research groups and companies have developed EEG devices that use so-called dry-cap technology (or 'dry electrodes') and they are exploring the possibility of having a wireless connection between device and computer, so that the user can move more freely. And sometimes, depending on the application, rather than 256 electrodes it is sufficient to have a device with two, eight or sixteen electrodes, allowing companies to develop fancy and portable headsets that can be used for game or domestic applications and in which other sensors, such as accelerators measuring head movements, can be integrated. Companies that are exploring the market for portable headsets for BCI applications can now be identified.

3 BCI, Computer Science, and Human-Computer Interaction

We mentioned the main reasons why BCI research has become visible for the CS and HCI community and why we are now seeing attempts to integrate this research into HCI research in general, research on multimodal interaction, research on using (neuro)physical information aiming at improving interface design, and on real-time adaptation of the interface to the user, and on having BCI as an added modality to control devices and facilitate communication.

When BCI became visible for the CS community one could expect that both start-up companies and R&D departments of large ICT companies would try to exploit and investigate the commercial possibilities of this new technology. That has indeed happened. New companies such as Emotiv and NeuroSky, just to mention the most

influential, and companies such as IBM and Microsoft, have become active in this field. Rather than aiming at medical applications they look at the much bigger market of non-disabled and healthy persons. Consumer products are being offered, but until now these are mainly games, toys, and gadgets. This is not bad, the game market is a billion dollar market and still growing. But clearly, companies also see possibilities to introduce BCI into domestic and professional environments where an added modality to interact with an application will make the interaction more safe or precise.

4 Integrating BCI in Human-Computer Interaction

It is useful to be more explicit about what BCI can offer to the HCI community. Any interface to an application can profit from knowing and learning as much as possible of the persons that are using it. Knowing about activity in particular regions of the human brain provides information that cannot be obtained from other sensors and that can complement that information. The information can be used to inform the interface about the mental and affective state of the user and that knowledge can be used to provide more adequate feedback and also to adapt the interface and the application to a particular user. In recent years this type of BCI has been called passive BCI. It is the system that decides how to use the information. There is no attempt by the user to control the system by consciously 'playing' with this brain activity.

The second reason that BCI is interesting for HCI is similar to the reason that BCI has been exploited for disabled persons: a user can use his or her brain activity as an input modality, maybe in combination with other modalities, to directly control an interface and its application. By performing certain mental tasks the associated brain activity in various regions of the brain can be distinguished and mapped onto commands that control the application. Applications include, among other things, navigating in a virtual world, controlling a robot, or cursor and menu control. Clearly, it is preferable that this mapping is 'natural' or 'intuitive'. Hence, the mental task that has to be performed should be related to the task that has to be performed in the real world or in the graphical user interface. This type of BCI has been called active BCI. A nice example is imagined movement. Brain activity related to imagining a movement (whether it is the tongue, a finger, or a limb) can be distinguished from other activity and can be used to steer a robot around, to control a menu and a cursor, or to have an avatar perform a movement in a virtual world.

Other types of BCI have been introduced or have been included in the definitions of active and passive BCI. For example, there are visual, auditory or tactile evoked potentials. That is, events can be designed in an application to evoke distinguishable brain activity. This allows the user to make clear to the system in which of the available alternatives he or she is interested just by paying attention to it. A possible name for this type of BCI is reactive BCI. Of course, evoked potentials can also happen and be measured when a user is in a situation where he or she has to perform a routine task where once in a while an interesting event happens and this is noticed by the BCI because of a change in brain activity in a particular brain region. Again, an implicit command from the user to the application can be issued if this particular brain activity is detected and the application is ready to accept this command.

There are no clear-cut distinctions between the different types of BCI we have mentioned. For example, we can measure an affective state or rather changes in an affective state in order to adapt the interface to these changes. But it is also possible to design applications where the user is expected to change his or her affective state in order to issue a command. For example, in a game situation we can have a natural change, caused by events happening in the game world, from a user being relaxed to a user being stressed, or the other way around. But issuing commands in such a situation by affective state changes that are consciously aimed at by a gamer can work as well. In the latter situation a gamer has, for example, to decide to become aggressive or to become relaxed knowing what the effect will be in the game world.

We mentioned ways in which brain activity can be used to issue commands or to adapt the interface, the interaction or the particular task a user is expected to perform. A final observation that can be made is that there can be designed or naturally occurring stimuli (physical, visual, auditory, tactile, olfactory) that help the user to perform a mental task or to make a transition from one affective state to another.

5 Multimodal and Hybrid Brain-Computer Interfacing

In HCI and CS research environments physical or mental disabilities of users are hardly taken into consideration. This is quite a contrast with a BCI application that aims at providing an ALS patient with a communication device. These patients are 'locked-in', their brain is functioning, but there is no way that brain activity can control the patient's muscles and movements. It also means that measuring the brain activity related to what a patient wants to communicate will not be interfered with by brain activity evoked by the actual movements of the patient, including involuntary movements such as eye blinks. Clearly, if we want to design BCI applications for non-disabled users we cannot assume that users will not move and therefore we should be able to distinguish brain activity that is meant to control a device from brain activity that has other causes. Similarly, we need to be able to distinguish brain activity that we want to use to adapt the interface to a particular mental state of the user from brain activity that is caused by intended or performed movements. These are quite complicated issues and they lead to quite complicated research issues.

These issues are addressed in hybrid BCI research [4] and in multimodal interaction research [2], where BCI is one of the modalities whose role in the interaction is supported by other modalities or whose role provides support to the other modalities. That is, information obtained from measuring brain activity during interaction can help to reduce ambiguity that is still there after analyzing the role of the more traditional modalities. On the other hand, it can also be the case that brain activity plays the leading role and that the other modalities are there to support and complement the information that can be extracted from the brain activity and that is meant to control devices or activities in a (virtual) environment or that is meant to inform the environment about how to adapt to a particular user, including the user's preferences, cognitive load, and affective mental states. Designing 3D game environments that include BCI as an interaction modality allows us to experiment with multimodal (including BCI) interaction modalities, without necessarily being bothered or limited in creativity by questions about robustness and efficiency [2,5].

6 Multi-party and Social Brain-Computer Interfacing

Nevertheless, robustness is an issue and will probably remain an issue for a long time. Having robust BCI also means that we have the possibility to control and to decrease robustness in order to introduce challenges in game and entertainment situations. But mainly robustness is necessary in noisy environments, environments with lots of distractions, and environments that require the user to spend his or her attention to non-BCI related issues. However, these issues are being addressed in current BCI research. To mention some of the situations that are being addressed: walking in a city environment while using BCI, measuring and distinguishing mental activity while driving a car, playing pinball with motor imagery, playing World of Warcraft with BCI control, or playing a BCI version of Pong with a friend while others are watching and commenting. All this has been done and certainly not in an unsatisfactory way.

Clearly, when we look at nowadays video games, then most interest goes to competitive and cooperative games. There are others involved, sometimes in a mediated manner, sometimes physically present as in LAN parties with lots of verbal and nonverbal interactions or in a situation comparable to a traditional board game. We are investigating the possibilities of such cooperative and competitive multimodal and multi-party social games that use BCI [3]. A next step, in which we look at mediating brain activity from one person to another is much further away, but scientific discussion about it is becoming possible [1]. That will lead us to a social media change from FaceBook to BrainBook.

Acknowledgments. We gratefully acknowledge the support of the BrainGain Smart Mix Programme of the Netherlands Ministry of Economic Affairs and the Ministry of Education, Culture and Science.

References

1. Chorost, M.: *World Wide Mind: The Coming Integration of Humanity, Machines, and the Internet*. Free Press, New York (2011)
2. Gürkök, H., Nijholt, A.: Brain-computer interfaces for multimodal interaction: a survey and principles. *International Journal of Human-Computer Interaction* (2011) (to appear)
3. Obbink, M., Gürkök, H., Plass-Oude Bos, D., Hakvoort, G., Poel, M., Nijholt, A.: Social Interaction in a Cooperative Brain-computer Interface Game. In: Camurri, A., Costa, C., Volpe, G. (eds.) *INTETAIN 2011*. LNICST, vol. 78, pp. 179–188. Springer, Heidelberg (2011)
4. Pfurtscheller, G., Allison, B.Z., Brunner, C., Bauernfeind, G., Solis-Escalante, T., Scherer, R., Zander, T.O., Mueller-Putz, G., Neuper, C., Birbaumer, N.: The hybrid BCI. *Frontiers in Neuroscience* 2, article 3, 1–11 (2010)
5. Tan, D., Nijholt, A. (eds.): *Brain-Computer Interfaces: Applying our Minds to Human-Computer Interaction*. Human-Computer Interaction Series. Springer, London (2010)

Brain-Computer Interfaces: Proposal of a Paradigm to Increase Output Commands

Ricardo Ron-Angevin, Francisco Velasco-Álvarez, and Salvador Sancha-Ros

Dpto. Tecnología Electrónica, ETSI Telecomunicación, Universidad de Málaga,
Campus de Teatinos, 29071, Málaga, Spain
{rra, fvelasco, ssancha}@dte.uma.es

Abstract. A BCI (Brain-Computer Interface) is based on the analysis of the brain activity recorded during certain mental activities, to control an external device. Some of these systems are based on discrimination of different mental tasks, matching the number of mental tasks to the number of control commands and providing the users with one to three commands. The main objective of this paper is to introduce the navigation paradigm proposed by the University of Málaga (UMA-BCI) which, using only two mental states, offers the user several navigation commands to be used to control a virtual wheelchair in a virtual environment (VE). In the same way, this paradigm should be used to provide different control commands to interact with videogames. In order to control the new paradigm, subjects are submitted in a progressive training based in different VEs and games. Encouraging results supported by several experiments show the usability of the paradigm.

Keywords: Brain-Computer Interfaces (BCI), Motor Imagery, Navigation commands, Virtual Environment (VE), Motivation, Games.

1 Introduction

A Brain -Computer Interface (BCI) is a system that enables a communication that is not based on muscular movements but on brain activity. One of its main uses could be in the field of medicine and especially in rehabilitation. It helps to establish a communication and control channel for people with serious motor function problems but without brain function disorder [1].

Most non-invasive BCI systems use the brain activity recorded from electrodes placed on the scalp, i.e., the electroencephalographic signals (EEG). Different features of the EEG signals can be extracted in order to encode the intent of the user. The most common EEG signal features used in current BCI systems include [2] slow cortical potentials [3], P300 potentials [4] or sensorimotor rhythms (SMRs) [5]. SMRs are based on the changes of μ (8-12 Hz) and β (18-26 Hz) rhythm amplitudes, which can be modified by voluntary thoughts through some specific mental tasks, as the motor imagery (MI) [6]. When a person performs a movement, or merely imagines it, it causes an increase or a decrease in μ and β rhythm amplitudes, which are referred to as event-related synchronization (ERS) or event-related desynchronization (ERD) [7].

People can learn to use motor imagery to change SMR amplitudes, and this relevant characteristic is what makes SMR suitable to be used as input for a BCI.

Although for a long time BCI research has been dedicated to the medical domain, in recent years, new BCI applications are focused toward healthy users, for example BCI games [8]. Effectively, BCIs can offer a new means of playing videogames or interacting with virtual environments [9]. However, researchers can use virtual reality (VR) technologies, not only to develop games controlled by brain activity, but also to study and improve brain-computer interaction. The positive impact that the use of VR has in the subjects' performance due to motivation, realism, vivid feedback or ease of use has been reported in several studies [10], [11]. In brain-computer interface research, it is necessary to provide some type of visual feedback allowing subjects to see their progress. VR is a powerful tool with graphical possibilities to improve BCI-feedback presentation and has the capability of creating immersive and motivating environments, which are very important in guaranteeing a successful training [12].

Many BCI applications are focused on the control of a wheelchair; however, before people can use a wheelchair in a real situation, it is necessary to guarantee that they have enough control to avoid dangerous scenarios. VR is a suitable tool to provide subjects with the opportunity to train and test the application. In this way, MI-based BCIs have been used to explore VEs. Some studies that use VR describe a system in which a virtual wheelchair moves in only one direction (forward) [13, 14]. Because of this restricted movement, only one command (and therefore one mental task) is needed. Other systems let the subjects choose among more commands. In [15], a simulated robot performs two actions ('turn left then move forward' or 'turn right then move forward') in response to left or right hand MI. A more versatile application can be found in [16] with three possible commands (turn left, turn right, and move forward) selected with three MI tasks (chosen among left-hand, right-hand, foot, or tongue). These BCIs typically provide the user with one to three commands, each associated with a given task. Having a higher number of commands makes it easier to control the virtual wheelchair, since the subject has more choices to move freely (by means of an information transfer rate increase). Nevertheless, it has been reported in several studies [17, 18] that the best classification accuracy is achieved when only two classes are discriminated. In an application focused on the control of a wheelchair, a classification error (a wrong command) can cause dangerous situations, so it is crucial to guarantee a minimum error rate to keep the users safe. For this purpose, the use of a BCI system based on classification of different mental tasks to provide different commands (associating each command with a mental task) is not the best solution, increasing the probability of misclassification and requiring a very good control.

The main objective of this paper is to introduce the navigation paradigm proposed by the University of Málaga (UMA-BCI) which, using only two mental states, offers the user several navigation commands to be used to control a virtual wheelchair in a VE. In the same way, this paradigm should be used to provide different control commands to interact with videogames. In order to control the new paradigm, subjects are submitted in a progressive training based in different VEs and games.

2 Methods

In this section we will provide an overview on the methods usually used in the UMA-BCI.

2.1 Data Acquisition

The EEG is recorded from two bipolar channels with electrodes placed over the right and left hand sensorimotor area. Active electrodes are placed 2.5cm anterior and posterior to electrode positions C3 and C4 according to the 10/20 international system. The ground electrode is placed at the FPz position. Signals are amplified by a 16 channel biosignal g.BSamp (Guger Technologies) amplifier and then digitized at 128 Hz by a 12-bit resolution data acquisition NI USB-6210 (National Instruments) card. To assure low impedances between the electrodes and the scalp (desired below 5K Ω), electrolyte gel is filled into each electrode before experiments start.

2.2 Training Protocol

Usually, the subjects who participate in the experiments have no previous BCI experience. They all undergo a training protocol for calibration and training purposes.

This training is based on the paradigm proposed by our group (UMA-BCI) in [11], which is based in a videogame. Subjects, immersed in a VE, have to control the displacement of a car to the right or left, according to the mental task carried out, in order to avoid an obstacle. The training protocol generally consists of two sessions, the first without feedback and the second providing continuous feedback. In each session, subjects are instructed to carry out 4 experimental runs, consisting of 40 trials of 8 seconds each. The first session is used to set up classifier parameters (weight vector) for the next feedback session and the future navigation sessions. The training is carried out discriminating between two mental tasks: mental relaxation and imagined right hand movements. The feedback consists in the movement of a car to the right (hand MI) or to the left (relaxation state) depending on the classification result (Figure 1).

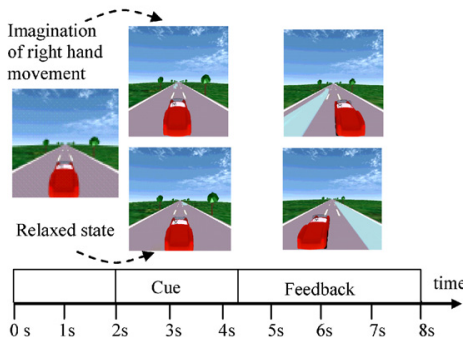


Fig. 1. Timing of one trial of the training with feedback

These two sessions are the same for every participant, and they allow to select those subjects who will continue with the navigation experiments, depending on the obtained results in relation with the classification error.

2.3 Signal Processing

For signal processing, the scheme used is that proposed by Guger et al. [19]. The feature extraction consists of estimating the average band power ($PC3$ and $PC4$) of each EEG channel in predefined, subject specific reactive frequency bands by: (i) digitally band-pass filtering the EEG using a fifth-order Butterworth filter, (ii) squaring each sample, and (iii) averaging over several consecutive past samples. A total of 64 samples are averaged, getting an estimation of the band power for an interval of 500ms. The reactive frequency band is manually selected for each subject, checking the largest difference between the power spectra of two 1s intervals (a full description about how to determine the frequency band can be found in [20]): a reference interval (0.5–1.5s) and an active interval where a mental task takes place (6–7s).

In sessions without feedback, the extracted feature parameters of the classification time points with the lowest classification error are used to set up the classifier parameters for the following session with feedback. The classification is based on the linear discriminant analysis (LDA). In the feedback sessions, the LDA classification result is converted online to the length distance L that the car moves in one or the other direction. The distance L is updated on the screen every four samples, that is, every 31.25 ms, to make feedback continuous to the human eye. The trial paradigm and all the algorithms used in the signal processing are implemented in MATLAB.

3 Navigation Paradigm

The main objective of the BCI research at the University of Málaga is to provide an asynchronous BCI system (UMA-BCI) which, by the discrimination of only two mental tasks, offers the user several output commands. These commands could be used to interact with videogames, as navigation commands to control an external device (robot, wheelchair) or be used in a VE. An asynchronous (or self-paced) system must produce outputs in response to intentional control as well as support periods of no control [21]; those are the so-called intentional control (IC) and non-control (NC) states, respectively. Both states are supported in the paradigm proposed: the system waits in a NC state in which an NC interface is shown (Figure 2a). The NC interface enables subjects to remain in the NC state (not generating any command) until they decide to change to the IC state, where the control is achieved through the IC interface (Figure 2b). The signal processing used to control both interfaces is the same as the one the described in section 2.3.

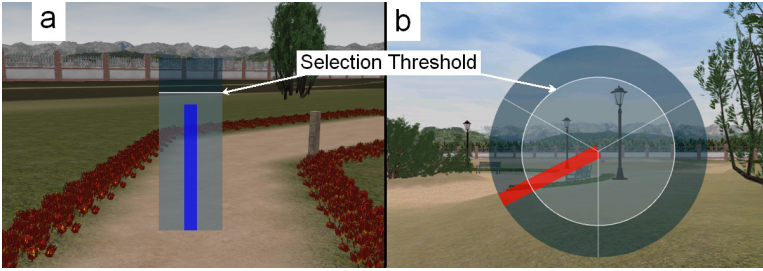


Fig. 2. NC interface (a) and IC interface (b)

The NC interface consists of a semi-transparent vertical blue bar placed in the centre of the screen. The bar length is computed every 62.5 ms (8 samples) as a result of the LDA classification. As preliminary study, the two mental tasks used are the same than the one used during the training phase (training protocol): right-hand MI versus relaxed state: if the classifier determines that the mental task is right-hand MI, the bar extends; otherwise (relaxation state), the bar length remains at its minimum size. In order to change from the NC to the IC state, the subject must extend the bar (carrying out the MI task) over the “selection threshold” and accumulate more than a “selection time” with the bar over this “selection threshold”. If the length is temporarily (less than a “reset time”) lower than the selection threshold, the accumulated selection time is not reset, but otherwise it is set to zero. All these parameters (“selection time”, “selection threshold” and “reset time”) are manually selected for each subject.

The IC interface to select a specific command is based on the methodology used in the design of the typewriter Hex-o-spell developed within the BBCI project [22]. This one consists of a circle divided into several parts, which correspond to the possible navigation commands. The IC interface showed in Figure 2b allows to select 3 commands: move forward, turn right and turn left. A circle divided into four parts allows to select, furthermore, the “move back” command. A bar placed in the centre of the circle is continuously rotating clockwise. The subject can extend the bar carrying out the MI task to select a command when the bar is pointing at it. The way the selection works in this interface is the same as in the NC interface, with the same selection and reset time and the same selection threshold. In the IC interface, another threshold is defined: stop threshold, which is lower than the selection threshold, and not visible to the subject. When it is exceeded, the bar stops its rotation in order to help the subject in the command selection.

Subjects receive audio cues while they interact with the system. When the state changes from IC to NC they hear the Spanish word for ‘wait’; the reverse change is indicated with ‘forward’, since it is the first available command in the IC state. Finally, every time the bar points to a different command, they can hear the correspondent word (‘forward’, ‘right’, ‘back’ or ‘left’).

This navigation paradigm is not to be applied only in VR; it can be used in other scenarios, for example, to control a robot in an experimental situation, or a real wheelchair. In such a scenario, the need for a graphical interface to control the system may not be adequate, as it could limit the subject's field of view, for having to look at a computer screen and, at the same time, distract him from the task of controlling the device (wheelchair, robot...). If a BCI system is to be proposed that allows a subject to control a wheelchair, it should let the user watch the environment at all times.

It is for this reason that the recent work of our group is also focused on an adaptation of this system in which, after training with the graphical interface, subjects could switch gradually to an audio-cued interface. In fact, in the graphical interface proposed, the visual feedback is not necessary, as the only essential information that subjects need to receive is the cue that indicates which command is being pointed by the bar. Subjects hear an audio cue which signals them which navigation command can be selected, so they decide whether to carry out the MI task to select it, or to wait for the next command. Regarding the feedback, the actual movement of the virtual wheelchair (or of the external device) represents how subjects are performing in the control of their mental task.

4 Use of the System

In order to help subjects to control the proposed paradigm, a progressive training must carry out. During the first phase of the training, subjects use the paradigm combining visual and audio-cued interface together. In a second phase, only the audio cue interface is used to select the different navigation commands.

It is accepted that a more immersive environment can help keep the subject's motivation, and, as a consequence, it could lead to better results [11]. For this reason, our group works on the development of different VEs in order to help subjects to get control of the proposed paradigm (Figure 3). In order to get immersive VEs, crucial elements to take into account are realism and stereoscopic vision. Therefore, the navigation paradigm is being applied in 3D environments that faithfully reproduce real-world scenarios in their look (textures, shininess, transparency and translucency), physics (collisions, gravity and inertia) and weather conditions as rain, snow and wind. VEs can be configured to disable some of the simulation features, so ease of navigation can be adjusted to the ability and expertise of the user. Immersion is further achieved with the addition of 3D sounds, which take into account the distance, power and speed (Doppler Effect is included) of the source.

To increase the degree of immersion, the VEs are projected on a large screen. The VEs are created with OpenGL for the graphics, OpenAL for the 3D audio, and ODE for physics simulation. The C programming language is used. Interaction between MATLAB and the VE is achieved with TCP/IP communications, which allowed us to use different machines for data acquisition and processing, and environment simulation and display.

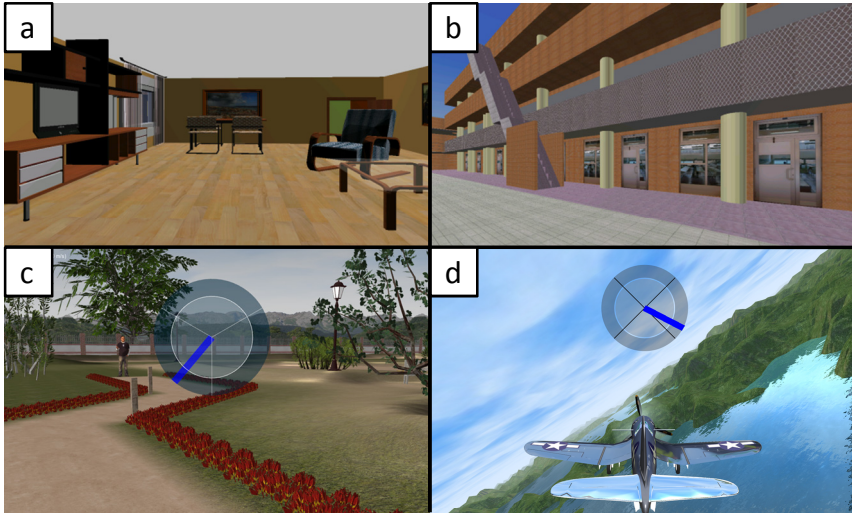


Fig. 3. Several VEs: a) Apartment, b) Engineering School of the University of Málaga, c) Park and corridor, d) BCI-controlled plane

By means of that versatility, users start navigating in an easy and attractive VE (Figure 3d): an environment without obstacles consisting of the control of a plane with 4 possible commands (rise, descend, turn right and left). Regarding the IC interface, the rotation speed of the feedback bar is fixed at 2.5 degrees per computation iteration (62.5 ms), so it takes 9 s to complete a turn if there is not any stop. The selection time changes among subjects, even between among sessions, in a range of 1-2 s. In fact, this VE is like a videogame but no instruction is provided. Subjects play and learn to control the plane using the graphical and the audio-cued interface.

Once they get used to the paradigm (firstly with visual and audio-cued interface, and then with only audio-cued) they progressively change to more sophisticated scenarios. These scenarios have been created in order to be recognized by the users as familiar. One of these scenarios is a virtual apartment to explore (Figure 3a). In this virtual apartment subjects can freely decide where to go, however, some obstacles must be avoided (furniture, walls,...). Another scenario is a known place, such as, the engineering school of the University of Málaga (Figure 3b), where most subjects come from. With this scenario, subjects are instructed to go to specific places, for example the bar of the school. In this second phase of the training, subjects can choose between the virtual apartment and the engineering school to navigate.

Finally, once the subjects got some control to navigate using the audio-cued interface, they participate in an experiment in which they have to follow a prefixed path to reach, as fast as they could, an avatar placed at the end of it (Figure 3 c). This path is located in a 3D virtual park. If the movement leads the subjects out of this path, the wheelchair collides with an invisible wall, so the movement finishes. During

the experiments, subjects are looking at a large stereoscopic screen (2 x 1.5 m) placed at a distance of 3 m, wearing polarized glasses and earphones.

Figure 4 shows the different paths followed by a single subject in 3 different runs (the starting point is on the right side). In order to establish a criterion to compare the performance of the subject, a reference path is presented in the figure with a white line. This path is achieved with the same paradigm, but an operator uses a function generator to manually emulate the brain activity, so the bar length could be easily controlled. This path can be considered close to the optimal path that can be achieved with this paradigm. Every point where a collision happened is signalled with an arrow, and each command with a symbol in the paths. This subject collided once in run 1 and twice in run 3. Run 2 was carry out without collisions. The number of commands used is 18.3 (average between the 3 runs), that is, only 2 times the number of commands using a manual control (9).

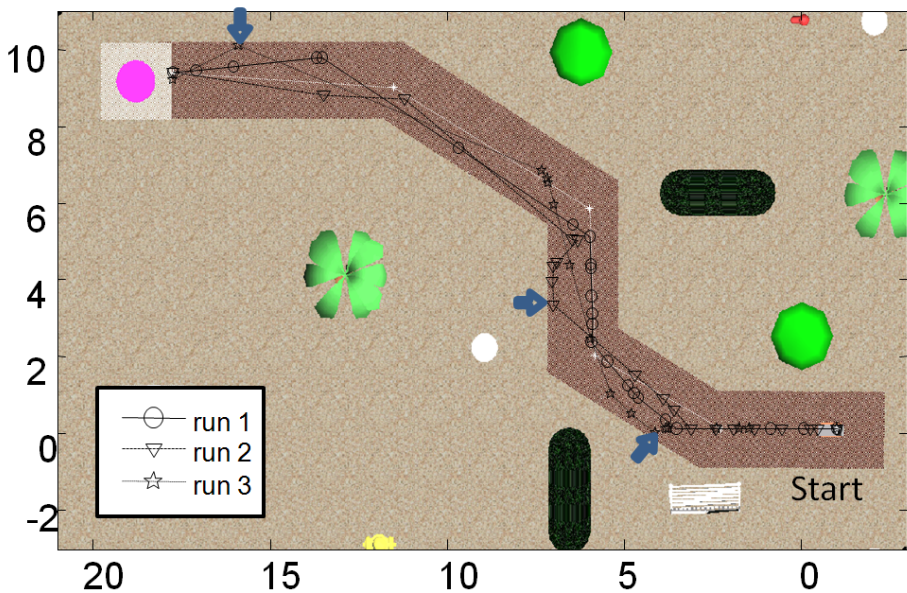


Fig. 4. Paths followed by a subject in a virtual park

5 Discussion and Conclusion

A new paradigm has been proposed to navigate through a VE using only two mental tasks, which keeps the classification accuracy at its maximum. The mapping of these mental tasks into a higher number of commands makes it possible to freely move with a friendly paradigm of interaction. This paradigm can easily be modified to let the subjects choose among a higher number of commands (for example, it could be included a fourth command to move backwards).

The subjects' motivation is a very important factor in their performance. For this reason, the use of VE with a higher degree of immersion could improve the results. Different applications of VR to BCI systems have been presented, showing how it not only helps to keep user's interest and motivation, but actually has a positive effect in the user's performance and training. Among these applications, we have focused on those oriented to the use of VR as a tool to test and train with several navigation paradigms, especially on the UMA-BCI. This last paradigm has shown its usability with encouraging results supported by several experiments. This navigation paradigm does not need to be applied only in VR, it can be used in other scenarios, for example, to control a robot in a experimental situation, or a real wheelchair.

Acknowledgments. This work was supported in part by the Innovation, Science and Enterprise Council of the Junta de Andalucía (Spain), project P07-TIC-03310.

References

1. Wolpaw, J.R., Birbaumer, N., McFarland, D.J., et al.: Brain-Computer Interfaces for Communication and Control. *Clinical Neurophysiology* 113, 767–791 (2002)
2. Mak, J.N., Wolpaw, J.R.: Clinical Applications of Brain-Computer Interfaces: Current State and Future Prospects. *IEEE Reviews in Biomedical Engineering* 2, 187–199 (2009)
3. Birbaumer, N., Kübler, A., Ghanayim, N., et al.: The Thought Translation Device (TTD) for Completely Paralyzed Patients. *IEEE Transactions on Rehabilitation Engineering* 8, 190–193 (2000)
4. Farwell, L.A., Donchin, E.: Talking Off the Top of Your Head: Toward a Mental Prosthesis Utilizing Event-Related Brain Potentials. *Electroencephalogr. Clin. Neurophysiol.* 70, 510–523 (1988)
5. Wolpaw, J.R., McFarland, D.J., Vaughan, T.M.: Brain-Computer Interface Research at the Wadsworth Center. *IEEE Trans. Rehabil. Eng.* 8, 222–226 (2000)
6. Kübler, A., Müller, K.R.: An introduction to brain-computer interfacing. In: Dornhege, G., Millán, J.d.R., Hinterberger, T., et al. (eds.) *Toward Brain-Computer Interfacing*, pp. 1–25. MIT Press, Cambridge (2007)
7. Neuper, C., Pfurtscheller, G.: Motor imagery and ERD. In: Pfurtscheller, G., Lopes da Silva, F.H. (eds.) *Event-Related Desynchronization. Handbook of Electroencephalography and Clinical Neurophysiology. Revised Series*, vol. 6, pp. 303–325. Elsevier, Amsterdam (1999)
8. Games and Brain-Computer Interfaces: The state of the Art, Boris Reuderink
9. Lécuyer, A., Lotte, F., Reilly, R.B., et al.: Brain-Computer Interfaces, Virtual Reality, and Videogames. *Computer* 41, 66–72 (2008)
10. Leeb, R., Scherer, R., Keinrath, C., et al.: Combining BCI and Virtual Reality: Scouting Virtual Worlds. In: Dornhege, G., Millán, J.d.R., Hinterberger, T., et al. (eds.) *Toward Brain-Computer Interfacing*, pp. 393–408. MIT Press, Cambridge (2007)
11. Ron-Angevin, R., Díaz-Estrella, A.: Brain-Computer Interface: Changes in Performance using Virtual Reality Techniques. *Neurosci. Lett.* 449, 123–127 (2009)
12. Leeb, R., Lee, F., Keinrath, C., et al.: Brain-Computer Communication: Motivation, Aim, and Impact of Exploring a Virtual Apartment. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 15, 473–482 (2007)

13. Leeb, R., Friedman, D., Müller-Putz, G.R., et al.: Self-Paced (Asynchronous) BCI Control of a Wheelchair in Virtual Environments: A Case Study with a Tetraplegic. *Computational Intelligence and Neuroscience* (2007)
14. Leeb, R., Settgast, V., Fellner, D., et al.: Self-Paced Exploration of the Austrian National Library through Thought. *International Journal of Bioelectromagnetism* 9, 237–244 (2007)
15. Tsui, C.S.L., Gan, J.Q.: Asynchronous BCI Control of a Robot Simulator with Supervised Online Training. In: Yin, H., Tino, P., Corchado, E., Byrne, W., Yao, X. (eds.) *IDEAL 2007*. LNCS, vol. 4881, pp. 125–134. Springer, Heidelberg (2007)
16. Scherer, R., Lee, F., Schlögl, A., et al.: Toward Self-Paced Brain-Computer Communication: Navigation through Virtual Worlds. *IEEE Transactions on Biomedical Engineering* 55, 675–682 (2008)
17. Obermaier, B., Neuper, C., Guger, C., et al.: Information Transfer Rate in a Five-Classes Brain-Computer Interface. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 9, 283–288 (2001)
18. Kronegg, J., Chanel, G., Voloshynovskiy, S., et al.: EEG-Based Synchronized Brain-Computer Interfaces: A Model for Optimizing the Number of Mental Tasks. *IEEE Transactions on Neural Systems and Rehabilitation Engineering: A Publication of the IEEE Engineering in Medicine and Biology Society* 15, 50–58 (2007)
19. Guger, C., Edlinger, G., Harkam, W., et al.: How Many People are Able to Operate an EEG-Based Brain-Computer Interface (BCI)? *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 11, 145–147 (2003)
20. Pfurtscheller, G.: Quantification of ERD and ERS in the time domain. In: Pfurtscheller, G., Lopes da Silva, F.H. (eds.) *Event-Related Desynchronization*. Handbook of Electroencephalography and Clinical NeuroPhysiology. Revised Series, vol. 6, pp. 89–105. Elsevier, Amsterdam (1999)
21. Schlögl, A., Kronegg, J., Huggins, J.E., et al.: Evaluation Criteria for BCI Research. In: Dornhege, G., Millán, J.d.R., Hinterberger, T., et al. (eds.) *Toward Brain-Computer Interfacing*, pp. 327–342. The MIT Press, Cambridge (2007)
22. Blankertz, B., Dornhege, G., Krauledat, M., et al.: The Berlin Brain-computer interface presents the novel mental typewriter Hex-o-Spell. In: *3rd International Brain-Computer Interface Workshop and Training course*, Graz, Austria, pp. 108–109 (2006)

Steady State Visual Evoked Potential Based Computer Gaming – The Maze

Nikolay Chumerin, Nikolay V. Manyakov, Adrien Combaz, Arne Robben,
Marijn van Vliet, and Marc M. Van Hulle

KU Leuven, Laboratorium voor Neuro- en Psychofysiologie, Campus Gasthuisberg,
Herestraat 49, B-3000 Leuven, Belgium

{Nikolay.Chumerin,NikolayV.Manyakov,Adrien.Combaz,
Arne.Robben,Marijn.vanVliet,Marc.VanHulle}@med.kuleuven.be

Abstract. We introduce a game, called “The Maze”, as a brain-computer interface (BCI) application in which an avatar is navigated through a maze by analyzing the player’s steady-state visual evoked potential (SSVEP) responses recorded with electroencephalography (EEG). The same computer screen is used for displaying the game environment and for the visual stimulation. The algorithms for EEG data processing and SSVEP detection are discussed in depth. We propose the system parameter values, which provide an acceptable trade-off between the game control accuracy and interactivity.

1 Introduction

With a brain-computer interface (BCI) brain activity is read and used for enabling a subject to interact with the external world, without involving any muscular activity or peripheral nerves. BCI is now widely regarded as one of the most successful applications of the neurosciences and is in a position to significantly improve the quality of life of patients suffering from amyotrophic lateral sclerosis, stroke, brain/spinal cord injury, cerebral palsy, muscular dystrophy, etc [1].

In this work we consider non-invasive, electroencephalography (EEG)-based BCI method based on the steady-state visual evoked potential (SSVEP). SSVEP is a response recorded from the occipital pole of a brain on the repetitive presentation of visual stimuli (*i.e.*, flickering stimuli). When stimulation is at a sufficiently high rate (starting from 6 Hz), the individual transient EEG responses overlap, leading to a steady state signal: the signal resonates at the stimulus rate and its multipliers [2]. This means that, when a subject is looking at a stimulus flickering at frequency f , one can detect f , $2f$, $3f$, ... in the recorded EEG data. Since the amplitude of a typical EEG signal decreases as $1/f$ in the spectral domain [3], the higher harmonics become less prominent. Furthermore, SSVEP is embedded in other on-going brain activity and (recording) noise. Thus, when considering a too small recording interval, erroneous detections are quite likely to occur. To overcome this problem, averaging over several

recording intervals [4], or recording over longer time intervals [5] are often used for increasing the signal-to-noise ratio (SNR) in the spectral domain. Finally, in order to increase the usability and the information transfer rate of the SSVEP-based BCI, the user should be able to select one of several commands, which means that the system should be able to reliably detect several (n_f) frequencies f_1, \dots, f_{n_f} . This makes the frequency detection problem more complex, calling for an efficient signal processing and decoding algorithm.

BCIs were initially aimed for medical purposes, but currently they attract a lot of attention from the entertainment community [6], since they can be used as a new interface, *e.g.*, for mind-controlled games, or for remotely controlling devices. Several studies on SSVEP BCI gaming were published during the last few years [7,8].

In this paper, we present a novel BCI SSVEP game, which achieves good performance thanks to an appropriate detection algorithm combined with spatial filtering. We also discuss some necessary modifications to the game strategy, which can make this brain game more easy to use and more attractive.

2 Methods

2.1 EEG Data Acquisition

The EEG recordings were performed using a prototype of an ultra low-power 8-channels wireless EEG system, which was developed by imec¹, and built around their ultra-low power 8-channel EEG amplifier chip [9]. The data are transmitted at a sampling rate of 1000 Hz, for each channel. We used an electrode cap with large filling holes and sockets for mounting of active Ag/AgCl electrodes (Acti-Cap, Brain Products). The recordings were made with eight electrodes located on the occipital pole (covering the primary visual cortex), namely at positions P3, Pz, P4, PO9, O1, Oz, O2, PO10, according to the international 10–20 electrode placement system. The reference electrode and ground were placed on the left and right mastoids, respectively.

The raw EEG signals are filtered above 3 Hz, with a fourth order zero-phase digital Butterworth filter, so as to remove the DC component and the low frequency drift. A notch filter is also applied to remove the 50 Hz powerline interference.

2.2 Calibration Stage

The game uses only four commands for navigating the avatar through the maze: “left”, “up”, “right” and “down”, hence, four stimulation frequencies are needed. During our preliminary experiments, we noticed that the optimal set of stimulation frequencies is very subject dependent. This motivated us to introduce a calibration stage, preceding the actual game play, for locating the frequency band, consisting of four frequencies, that evoke prominent SSVEP responses in

¹ <http://www.imec.be>

the subject’s EEG signal. To this end, we propose a “scanning” procedure, consisting of several blocks. In each block, the subject is visually stimulated for 15 seconds by a flickering screen ($\approx 28^\circ \times 20^\circ$), after which a black screen is presented for 2 seconds. The number of blocks in the calibration stage is defined by the number of available stimulation frequencies. We have used a laptop with a bright 15,4" LCD screen with a 60 Hz refresh rate. In order to arrive at a visual stimulation with stable frequencies, we show an intense stimulus for k frames, and a less intense stimulus for the next l frames, hence, the flickering period of the stimulus is $k + l$ frames and the corresponding stimulus frequency is $r/(k + l)$, where r is the screen’s refresh rate. Using this simple strategy, one can stimulate the subject with the frequencies that are dividers of the screen refresh rate: 30 Hz (60/2), 20 Hz (60/3), 15 Hz (60/4), and so on. We grouped these frequencies into overlapping bands, for which each band contains four consecutive stimulation frequencies (*e.g.*, band 1: [6 Hz, 6.66 Hz, 7.5 Hz, 8.57 Hz], band 2: [6.66 Hz, 7.5 Hz, 8.57 Hz, 10 Hz], and so on). After stimulation, we visually analyze the spectrograms of the recorded EEG signals, and select the “best” band of frequencies to be used in the game. We have to admit that this frequency selection procedure is subjective, and probably not optimal, calling for an automated procedure.

2.3 Spatial Filtering

Following the minimum energy combination method proposed in [10], we use a spatial filter designed in the following way: a linear combination of the channels is sought that decreases the noise level of the resulting weighted signals at the specific frequencies we want to detect (namely, the frequencies of the oscillations evoked by the periodically flickering stimuli, and their harmonics). This can be done in two steps. In the first step, all information related to the frequencies of interest must be eliminated from the recorded signals. The resulting signals contain only information that is “uninteresting” in the context of our application, and, therefore, could be considered as noise components of the original signals. In the second step, we look for a linear combination that minimizes the variance of the weighted sum of the “noisy” signals obtained in the first step. Eventually, we apply this linear combination to the original signals, resulting in signals with a lower level of noise.

The first step can be done by subtracting from the EEG signal all the components corresponding to the stimulation frequencies and their harmonics. Formally, this can be done in the following way. Let us consider the input signal, sampled over a time window of duration T with sampling frequency F_s , as a matrix \mathbf{X} with channels in columns and samples in rows. Then, one needs to construct a matrix \mathbf{A} , which should have the same number of rows as \mathbf{X} and as the number of columns twice the number of all considered frequencies (including harmonics). For a given time instant t_i (corresponding to the i -th sample in \mathbf{X}) and frequency f_j (from the full list of stimulation frequencies including the harmonics), the corresponding elements $a_{i,2j-1}$ and $a_{i,2j}$ of the matrix \mathbf{A} are computed as $a_{i,2j-1} = \sin(2\pi f_j t_i)$ and $a_{i,2j} = \cos(2\pi f_j t_i)$. For example,

considering only $n_f = 2$ frequencies with their $N_h = 2$ harmonics and a time interval of $T = 2$ seconds, sampled at $F_s = 1000$ Hz, the matrix \mathbf{A} would have $2 \times n_f \times (1 + N_h) = 2 \times 2 \times 3 = 12$ columns and $T \times F_s = 2000$ rows. The most “interesting” components of the signal \mathbf{X} can be obtained from \mathbf{A} by a projection determined by the matrix $\mathbf{P}_\mathbf{A} = \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T$. Using $\mathbf{P}_\mathbf{A}$ the original signal without the “interesting” information is estimated as $\tilde{\mathbf{X}} = \mathbf{X} - \mathbf{P}_\mathbf{A} \mathbf{X}$. Those remaining signals $\tilde{\mathbf{X}}$ can be considered as noise components of the original signals (*i.e.*, the brain activity not related to the visual stimulation).

In the second step, we use an approach based on Principal Component Analysis (PCA) to find a linear combination of the input data for which the noise variance is minimal. A PCA transforms a number of possibly correlated variables into uncorrelated ones, called principal components, defined as projections of the input data onto the corresponding principal vectors. By convention, the first principal component captures the largest variance, the second principal component the second largest variance, and so on. Given that the input data comes from the previous step, and contains mostly noise, the projection onto the last principal component direction is the desired linear combination of the channels, *i.e.*, one that reduces the noise in the best way (*i.e.*, making the noise variance minimal).

The conventional PCA approach estimates the principal vectors as eigenvectors of the covariance matrix $\Sigma = E\{\tilde{\mathbf{X}}^T \tilde{\mathbf{X}}\}$, where $E\{\cdot\}$ denotes the statistical expectancy². Since the considered EEG signal has 8 channels, Σ has size 8×8 , is positive semidefinite and, therefore, it is possible to find a set of 8 orthonormal eigenvectors (represented as columns of a matrix V), such that $\Lambda = V \Sigma V^T$, where Λ is a diagonal matrix of the corresponding eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_8 \geq 0$. Then, the K last (smallest) eigenvalues are selected such that K is maximal, and $\sum_{k=1}^K \lambda_{9-k} / \sum_{j=1}^8 \lambda_j < 0.1$ is satisfied. The corresponding K eigenvectors, arranged as columns of a matrix V_K , specify a linear transformation that efficiently reduces the noise power in the signal $\tilde{\mathbf{X}}$. The same noise-reducing property of V_K is valid for the original signal \mathbf{X} . Assuming that V_K would reduce the variance of the noise more than the variance of the signal of interest, the signal that is spatially filtered in this way, $\mathbf{S} = V_K \mathbf{X}$, would have greater (or, at least, not smaller) SNR [10].

2.4 Classification

The straight-forward approach to select one frequency (among several possible candidates) present in the analyzed signal is based on a direct analysis of the signal power function $P(f)$ that is defined as follows:

$$P(f) = \left(\sum_t s(t) \sin(2\pi ft) \right)^2 + \left(\sum_t s(t) \cos(2\pi ft) \right)^2,$$

² Since the original signal is high-pass filtered above 3 Hz, the DC component is removed and, therefore, the filtered data are centered (the mean is close to zero).

where $s(t)$ is the signal after spatial filtering. Note that the right-hand part of this equation is the squared Discrete Fourier Transform magnitude at the frequency of interest [10]. The “winner” frequency f^* can then be selected as the frequency with maximal (among all considered frequencies f_1, f_2, \dots, f_{n_f}) power amplitude:

$$f^* = \arg \max_{f_1, \dots, f_{n_f}} P(f).$$

Unfortunately, in our case, this direct method is not applicable due to the nature of the EEG signal: the corresponding power function decreases (similarly to $1/f$) with increasing f . In this case, the true dominant frequency could have an power amplitude less than the other considered lower frequencies. In [5] it was shown that the SNR does not decrease with increasing frequency, but remains nearly constant. Relying on this finding, one can select the “winner” frequency as the one for which the SNR is maximal, $P(f)/\sigma(f)$, where $\sigma(f)$ is an estimation of the noise power for frequency f .

The noise power estimation is not a trivial task. One way to do this is to record extra EEG data from the subject, without visual stimulation. In this case, the power of the considered frequencies in the recorded signal should correspond to the noise level. Despite its apparent simplicity, this method has at least two drawbacks: 1) an extra (calibration) EEG recording session is needed, and 2) the noise level changes over time and the pre-estimated values could significantly deviate from the actual ones. To overcome these drawbacks, we need an efficient on-line method of noise power estimation. As a possible solution, one can try to approximate the desired noise power $\sigma(\tilde{f})$ for a frequency of interest \tilde{f} using values of $P(f)$ from a close neighborhood $O(\tilde{f})$ of the considered frequency \tilde{f} . A simple averaging $\sigma(\tilde{f}) \approx E\{P(f)\}_{f \in O(\tilde{f}) \setminus \tilde{f}}$ produces unstable (jittering) estimates if the size of the neighborhood $O(\tilde{f})$ is small. Additionally, a large neighborhood could contain several frequencies of interest that could bias the estimate of $\sigma(\tilde{f})$.

In our work, we have used an approximation of noise based on an autoregressive modeling of the data, after excluding all information about the flickering, *i.e.*, of signals $\tilde{\mathbf{S}} = V_K \tilde{\mathbf{X}}$ (see previous subsection). The rationale behind this approach is that the autoregressive model can be considered as a filter (working through convolution), in terms of ordinary products between the transformed signals and the filter coefficients in the frequency domain. Since we assume that the prediction error in the autoregressive model is uncorrelated white noise, we have a flat power spectral density for it with a magnitude that is a function of the variance of the noise. Thus, the Fourier transformations of the regression coefficients a_j (estimated, for example, with the use of the Yule-Walker equations) show us the influence of the frequency content of particular signals on the white noise variance ($\tilde{\sigma}$). By assessing such transforms, we can obtain an approximation of the power of the signal $\tilde{\mathbf{S}}$.

More formally, we have:

$$\sigma(f) = \frac{\pi T}{4} \frac{\tilde{\sigma}^2}{|1 - \sum_{j=1}^p a_j \exp(-2\pi i j f / F_s)|},$$

where T is the length of the signal, $i = \sqrt{-1}$, p is the order of the regression model and F_s is the sampling frequency. Since for the detection of each stimulation frequency, we use several channels and several harmonics, we could combine separate values of the SNR as:

$$Q(f) = \frac{1}{K(N_h + 1)} \sum_{k=1}^K \sum_{h=1}^{N_h+1} P_k(hf) / \sigma_k(hf),$$

where K is the dimensionality of the signal \mathbf{S} and $(N_h + 1)$ is the number of the multipliers of the considered frequency f (one fundamental frequency plus its N_h harmonics).

The “winner” frequency f^* is defined as the frequency with the largest index $Q(\cdot)$ among all frequencies of interest:

$$f^* = \arg \max_{f_1, \dots, f_{n_f}} Q(f).$$

2.5 Game Design and Implementation

We have developed a SSVEP-based BCI game “The Maze”, in which the player can control an avatar in a simple maze-like environment. The task is to navigate the avatar (depicted as Homer Simpson’s head) to the target (*i.e.*, a donut) through the maze (see Fig. [11](#)). The game has several pre-defined levels of increasing complexity. A random maze mode is also available. The player can control the avatar by looking at flickering arrows (showing the direction of the avatar’s next move) placed in the periphery of the maze. Each arrow is flickering with its own unique frequency taken from the selected frequency band (see Section [2.2](#)). The selection of the frequencies can be predefined or set according to the player’s preferences.

The game is implemented in Matlab as a client-server application and can run either in parallel Matlab mode (as two labs) or on two Matlab sessions started as separate applications. The server part is responsible for the EEG data acquisition, processing and classification. The client part is responsible for the game logic, user interface and rendering. The client-server communication is implemented using sockets and due to a minimal data transfer rate (during the game only commands are sent from the server to the client) it can work over a regular network, allowing also (optionally) to run the game on two different computers. For the accurate (in terms of timing) visualization of the flickering stimuli, we have used Psychtoolbox 3 (<http://psychtoolbox.org>).

To reach a decision, the server needs to analyze the EEG data acquired over the last T seconds. In the game, T is one of the tuning parameters (must be set

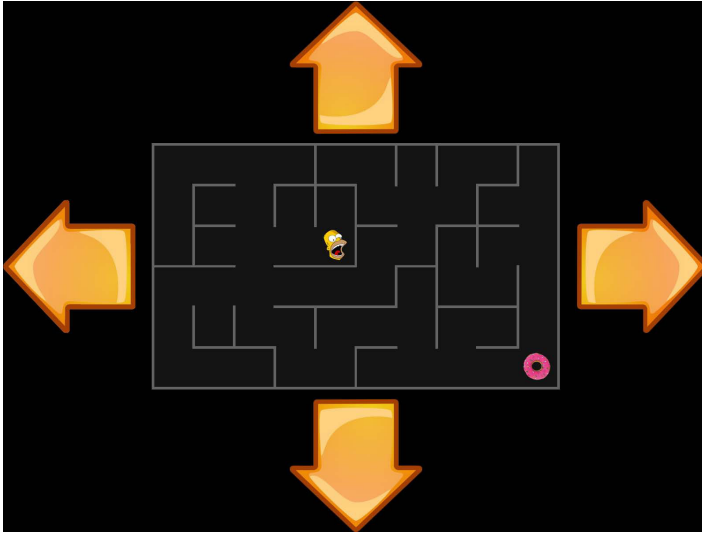


Fig. 1. Snapshot of “The Maze” game

before the game starts), which controls the game latency. Decreasing T makes the game more responsive, but in the same time it makes the interaction less accurate, resulting in wrong navigation decisions. By default, a new portion of the EEG data is collected every 200 ms. The server analyzes the new (updated) data window and detects the dominant frequency using the method described above. The command corresponding to the selected frequency is sent to the client also every 200 ms, thus, the server’s update frequency is 5 Hz. The final decision (the command that is executed) is made by the client using the history of the last m frequency detections: if in the queue of the last m detected frequencies there is a frequency with more than $m/2$ occurrences, then this frequency is considered to be the “final winner”, otherwise no decision is made.

As mentioned above, the game control has an unavoidable time lag. In order to “hide” this latency, we let the avatar change its navigation direction only in so-called decision points: as the avatar starts to move, it will not stop until it reaches the next decision points on its way. This allows the player to use this period of “uncontrolled avatar movement” for planning (by looking on appropriate flickering arrow) the next navigation direction. By the time the avatar reaches the next decision point, the EEG data window, which is to be analyzed, already contains the SSVEP response corresponding to the new navigation direction.

2.6 Influence of Window Size and Decision Queue Length on Accuracy

To assess the best combination of the window size T and the decision queue length m , we have studied their influence on the classification accuracy. Six

healthy subjects (all male, aged 24–34 with average age 28.3, four righthanded, one lefthanded and one bothhanded) participated in the experiment. Only one subject had prior experience with SSVEP-based BCI. For each subject, several sessions with different stimulation frequency sets were recorded, but we present the results only for those sessions, for which the stimulation frequencies coincide with the ones that are determined with the calibration stage. Each subject was presented with a specially designed level of the game, and was asked to consequently look at each one of four flickering arrows for 20 seconds followed by 10 seconds of rest, so the full round of four stimuli (flickering arrows) was $4 \times (20 + 10) = 120$ seconds. The stimulus to attend to was marked with the words “look here”. Each recording session consisted of two rounds and, thus, lasted 4 minutes. The recorded EEG data were then analyzed off-line using exactly the same mechanism as in the game: for each position of the sliding window (of size T) the detected frequency was pushed in the queue (of length m), and the final decision was based on reaching more than 50% of the votes. Due to the design of the experiment, the true winner frequency is known for each moment of time, which enables us to estimate the accuracy.

3 Results and Discussion

The results of the experiment described in Section 2.6 are shown in Table 1. With the accuracy of the frequency classification we mean the ratio of the correct decisions with respect to all decisions made by the classifier. Note, that the chance level of accuracy in this experiment is 25%.

From Table 1 it can be seen that, in general, the longer queues of the decision making mechanism lead to a better accuracy of the game control. The drawback of the longer queues is an additional latency. To reduce the later, the server’s update frequency (the actual one is 5 Hz) can be increased. This, in turn, increases the computational load (mostly on the server part).

Based on our experience (also supported by the data from Table 1), we can recommend to use the window size $T = 3$ and the queue length $m = 5$ (or more) as default values for an acceptable gameplay.

Unfortunately, the information transfer rate (ITR) commonly used as a performance measure for BCIs, is not relevant for the game, at least in its actual form. By design, the locations of the decision points depend on the (randomly generated) maze, and, therefore, the decisions themselves are made at an irregular rate, which, in turn, does not allow for a proper ITR estimation.

A few more issues concerning the visual stimulation and the game design need to be discussed. Even though the visual stimulation in the calibration stage (one full-screen stimulus) differs from the one used in the game (four simultaneously flickering arrows, see Figure 1), we strongly believe that the frequencies selected in such a way are also well suited for the game control. This belief has been indirectly supported during our experiments (see Section 2.6): the frequency sets, different from the ones selected during the calibration stage, in most cases yield less accurate detections.

Table 1. Classification accuracy as a function of window size T and decision queue length m

T (s)	m	subject 1	subject 2	subject 3	subject 4	subject 5	subject 6
1	1	57.14%	91.96%	75.22%	53.35%	49.78%	47.32%
	3	58.80%	93.06%	78.24%	52.31%	49.54%	48.38%
	5	59.13%	94.71%	81.49%	53.61%	50.00%	48.08%
2	1	70.83%	99.51%	88.97%	69.36%	64.71%	53.92%
	3	72.70%	100.00%	89.03%	69.39%	65.56%	54.34%
	5	73.14%	100.00%	89.36%	71.01%	66.22%	54.52%
3	1	74.18%	100.00%	92.66%	81.79%	69.84%	62.50%
	3	75.28%	100.00%	92.05%	82.39%	71.31%	62.50%
	5	76.19%	100.00%	92.26%	82.74%	72.02%	62.20%
4	1	73.17%	100.00%	94.82%	86.59%	70.43%	63.11%
	3	74.68%	100.00%	94.55%	87.18%	70.19%	63.14%
	5	75.68%	100.00%	94.93%	89.19%	70.95%	63.51%
5	1	65.28%	100.00%	94.10%	88.89%	65.63%	63.89%
	3	65.81%	100.00%	94.49%	88.97%	65.07%	62.87%
	5	65.63%	100.00%	93.36%	89.45%	64.45%	63.28%

One of the drawbacks of SSVEP-based BCIs with dynamic environment and fixed locations of stimuli is the frequent change of the subject’s gaze during the gameplay, which leads to a discontinuous visual stimulation. To avoid this, we introduced an optional mode where the stimuli (arrows) are locked close to the avatar and move with it during the game, which might make the game more comfortable to play.

Several subjects have noticed that the textured stimuli are easier to concentrate on than the uniform ones. Some of our subjects preferred the yellow color of the stimuli to the white color, which partially might be explained by a characteristic feature of the yellow light stimulation: it elicits an SSVEP response of a strength that is less dependent on the stimulation frequency than other colors [11].

BCI-based gaming is research direction that is still in its infancy, and still a lot of issues to be tackled before it could become accepted in the gaming community. All these issues, including the ones discussed above, clearly indicate the necessity of further BCI research, in general, and the development of suitable applications for interactive entertainment, in particular.

Acknowledgments. NC is supported by IST-2007-217077, NVM is supported by research grant GOA 10/019, AC and AR are supported by IWT doctoral grants, MvV is supported by G.05809, MMVH is supported by PFV/10/008, CREA/07/027, G.0588.09, IUAP P6/29, GOA 10/019, IST-2007-217077, and the SWIFT prize of the King Baudouin Foundation of Belgium. The authors grateful to Refet Firat Yazicioglu, Tom Torfs and Cris Van Hoof from imec, Leuven, Belgium, for providing the wireless EEG system.

References

1. Mak, J.N., Wolpaw, J.R.: Clinical Application of Brain-Computer Interface: Current State and Future Prospects. *IEEE Rev. Biomed. Eng.* 2, 187–199 (2009)
2. Herrmann, C.S.: Human EEG Responses to 1–100 Hz Flicker: Resonance Phenomena in Visual Cortex and Their Potential Correlation to Cognitive Phenomena. *Exp. Brain Res.* 137, 346–353 (2001)
3. Allegrini, P., Menicucci, D., Bedini, R., Fronzoni, L., Gemignani, A., Grigolini, P., West, B.J., Paradisi, P.: Spontaneous brain activity as a source of ideal $1/f$ noise. *Phys. Rev. E* 80, 061914 (2009)
4. Manyakov, N.V., Chumerin, N., Combaz, A., Robben, A., Van Hulle, M.M.: Decoding SSVEP Responses Using Time Domain Classification. In: International Conference on Fuzzy Computation and 2nd International Conference on Neural Computation, pp. 376–380 (2010)
5. Wang, Y., Wang, R., Gao, X., Hong, B., Gao, S.: A practical VEP-based brain-computer interface. *IEEE TNSRE* 14(2), 234–239 (2006)
6. Nijholt, A., Plass-Oude Bos, D., Reuderink, B.: Turning shortcomings into challenges: Brain-computer interfaces for games. *Entertainment Comp.* 1(2), 85–94 (2009)
7. Martinez, P., Bakardjian, H., Cichocki, A.: Fully online multicommand brain-computer interface with visual neurofeedback using SSVEP paradigm. *Computational Intelligence and Neuroscience* vol. 2007, article ID 94561 (2007)
8. Lalor, E.C., Kelly, S.P., Finucane, C., Burke, R., Smith, R., Reilly, R.B., McDarby, G.: Steady-state VEP-based brain-computer interface control in an immersive 3D gaming environment. *EUEASIP J. on Appl. Sign. Proc.* 19, 3156–3164 (2005)
9. Yazicioglu, R.F., Torfs, T., Merken, P., Penders, J., Leonov, V., Puers, R., Gyssels, B., Van Hoof, C.: Ultra-low-power biopotential interfaces and their applications in wearable and implantable systems. *Microel. J.* 40(9), 1313–1321 (2009)
10. Friman, O., Volosyak, I., Gräser, A.: Multiple channel detection of steady-state visual evoked potentials for brain-computer interfaces. *IEEE TBE* 54(4), 742–750 (2007)
11. Regan, D.: An effect of stimulus colour on average steady-state potentials evoked in man. *Nature* 210(5040), 1056–1057 (1966)

Single Value Devices

Angelika Mader, Edwin Dertien, and Dennis Reidsma

University of Twente, The Netherlands

{a.h.mader,e.c.dertien,d.reidsma}@utwente.nl

Abstract. We live in a world of continuous information overflow, but the quality of information and communication is suffering. Single value devices contribute to information and communication quality by focussing on one explicit, relevant piece of information. The information is decoupled from a computer and represented in an object, integrated into daily life.

The contribution of this paper is on different levels: Firstly, we identify single value devices as a class, and, secondly, illustrate it through examples in a survey. Thirdly, we collect characterizations of single value devices into a taxonomy. The taxonomy also provides a collection of design choices that allow one to more easily find new combinations or alternatives, and that facilitate the design of new, meaningful, effective and working objects. Finally, when we want to step from experimental examples to commercializable *products*, a number of issues become relevant that are identified and discussed in the remainder of this paper.

1 Introduction

After the quantity explosion of information and communication, the desire for quality arises, as also expressed by the slow media movement [19]. *Single value devices* are objects that filter one item out of the constant cloud of information, and display it in isolation on a physical object, making this information much more accessible and prominent and integrating it in our daily lives. These key properties give single value devices the potential to increase the quality of information.

A single value can carry a huge amount of information. The single bit of information that a friend is online on ICQ creates an awareness of the other person, an emotion of sharing presence and activity, and may suggest an action, which is to contact the friend. Embodying the representation in a dedicated (everyday, or especially designed) object has additional advantages. Firstly, it brings more immediacy to the information, compared with opening a laptop, connecting to the internet and searching for the information. Secondly, dedicated objects allow for an almost unlimited variety of designs to represent the information and interact with the user, such as sound, touch, light, movement, whatever can be invented using actuators and sensors, and what people find easy and pleasant to perceive. As we will discuss later, it also gives more possibilities to design for *emotion*. Finally, dedicated objects, more than traditional screen-based devices, allow the technology and the information representation to move into the

background or periphery. The information comes only into focus when needed, and the user is not overburdened with information (cf. *ubiquitous computing* and *calm technology* [29,8,30,27]).

Most existing single value devices come from conceptual experiments and from art and exist only as prototypes. In order to get to mature *products* and to design meaningful, effective and working objects, an understanding of the design choices and their consequences is necessary, which is the core contribution of this paper. Our fundamental question is: *How to design meaningful and effective single value devices?* We will, first, approach this question by investigating the possible characteristics of single value devices. To this end, we present a survey of existing single value devices in Section 2; subsequently, we suggest a taxonomy for single value devices in Section 3. When taking the step from the proof-of-concept or artistic-exploration nature of most existing single value devices to commercially feasible products, a number of design issues becomes relevant, which are discussed in Section 4.

2 Survey of Single Value Devices

In this section we present a chronological survey of single value displays, in order to unfold the space of possible applications and approaches. The objects presented here are often prototypes and results from art projects.

- *Feather, Scent* and *Shaker* [23] are pairs of objects shared by two people. In “Feather” and “Scent”, one partner has a picture frame, and shows (s)he thinks of the other by shaking the frame. This message of connectedness is communicated to the partner at home in a manner reflecting the transience of thought: through a feather in a cylinder that is lifted by a little fan, or by vaporising essential oil in an aluminium bowl using a heating element. “Shaker” is meant for less intimate friends, and consists of a pair of handsized devices that, when shaken, cause a vibration of the other object.
- The *Dangling String* [30] is an installation for an office environment. It consists of one and a half meter of plastic spaghetti hanging from the ceiling, mounted to a small electric motor. The motor is triggered by the activity on an Ethernet cable. A very busy network causes a madly whirling string with a characteristic noise; a quiet network causes only a small twitch every few seconds. Placed in an unused corner of a hallway, the long string is visible and audible from many offices without being obtrusive.
- The level of web activity is displayed in [18] using ripples in a water tank. A solenoid-driven float triggered by “bits” of web activity creates ripples on the surface of the water; these are reflected on the ceiling using a strong light.
- Also for a working environment is the light installation of [15]. Posters of research projects on the corridor walls are illuminated by spotlights. The light intensity of each spot is determined by the number of hits on the corresponding project webpage over a period of time.
- The *Peek-A-Boo Surrogate*, as one of many examples in [16], is also for a working environment. It consists of a little figure that turns its face to the wall if the

person to which it is connected is not present in her or his office, and turns it front to the room, if the person is available.

- Using touch or temperature sensors, a *White Stone* [26] can detect when one partner takes it in her or his hand, which causes a coupled remote White Stone to produce a sound. A message can be sent back by triggering an internal heating device in the other White Stone.
- *Soft Air Communication* [26] refers to a pair of inflatable chairs that can sense weight and movements. At the corresponding chair these signals are then displayed by light and sound.
- The *Frame* [26] indicates presence or absence of family members. A photo in the Frame rises when the respective person is at home, or is dimmed otherwise. A receptor on a key ring or in a wallet captures the presence of the person.
- The *Kiss Communicator* [7] includes two devices, wireless connected by internet. The sender can blow on her device, which is displayed as a colour sequence on the other device.
- Also designed for partners is *LumiTouch* [9], is a pair of picture frames that are connected to a computer and may contain the photo of the partner. The frames allow for two modes of interaction: in one mode movement is detected which makes the other frame glowing. In the other mode a frame can be squeezed giving a different light effect on the other frame.
- The *Internet Tea Kettle* [2] tells adult children whether parents make tea, i.e. use the tea kettle, which is an inherent element of normal daily activity in the Japanese culture, and indicating that the parents are acting well.
- The *Ambient Orb* [13] is a light ball indicating stock market activity. It can be configured via a company website to display other information. It is a commercial product. Connection is through a radio data network.
- The *Data Fountain* [5] can display stock market information by the height of a water fountain. In the example given, the different heights of three vertical water jets reflect the relative exchange rates of the Yen, Euro, and Dollar.
- The *Fishtank* [25] is designed for motivating people to move more. Employees in an office wear a pedometer, to measure their movement. A fish representing them, displayed on a public screen, grows with their amount of movement.
- *Nabaztag* [3] is a networked robot rabbit with speech, movable ears, and colored LEDs. It has been used as a single value device in applications for communication with a spouse, display of aggregated weather information, and others.
- The *Flower Lamp* [4] opens up to bloom depending on the energy consumption at home: the less energy is consumed the more the flower opens.
- The *Hug Shirt* [21] allows to send a hug via SMS to a mobile phone of the person wearing the hug shirt, and via blue tooth the hug is transmitted to the shirt. Sensors in a hug shirt can capture heart beat, skin temperature and strength of a hug, actuators can physically reproduce these information.
- *Blossom* [22] is a very personal object for a woman reflecting her connectedness to her family roots in cyprus. The blossom is made of stamps that were sent from Cyprus to England at the same time as her family emigrated. It opens when a predetermined amount of rain is detected by a sensor on the family land

in Cyprus. The blossom opens only once to reflect the uniqueness of events in contrast to continuous availability of services.

- *Journeys between ourselves* [22] is a pair of necklaces made for a mother and adult daughter. When one touches her necklace the other's necklace starts trembling softly. The necklaces are very personal objects made for specific persons, where the design refers exclusively to shared memories of mother and daughter.
- The *Smart Umbrella* [28] gives a voice alert if its owner leaves the house while rain is predicted. It combines internet information about the weather (weather.com) with local information, the state of the door.
- The *Babbage Cabbage* [14] uses a red cabbage as display. The acidity level of the feeding water can be modified, changing the colour of the cabbage between violet, purple and green. The cabbage has been used as single value display, with the colour of the cabbage representing information such as health of family members, or the quality of global climate and environment.
- A playful competition device is *ikWin: google battle* [20], consisting of two platforms that can be extended to a couple of meters height. Two people can compete by getting on a platform each, and then giving their name as input. The number of google hits will move the platform up, and the one with more google hits will end up in a higher position.
- *Pairs* [10] is also meant for partners. Two paired objects tremble with increasing intensity when they come closer to each other, and stop if they are put together. Much attention is put on the objects themselves, made from wood, such that it is a pleasure to touch and put them together. Additionally, much effort was put in giving them an individual look and personal history. Technology used includes arduino and wireless internet connection.
- *Scottie* [6] is designed for communication between children in a hospital and their relatives. Each participant has a doll, sending messages is done by shaking the doll or knocking on it, messages received are transformed into vibration and colors. Technology used includes arduino, Bluetooth, and mobile phone.
- Tactile communication between remote parents and their children is supported by the *Huggy Pajama* [24]. It consists of a doll equipped with pressure sensors, to be hugged by a parent, and a haptic jacket, where, by air pressure, the sensation of a hug can be reproduced.
- The *Internet Enabled Furby* [12] is an example of an instrumented toy, which functions as room observer (light sensor) and primitive communication device (ears). It has an ethernet connection and is controlled by an arduino.
- The *CoConatch* [1] is designed as physical warning device for twitter. It can alert a user with sound, movement and light about new messages. The device is connected to a PC using a USB connection. The PC runs a tiny server application to connect the device with Twitter through internet.

3 Taxonomy of Single Value Devices

The single value devices discussed in the survey above are often highly individual projects, stemming as much from an artistic idea as from technological developments. Many examples concern connection to your loved ones, ranging

from simple presence awareness devices (The Frame [26], The Internet Tea Kettle [2]), to active communication devices (e.g., Journeys between ourselves [22], The Huggy Pajama [24]). Other examples focus on displaying practical information about one’s environment, such as The Smart Umbrella [28], and the Babbage Cabbage visualizing environmental issues [14]. Communication technologies range from internet to GSM text messages; some devices are realized as mass-producible objects whereas others are highly individual, one-time objects.

In the following we provide a taxonomy of single value devices. It also describes the design space: for each of the characteristics a design decision has to be taken. In this sense it can serve as a stimulation to reflect on choices made, to explore new combinations, to find white areas in the space of possible designs, and invent new characteristics for meaningful, surprising and playful applications.

1 What Are the Characteristics of the Information Displayed?

1.1 Information Direction and Communicative Intent. The flow of information may be between two humans (social information), or human and machine [8]. When an explicit **Communicative Intent** is involved, the connection will always be between humans, and may be unidirectional or bidirectional. Examples are Scottie [6] and the necklaces [22]. In contrast, **Status Information** is unidirectional, and may involve human-machine as well as human-human connections (note that the elderly relatives using the Internet Tea Kettle do not make tea *in order to* communicate this fact to their family members). Other examples are the Data Fountain [5] and Smart Umbrella [28].

1.2 Information Distance represents, for social information, the social distance between the information source and the receiver. It can be described as shells spanning from self to family to society and world [14].

1.3 Information Privacy. The information represented by the single device may be **Public**, e.g., taken from the internet, or **Private** (everything that has (or should have) only personal use). Most human-human connections fall in the latter category. The Smart Umbrella [28] combines both: the information that someone is leaving the house is private, the weather information is public.

1.4 Information Decoupling. Single value devices typically involve one or more aspects of decoupling. **Physical decoupling:** The displayed quantity is not necessary one single physical measurable phenomenon such as temperature, but can also be an aggregate value. For example, the “state of the global environment” displayed by the Babbage Cabbage is a complex aggregate of many information sources. **Geographic decoupling:** network communications allow us to completely decouple the display and the measured data geographically. Blossoms [22] are an good example: the blossom (in England) opens depending on rain quantity on Cyprus. **Temporal decoupling:** the values displayed need not be strictly related to real-time (as in ‘here and now’). The project poster spot lights [15] display historical data. Other devices might target, e.g., awareness of

changes in bodily health over time, or engender historical awareness by showing the climate at a certain location, ten years in the past.

1.5. Information Source. Communication devices typically obtain their information from **Sensors** in the paired object (e.g., LumiTouch [9], the White Stones [26], and the Kiss Communicator [7]). **Databases and statistics** as available on the internet are another information source. Examples are the Data Fountain [5] displaying currency exchange values, or the Nabaztag [3] indicating the weather by light patterns. As said before, the single values that are displayed need not correspond to a single value that is measured. **Aggregators** can combine information in various ways, ranging from very straightforward to highly complicated information fusion using intelligent learning algorithms.

2 What Is the Intended Impact of Displaying the Information?

In the first place all devices create **awareness**: your partner thinks of you; your parents make tea (and therefore apparently are active [2]); or the device makes you aware of the CO₂ emission 10 years ago, compared to today. The consequence of awareness is an action or an emotion [11]. **Action.** Some information suggests an action, such as to water the dry plant, going to the coffee room when others are there, make a break, read your tweets, phone your parents. Measuring and displaying personal health status can motivate people to live healthier, as with the Fishtank [25]. **Emotion.** Other information mostly aims to trigger emotions. This often concerns relations between people – as with the various partner devices – but another emotion might be, e.g., *feeling rich or important*, through a personal stock market indicator, or the number of tweets received.

3 What Physical Object Is Used for the Single Value Device?

3.1 Is the information displayed through a dedicated object? Most examples use dedicated objects, already existing or created for this purpose. A few, however, use walls and surfaces [18] or dedicated screens [25] for display.

3.2 What Is the Modality Used for Displaying the Information? Any modality can be used (and: has been used) to represent the information in a single value device: light intensity [15] or pattern [18,3], sound [26], smell [23], motion [30], Bubbles in a tube [17], trembling [10,22], etc.

3.3 How personal Is the Object? Some of most evocative examples of single value devices are completely **Individual** objects. For example, the necklaces and Blossom [22] are pieces of art made for individual persons, by exploring what is meaningful to these persons and their relationships and transferring that into a very personal object. **Configurable** objects allow one, to some extent, to personalize a mass produced object. For the Nabaztag differently patterned ears could be chosen, and there was a great variety of costumes for Nabaztags. A few single value devices are based on **mass produced** consumer electronics gadgets, and their physical appearance is hardly configurable.

4 What Hardware Technology Is Used?

Single value devices use a wide range hardware technologies. **Actuators.** In our examples, (LED) lamps are used [9,18], dimmers [15], speakers [3], inflatable components for haptic sensations [24,21], motors [30], vibraton motors [10,22,3,16], heaters [23,26], also using bimetal [22], and pumps [5,17]. **Sensors** in the examples measure quantities such as location (GPS), displacement (accelerometers, distance sensors), presence of objects (RFID), sound intensity, light intensity, temperature, humidity, pressure (touch, air pressure, height), or time (DCF, GPS). **Information transport** is done through PCs, Arduinos or other microcontrollers, USB, WLAN, phones, wires, and web servers.

5 What Is Required for Using the Device in Daily Life?

When aiming at mature products, a number of pragmatic questions regarding actual use of the devices in daily life become important, too. **Setup and configuration.** How much technological expertise is required for setup and configuration (introduction to a network, coupling to a paired device, etc.)? Does the device require regular configuration updates? **Charging.** Does the device need batteries? Frequently having to recharge the device diminishes it's property of being a background service. **Services.** Which basic services, such as wireles internet or mobile phones, are used? Are there associated costs such as renewal of prepaid phone cards? Does the availability of the service depend on the availability of a company server?

4 Design Issues

Underlying single value devices is a fundamental tension. On the one hand, their most important characteristics center on being highly personal and context dependent objects. On the other hand, commercially feasible production requires very different design decisions, like uniformisation.

4.1 Single Value Devices Are Objects with Associated Emotions

Single value devices often represent information with personal meaning, and the object displaying the information should allow for emotional connotation. Consequently, there is the choice of either designing a new, dedicated object as carrier for this information, or taking an existing object with the emotional connotation already associated to it.

Attaching meaning to an object is an action of a person [11], it is not an inherent property of an object. Different people can attach different meanings to the same object. How to design an object that stimulates users to attach emotional meaning to it? In our examples we find two extreme approaches. The necklaces in [22] are designed personally for two people, taking their shared memories into account: e.g., elements from illustrations of fairy tales read together. The other extreme is to design a very neutral object and give space for projecting meaning to it. The white stones [26] and the CoConatch [1] have a tenuous design allowing for different connotations.

Another choice is to equip *existing* personal objects – that already carry emotional meaning for a user – with technology, as done with the tangible bits [18]. A very invasive strategy is followed with the internet enabled Furby [12]: after treatment it cannot be used anymore as a normal Furby. A more restrained approach is that of the Lumitouch photo frame [9] carrying a personal photo. To generalize the approach of non-invasive technology added to existing objects, we suggest to develop light-giving pedestals or small display cases in which one can put highly valued personal objects.

4.2 Customizability

Configurability is the answer of consumer electronics to individual needs when, at the same time, concentrating on cheap mass production. For single value devices it is crucial to keep them simple, Not all target groups will want to, or be able to, configure their device.

However, one key property of single value devices is that they are highly context dependent. Some people may want to see status information about global warming; others might prefer a single value device to display the severity of traffic jams, each user has his or her own loved ones whom (s)he wants to connect to. Functionalities may not be relevant in every context, e.g., the pollen status is mainly relevant in spring. From this point of view, single value devices need to be extensively configurable.

Aspects that need to be simple for each user are initialization, such as introducing to the local network, or pairing with another device.

Altogether, it is obvious that a range of devices is necessary to satisfy different needs. Configurability is a feature that should be included carefully balanced with simplicity. In order to meet requirements from producability we suggest a generic platform as discussed in the following section.

4.3 Building Blocks for Single Value Devices

In principle, it is not difficult to build prototypes of single value devices in all flavours. Still, from a practical point of view, the design of a prototype requires effort, knowledge and technological experience. For end users and their evaluation, prototypes easily suffer from lack of every-day convenience. Aspects like small size, simple chargability or connection to the internet, are typically not the first requirements in prototype development – but for usability these aspects are very important.

We suggest the development of a platform that can serve as a standard basic setup for single value devices. It should be easy to equip with a range of sensors and actuators, easily accessible for the software part, easily connect to internet, and aspects as charging sorted out. The advantage of such a platform would be: for the researching developer, who can efficiently develop prototypes, using a kind of universal building block, for the product developer, who can build on a flexible standard platform, for the end user, as basic usability of such a product is present, for the producer, who can produce such a platform in high numbers and use it for different products.

We have made a start with developing a standard basic setup for single value devices used in a *coffee rendez-vous application*, as an example application.

4.4 Service Dependence

Permanence of service availability is a topic that is often overlooked in the development of prototype single value devices, but which is very important when they are to become commercial products. We make things that depend for their whole life cycle on paid connectivity services such as telephony networks, radio datanetwork or internet. Additionally, many devices need a webserver for registration, configuration, storage of data, and dedicated applications. Quality of service, regular updates, and sufficient variation in the available applications is a crucial factor in the commercial success of a product.

5 Conclusion

Single value devices have a potential that is not yet realized. They can integrate into daily life in an unobtrusive and aesthetic way. The variety of applications is huge, from motivating to live healthier to reminding of everyday duties, from telepresence to simple playful communicators. But still, very few commercial products exist. Most of the examples available are prototypes exploring conceptual design choices. Aiming at commercial products a more integral view on single value devices is necessary, to which the work presented contributes.

We investigated the design choices by, first, exploring existing examples in a survey, and, in a second step, distilling the characteristics in a taxonomy of single value devices. The taxonomy in itself is already useful to identify unexplored areas in the design space. Furthermore, we contribute a critical discussion about how to design objects with emotional connotation, which is certainly underlying in many publications and prototype developments on single value devices. Further discussions address customizability, and the service concept that inherently gets introduced with single value devices. Our future efforts will aim at a more mature version of the hardware platform, and a variety of projects where we prove our platform and explore the possibilities of applications.

References

1. Coconatch, <http://www.coconatch.com>
2. Internet tea kettle, <http://www.mimamori.net/>
3. Nabaztag, <http://www.nabaztag.com>
4. Backlund, S., Gyllenswård, M., Gustafsson, A., Hjelm, S.I., Mazé, R., Redström, J.: STATIC! The aesthetics of energy in everyday things. In: Proc. Design Research Society Wonderground International Conference (2006)
5. Bakker, S., van den Hoven, E., Eggen, B.: Exploring Interactive Systems Using Peripheral Sounds. In: Nordahl, R., Serafin, S., Fontana, F., Brewster, S. (eds.) HAID 2010. LNCS, vol. 6306, pp. 55–64. Springer, Heidelberg (2010)
6. Bonn, B.: Scottie: Playful affective communication. In: Nijholt, A., Reidsma, D., Hondorp, H. (eds.) Proc. Intetain. Springer, Heidelberg (2009)
7. Buchenau, M., Suri, J.F.: Experience prototyping. In: Proc. 3rd Conf. on Designing Interactive Systems, pp. 424–433. ACM, New York (2000)

8. Buxton, W.: Integrating the periphery and context: A new taxonomy of telematics. In: Proc. Graphics Interface Conference, pp. 239–246. Morgan Kaufman (1995)
9. Chang, A., Resner, B., Koerner, B., Wang, X., Ishii, H.: LumiTouch: an emotional communication device. In: CHI 2001 Extended Abstracts on Human Factors in Computing Systems, pp. 313–314. ACM Press, New York (2001)
10. Cottam, M.: Wooden logic: In search of heirloom logics. Master's thesis, Umeå Institute of Design (2009)
11. Csikszentmihalyi, M., Rochberg-Halton, E.: The meaning of things: domestic symbols and the self. Cambridge University Press (1981)
12. Dertien, E.: Internet enabled furby, <http://hackaday.com/2009/08/31/internet-enabled-furby/>
13. A. Devices. Ambient orb, http://www.ambientdevices.com/cat/orb/MAN_AmbientOrb3-23-03.pdf
14. Fernando, O.N., Cheok, A.D., Merritt, T., Peiris, R.L., Fernando, C.L., Ranasinghe, N., Wickrama, I., Karunanayaka, K.: Babbage Cabbage: Biological Empathetic Media. In: VRIC Laval Virtual Proceedings, pp. 363–366 (April 2009)
15. Gellersen, H.-W., Schmidt, A., Beigl, M.: Ambient media for peripheral information display. *Personal and Ubiquitous Computing* 3, 199–208 (1999)
16. Greenberg, S., Kuzuoka, H.: Using digital but physical surrogates to mediate awareness, communication and privacy in media spaces. *Personal and Ubiquitous Computing* 3, 182–198 (1999)
17. Heiner, J.M., Hudson, S.E., Tanaka, K.: The information percolator: ambient information display in a decorative object. In: Proc. ACM Symposium on User Interface Software and Technology, pp. 141–148. ACM, New York (1999)
18. Ishii, H., Ullmer, B.: Tangible bits: towards seamless interfaces between people, bits and atoms. In: Proc. CHI 1997, pp. 234–241. ACM, New York (1997)
19. Köhler, B., David, S., Blumtritt, J.: Slow media manifest, <http://www.slow-media.net/manifest>
20. Roest, A., Claessen, S., Forbach, M., Pijls, B.: Google-battle: Ikwin, <http://www.mediamatic.net/page/52953>
21. Rosella, F., Genz, R.: Hug Shirt, <http://www.cutecircuit.com/products/thehugshirt/>
22. Seymour, S.: Social fabric: Jayne Wallace. In: Fashionable Technology, pp. 138–157. Springer, Vienna (2008)
23. Strong, A., Gaver, W.: Feather, Scent, and Shaker: Supporting simple intimacy. In: CSCW 1996 (1996)
24. Teh, J.K.S., Cheok, A.D., Choi, Y., Fernando, C.L., Peiris, R.L., Fernando, O.N.N.: Huggy pajama: a parent and child hugging communication system. In: Proc. IDC 2009, pp. 290–291. ACM Press, New York (2009)
25. Tellart: Humana fishtank, <http://www.tellart.com>
26. Tollmar, K., Junstrand, S., Torgny, O.: Virtually living together: A design framework for new communication media. In: Symposium on Designing Interactive Systems, pp. 83–91 (2000)
27. Tugui, A.: Calm technologies in a multimedia world. In: Ubiquity (2004)
28. Vazquez, J.L., Lopez-de-Ipina, D.: Social Devices: Autonomous Artifacts That Communicate on the Internet. In: Floerkemeier, C., Langheinrich, M., Fleisch, E., Mattern, F., Sarma, S.E. (eds.) IOT 2008. LNCS, vol. 4952, pp. 308–324. Springer, Heidelberg (2008)
29. Weiser, M.: The computer for the 21st-century. *Scientific American* 265(3), 94–104 (1991)
30. Weiser, M., Brown, J.S.: Designing calm technology. *PowerGrid Journal* 1 (1996)

A Kinect-Based Natural Interface for Quadrotor Control

Andrea Sanna, Fabrizio Lamberti, Gianluca Paravati, Eduardo Andres Henao
Ramirez, and Federico Manuri

Politecnico di Torino, Dipartimento di Automatica e Informatica,
C.so Duca degli Abruzzi 24, I-10129 Torino, Italy
{andrea.sanna,fabrizio.lamberti,gianluca.paravati,
eduardo.henaoramirez,federico.manuri}@polito.it

Abstract. The evolution of input device technologies led to identification of the natural user interface (NUI) as the clear evolution of the human-machine interaction, following the shift from command-line interfaces (CLI) to graphical user interfaces (GUI). The design of user interfaces requires a careful mapping of complex user “actions” in order to make the human-computer interaction (HCI) more intuitive, usable, and receptive to the user’s needs: in other words, more user-friendly and, why not, fun. NUIs constitute a direct expression of mental concepts and the naturalness and variety of gestures, compared with traditional interaction paradigms, can offer unique opportunities also for new and attracting forms of human-machine interaction. In this paper, a kinect-based NUI is presented; in particular, the proposed NUI is used to control the Ar.Drone quadrotor.

Keywords: Natural User Interface, Kinect, Quadrotor control, Interactive systems.

1 Introduction

Gestures are important factors in conversations between humans. Researchers have designed and implemented several human-computer interaction (HCI) paradigms based on gestures, thus creating the so called Natural User Interfaces (NUIs). NUIs have been investigated since early eighty’s (voice and gestures are used to control a GUI in [3]). Among NUIs, gesture-based interfaces always played a crucial role in human-machine communication, as they constitute a direct expression of mental concepts [16]. The naturalness and variety of hand and body gestures, compared with traditional interaction paradigms, can offer unique opportunities also for new and attracting forms of HCI [15]. Thus, new gesture-based solutions have been progressively introduced in various interaction scenarios (encompassing, for instance, navigation of virtual worlds, browsing of multimedia contents, management of immersive applications, etc. [20] [28]) and the design of gesture-based systems will play an important role in the future trends of the HCI.

Human-Robot Interaction (HRI) is a subset of the HCI and can be considered as one of the most important Computer Vision domains. In HRI-based systems,

especially in safe critical applications such as search-and-rescue and military, it is increasingly necessary for humans to be able to communicate and control robots in a natural and efficient way. In the past, robots were controlled by human operators using hand-controllers such as sensor gloves and electromechanical devices [23]. These devices limit the speed and simplicity of the interaction. To overcome the limitations of such electro-mechanical devices, vision based techniques [16] have been introduced. Vision based techniques do not require wearing of any contact devices, but use a set of sensors and computer vision techniques for recognizing gestures. Therefore, the type of communication based on gestures can provide an expressive, natural and intuitive way for humans to control robotic systems. One benefit of such a system is that it proposes natural ways to send geometrical information to the robot, such as: up, down, etc. As seen in [2], through the recognition of gestures, a natural language for human-machine interaction can be created, relying on non invasive methods such as a camera, to identify user gestures for comparison with a predefined gesture database. Gestures may represent a single command, a sequence of commands, a single word, or a phrase and may be static or dynamic. Such a system should be accurate enough to provide the correct classification of gestures in a reasonable time.

Although a lot of works of HRI by gestures are known in the literature (Section 2 briefly reviews the most appropriate) recent technological advances have opened new and challenging research horizons. In particular, controllers and sensors used for home entertainment can be affordable devices to design and implement new HRI forms.

The aim of this work is to create a human-robot interaction framework based on the use of body gestures. To achieve this, the main requisites are to extract spatial information from specific parts of the body and secondly to extract gestures from this information. In this work, Microsoft Kinect [12] is used as gesture tracking device; recognized gestures are then used to control the Ar.Drone quadrotor platform [1] (in the following of the paper the terms: Ar.Drone, quadrotor, and platform will be used interchangeably). The user is the “controller” and so a new form of HRI can be experienced. Tests proved that the platform can be easily controlled by a customizable set of body movements, consequently allowing for an exciting, fun, and safe experience even for non-skilled users.

The paper is organized as follows: Section 2 reviews the main HRI solutions and briefly introduces the Ar.Drone. Section 3 describes the system architecture and the mapping between gestures and commands. Finally, remarks about this experience and future investigation trends are presented in Section 4.

2 Background

The ability to recognize gestures is important for an interface developed to understand users intentions. Interfaces for robot control that use gesture recognition have deeply been studied as using gestures provides a formidable challenge. Several issues arise from environments with complex backgrounds, from dynamic

lighting conditions, from shapes to be recognized (in general, hands and the other parts of the human body can be considered as deformable objects), from real-time execution constraints, and so on.

A lot of work has been focused on hand gesture recognition for human robot interaction. For instance, an architecture of hand gesture-based control of mobile robots was proposed in [21]. The gestures were captured by a data glove and gesture recognition was done by Hidden Markov Model statistical classifiers. The interpreted gestures were translated into commands to control the robot. Later on the use of a data glove was replaced by the use of markers in [9]. Two cameras provided the info to triangulate the position of the hand markers, allowing gesture recognition to take place and control a 6DOF robot with a high precision. An alternative identification of the hand posture was also proposed in [5]. The hand posture is identified from the temporal sequence segmented obtained by the Hausdorff distance method. A real time vision based gesture recognition system for robot control was implemented in [2]. Gestures were recognized using rule based approach by comparing the skin like regions in a particular image frame with the predefined templates in the memory of the system. Another hand gesture recognition system for robot control, which uses Fuzzy-C-Means algorithm as gesture classifier to recognize static gestures, was proposed in [25] and [26]. Static and dynamic gestures are recognized by a Fuzzy-C-Means clustering algorithm in [19].

YCbCr segmentation to recognize hand gestures has been proposed in [22], whereas a real-time hand posture recognition using 3D range data analysis is presented in [10]. A background subtraction approach using video sequences is proposed in [18], whereas motion detection algorithms for gesture recognition are used in [11]. A trajectory-based hand gesture recognition, which uses kernel density estimation and the related mean shift algorithm, was presented in [17]. A method for detecting and segmenting foreground moving objects in complex scenes using clusters is used in [4]. Under the assumption that the target object occupies the entire image, the humans body proportions are considered and using (vertical and horizontal) histogram analysis the hand gesture is recognized by a webcam in [8].

In this paper, a novel method of interaction and control of quadrotors by whole body movement recognition is described. Microsoft Kinect allows users to experience a new type of HRI able to provide an intuitive, robust and fun interaction form.

2.1 Quadrotors and the Platform Ar.Drone

Quadrotors are used in a large spectrum of applications ranging from surveillance to environmental mapping. Quadrotors are used singularly as well as in swarm; in this last case, the task of coordination is always a critical issue. Quadrotors can be used both outdoor and indoor; outdoor platforms use, in general, autopilots for autonomous navigation whereas several localization techniques (mainly based on computer vision) are exploited to determine position and orientation of indoor platforms.

The human interface plays a key role when a quadrotor and, in general any flying platform, has to be directly controlled by the user. RC-transmitters and joysticks are the two most common input devices used to control quadrotors. Innovative solutions uses multitouch devices (e.g., the iPhone [1] and Microsoft Surface [24]) and game controllers (e.g., Nintendo Wiimote [27]). Initial attempts of Microsoft Kinect usage to control the Ar.Drone have been proposed in [31] and [32]. In both cases, hand gestures are translated in commands for the platforms.

The Parrot AR.Drone [1] is a quadrotor helicopter with Wi-fi link and two cameras: a wide angle front camera and a high speed vertical camera. Software clients to control the platform are available: Windows/Linux PC clients and an application for iPhone can be used to control the Ar.Drone by keyboard, joystick or a multitouch device. The Parrot AR.Drone provides automatic “procedures” for takeoff, landing, and hovering. A public SDK is available to implement custom applications for the quadrotor control; the Windows client has been used as the starting point to develop the proposed solution (see Section 3). The SDK can be used to connect to the AR.Drone ad-hoc Wi-fi network, send commands (takeoff, land, up/down, rotate, and so on), receive, decode and display live video stream from the two cameras, receive and interpret navigation data and battery status. Although the Ar.Drone is sold in Europe to a price of about 300 euros as *the flying video game*, an impressive number of users use this platform for technical and research purposes.



Fig. 1. A high-level description of the system

3 System Architecture

A high-level description of the system is provided in Fig. 1. The user’s body is tracked by the Microsoft Kinect [12], that is connect to a PC via USB; gestures (body poses) are translated in commands to be sent to the platform via Wi-Fi connection. The user will be able to completely control the quadrotor movements by using the body as a sort of natural controller; moreover, an ad-hoc developed GUI (Graphics User Interface) enables the user to remotely control the platform as flight attitudes, navigation data (telemetry), and the video stream from the onboard cameras are shown, thus releasing the user to directly see the quadrotor.

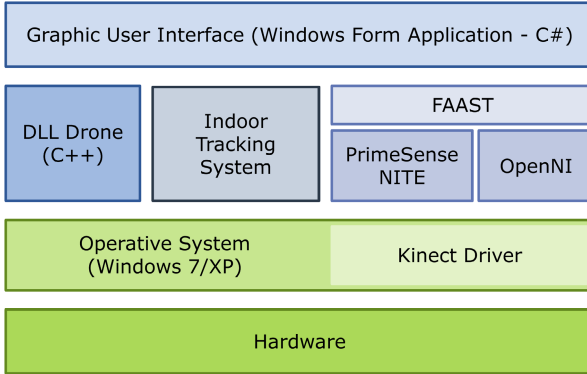


Fig. 2. Layers of the software architecture

From the software point of view, the architecture is shown in Fig. 2. The stack composed by FAAST (Flexible Action and Articulated Skeleton Toolkit [7]), OpenNI - PrimeSense Nite, and the Kinect drivers is used to capture and decode body poses. FAAST is a middleware to facilitate integration of full-body control with games and VR applications using OpenNI-compliant depth sensors (e.g., Microsoft Kinect). The toolkit incorporates a custom VRPN (Virtual-Reality Peripheral Network [29]) server to stream the user's skeleton over a network, allowing VR applications to read the skeletal joints as trackers using any VRPN client. FAAST can also emulate keyboard input triggered by body posture and specific gestures.

On the other hand, the OpenNI Framework [14] provides the interface for physical devices and for middleware components. APIs enable modules to be registered in the OpenNI framework and to be used to produce sensory data. OpenNI covers communication with both low level devices (e.g., Microsoft Kinect), as well as high-level middleware solutions (e.g., FAAST). OpenNI can interact with the Microsoft Kinect by the OpenKinect library [13]. Body poses detected by FAAST are used by the GUI to trigger a modified version of the keyboard-based Ar.Drone client (the DLL Drone module in Fig. 2), thus implementing an effective and robust command chain to control the platform. Moreover, the GUI has been designed to receive information about position and orientation of the platform from an optical tracking system. Information coming from the optical tracker (the affordable system proposed in [6] has been used for tests) allow to implement mechanisms of AI (Artificial Intelligence) to control the quadrotor, thus replacing the user.

Fig. 3 shows the exchanged data among system components. The Ar.Drone sends the GUI navigation data and the video stream, whereas it receives navigation commands. Each command is the *translation* of a body pose according to Table 1. This table is used by FAAST to trigger a set of keyboard events related to platform commands. Moreover, each pose (also called action) is associated with a threshold; for instance the syntax: `lean_forward 15` sets a lean forward of at least 15 degrees to activate the corresponding action. The thresholds define the

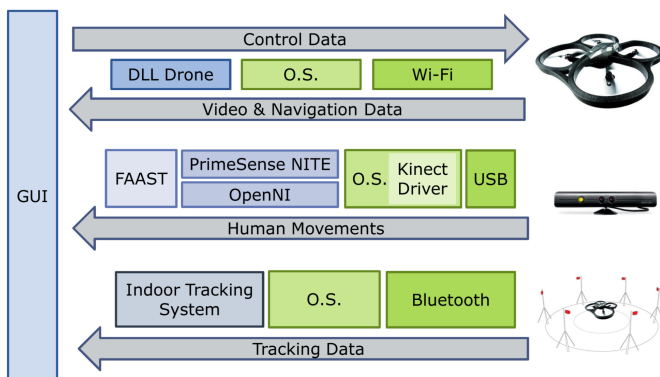


Fig. 3. Exchanged data among system components

Table 1. Correspondence between body poses and commands for the quadrotor

Body pose	Ar.Drone command
Right arm up	Takeoff
Right arm down	Landing
Lean forward	Go forward
Lean backward	Go backward
Lean right	Go right
Lean left	Go left
Left arm up	Go up
Left arm down	Go down
Left arm out	Turn left
Right arm out	Turn right
Rest position	Hovering

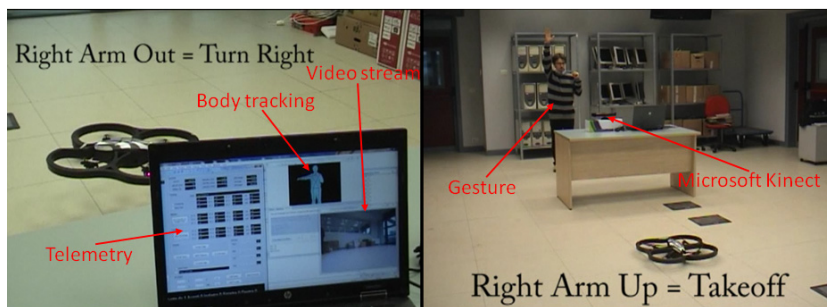


Fig. 4. Two pictures of the experimental setup. On the left the laptop console is shown, whereas the Microsoft Kinect is visible on the right.

sensibility in recognizing body poses and they can be thought as the joystick dead-zone, that is the region of movements which are not recognized by the device.

The user can customize the association between action and platform commands, thus choosing the body poses more intuitive and effective. Threshold values of 20-25 have been experienced as a good tradeoff between robustness (i.e., the system really detects the right pose) and sensibility (i.e., the size of the “deadzone”). A video showing an example of Ar.Drone control by body movements can be found in [30]. The video allows to appreciate both intuitiveness of the HRI and the graphics output the user can use to control the platform.

4 Conclusion and Remarks

This paper presents an example of NUI based on body gestures/movements to control a quadrotor. Although this work shows a challenging and exciting scenario, a more accurate and rigorous study is necessary to evaluate the efficiency of this kind of solution.

A comparative analysis involving different human machine interfaces is scheduled. The methodology that will be adopted for evaluation plans to propose a set of tests to be performed by a group of users. In particular, the tests will consist in repeating one or more navigation tasks by using a joystick, a multitouch device, and the proposed solution. The user will be asked to perform a complete flight session from takeoff to landing. The users will be trained on the execution of each test. The trainer will illustrate the features of the different interfaces. Thereafter, each user will be allowed to experience the basic flight commands with the different interfaces. The mission will consist in lift-off the quadrotor from a specific location, reach one or more checkpoints in the environment by performing a number of actions to change the flight attitudes, and finally try to drive the quadrotor to a well defined landing location to set down the platform. Different commands are involved in this test and it can represent a valid testbed to compare the different user interfaces. The results will be gathered in objective terms regarding the time needed to complete the task (from takeoff to landing). Moreover, subjective evaluations will be considered. Indeed, at the end of the set of tests each user will be asked to fill in a questionnaire about the usability of the proposed interface, including questions related to the perceived robustness of the system (e.g. to take into account the errors of classification of a posture, number of gesture repetitions due to the misclassified postures, and so on). The setup of the testbed could also cover precision evaluations by using the infrared-optical tracking system proposed in [6]. This system will be used to measure the position of the Ar.Drone in the environment. Indeed, by performing the flight tasks within the infrared-optical tracking system it is possible to integrate and cross-relate the previously described results (in terms of completion time) with a measure of the precision of the performed actions, e.g. it is possible to measure the distance between the “target” landing point (i.e. a well defined position where the user is asked to set down the platform) and the “real” landing point (i.e. the position where the platform is actually landed by using a specific human machine interface).

At this moment, the whole latency of the system has been measured: the term latency denotes, in this case, the delay between a user’s movement and the

execution of the corresponding command. The measure has been performed by analyzing the video sequence in [30] and counting the number of frames elapse between user and Ar.Drone movements. An average latency of 0.3 seconds has been experienced. Thus, about three commands can be executed in a second, that is fully consistent both with the platform's dynamic and the "user's dynamic".

Affordable devices such as Microsoft Kinect are opening new scenarios allowing to create innovative forms of HRI unthinkable until a few months ago. The evolution of devices designed to implement novel user centric forms of entertainment provides researchers alternative tools to re-design more intuitive, robust, and fun HRI paradigms.

References

1. Ar.Drone, <http://ardrone.parrot.com>
2. Bhuiyan, M.A., Ampornaramveth, V., Muto, S., Ueno, H.: Realtime vision based Gesture Recognition for Human Robot Interaction. In: The IEEE International Conference on Robotics and Biometrics, pp. 413–418 (2004)
3. Bolt, R.A.: Put-that-there: Voice and gesture at the graphics interface. *ACM Comput. Graphics* 14(3), 262–270 (1980)
4. Bugeau, A., Pérez, P.: Detection and segmentation of moving objects in complex scenes. *Computer Vision and Image Understanding* 113(4), 459–476 (2009)
5. Chao, N., Meng, M.Q., Xiaoping Liu, P., Wmg, X.: Visual gesutre recognition for human-machine interface of robot teleoperation. In: The 2003 IEEEIRSIJ Intl. Conference on Intelligent Robots and Systems, pp. 1560–1565 (2003)
6. De Amici, S., Sanna, A., Lamberti, F., Pralio, B.: A Wii Remote-based infrared-optical tracking system. *Entertainment Computing* 1(3-4), 119–124 (2010)
7. FFAST, <http://projects.ict.usc.edu/mxr/faast/>
8. Koceski, S., Koceska, N.: Vision-based gesture recognition for human-computer interaction and mobile robot's freight ramp control. In: The 32nd International Conference on Information Technology Interfaces, pp. 289–294 (2010)
9. Kofman, J., Wu, X., Luu, T.J., Verma, S.: Teleoperation of a robot manipulator using a vision-based human-robot interface. *IEEE Transactions on Industrial Electronics* 52(5), 1206–1219 (2005)
10. Malassiotis, S., Aifanti, N., Strintzis, M.G.: A gesture recognition system using 3D data. In: The 1st International Symposium on 3D Data Processing Visualization and Transmission, pp. 190–193 (2002)
11. Mariappan, R.: Video Gesture Recognition System. In: The International Conference on Computational Intelligence and Multimedia Applications, pp. 519–521 (2007)
12. Microsoft Kinect, <http://www.xbox.com/en-US/kinect/>
13. OpenKinect, http://openkinect.org/wiki/Main_Page
14. OpenNI, <http://www.openni.org/>
15. Pavlovic, V.I., Sharma, R., Huang, T.S.: Gestural interface to a visual computing environment for molecular biologists. In: 2nd Intern. Conf. on Automatic Face and Gesture Recognition, pp. 52–73. IEEE Computer Society, Los Alamitos (1996)
16. Pavlovic, V.I., Sharma, R., Huang, T.S.: Visual interpretation of hand gestures for human-computer interaction: a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19, 677–695 (1997)

17. Popa, D., Simion, G., Gui, V., Otesteanu, M.: Real time trajectory based hand gesture recognition. *WSEAS Transactions on Information Science and Applications* 5(4), 532–546 (2008)
18. Ribeiro, H.L., Gonzaga, A.: Hand Image Segmentation in Video Sequence by GMM: a comparative analysis. In: *The 19th Brazilian Symposium on Computer Graphics and Image Processing*, pp. 357–364 (2006)
19. Rao, V.S., Mahanta, C.: Gesture Based Robot Control. In: *The 4th International Conference on Intelligent Sensing and Information Processing*, pp. 145–148 (2006)
20. Selker, T.: Touching the future. *Commun. ACM* 51, 14–16 (2008)
21. Soshi Iba, C.J.P., Michael Vande Weghe, J., Khosla, P.K.: An Architecture for Gesture-based Control of Mobile Robots. In: *The IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 2, pp. 851–857 (1999)
22. Stergiopoulou, E., Papamarkos, N.: A New Technique for Hand Gesture Recognition. In: *The IEEE International Conference on Image Processing*, pp. 2657–2660 (2006)
23. Sturman, D.J., Zetler, D.: A survey of glove based input. *IEEE Computer Graphics and Applications* 14, 30–39 (1994)
24. Controlling the AR Drone with Microsoft Surface, <http://blogs.msdn.com/b/surface/archive/2011/01/27/controlling-the-ar-drone-with-surface.aspx>
25. Wachs, J.P., Stern, H., Eden, Y.: Parameter search for an image processing-Fuzzy c-Means hand gesture recognition system. In: *The IEEE Int. Conf. on Image Processing*, pp. 341–344 (2003)
26. Wachs, J.P., Stern, H., Edan, Y.: Cluster Labeling and Parameter Estimation for the Automated setup of a Hand gesture Recognition System. *IEEE Transactions Systems and Humans* 35(6), 932–944 (2005)
27. Controlling the AR Drone with Nintendo Wiimote, http://www.youtube.com/watch?v=zJ50H-_431w&feature=player_embedded
28. Wright, A.: Making sense of sensors. *Commun. ACM* 52, 14–15 (2009)
29. The Virtual-Reality Peripheral Network, <http://www.cs.unc.edu/Research/vrpn/>
30. Controlling the AR Drone with Microsoft Kinect, <http://www.youtube.com/watch?v=jDJpb4xXAJM>
31. Hand tracking to control the AR Drone with Microsoft Kinect, <http://dronehacks.com/2010/12/21/controlling-the-ar-drone-with-a-kinect-controller/>
32. Hand tracking to control the AR Drone with Microsoft Kinect, <http://www.youtube.com/watch?v=mREorv0hbY8>

Smart Material Interfaces: A Vision

Andrea Minuto¹, Dhaval Vyas^{1,2}, Wim Poelman², and Anton Nijholt¹

¹ Human Media Interaction, University of Twente, The Netherlands
a.minuto@utwente.nl

² Design, Production and Management, University of Twente, The Netherlands
d.m.vyas@utwente.nl

Abstract. In this paper, we introduce a vision called Smart Material Interfaces (SMIs), which takes advantage of the latest generation of engineered materials that has a special property defined “smart”. They are capable of changing their physical properties, such as shape, size and color, and can be controlled by using certain stimuli (light, potential difference, temperature and so on). We describe SMIs in relation to Tangible User Interfaces (TUIs) to convey the usefulness and a better understanding of SMIs.

Keywords: Tangible User Interfaces, Ubiquitous Computing, Smart Material Interfaces.

1 Introduction

Although the tangible representation allows the physical embodiment to be directly coupled to digital information, it has limited ability to represent change in many material or physical properties. Unlike malleable pixels on the computer screen, it is very hard to change a physical object in its form, position, or properties (e.g. color, size) in real time.

– Hiroshi Ishii [5]

Mark Weiser’s [18] vision of Ubiquitous Computing motivated researchers to augment everyday objects and environments with computing capabilities to provide reality-based [7] and more natural interaction possibilities. One of the most promising sub visions has been the tangible user interfaces (TUIs) [6]. In TUIs it is proposed to use physical handles to manipulate digital information. Some of the known examples of TUIs are Urp [16], actuated workbench [11], Illuminating Light [15], MediaBlocks [14], Siftables [8] and SandScape [4]. One of the major limitations of TUIs is that they focus more on the input mechanism and less on the output. As Ishii [5] explained, the incapability of making changes in the physical and material properties of output modalities is a major limitation of TUIs. Building on this limitation of TUIs [5], we propose a sub vision entitled Smart Material Interfaces (SMIs). The main focus of a SMI is being able to make changes in the physical and material properties of output modalities. SMI proposes the use of materials that have inherent or “self augmented” capabilities of changing physical properties such as color, shape and texture, under the

control of some external stimulus such as electricity, magnetism, light, pressure and temperature.

The purpose of this paper is to draw attention to this upcoming field of research. In this paper, we describe SMIs in relation to TUIs to convey the usefulness and a better understanding of SMIs. We first describe our motivation behind this work. Next, we describe the vision of SMI with reference to TUIs. In the end we provide future directions for SMIs.

1.1 Motivation

There are three main motivations for introducing such a vision, in the ever growing field of ubiquitous computing.

First, we believe that there is a need to make the vision of ubicomp, as conceived by Mark Weiser [18], more relevant. We see a trade-off in the ways this vision is applied in the current research. The central idea behind the vision of ubicomp is to seamlessly embed computing in the everyday used objects, both socially and procedurally. The material qualities of these everyday objects play a big role in the social and procedural practices of people. In the current ubicomp research, the material and the computation are seen detached from each other [17]. As Buechley and Coelho [2] suggest, electronic components are seldom integrated into objects' intrinsic structure or form. We believe that there is a need to highlight the blurring boundaries between the material qualities of an object and the computational functionalities it is supposed to support.

Second, the technology push from different fields of material sciences has provided new possibilities to integrate materials such as metals, ceramics, polymeric and biomaterials and other composite materials for designing products. A wide range of smart materials can be seen in the literature that can change their shape, size, color and other properties based on external stimuli. These properties of smart materials can be used to create new kind of interaction and interfaces. In section 2.1, we will provide a few examples of these materials.

Third, with the use of smart materials, as designers, we can introduce a new communication 'language' to users. Use of screen-based interfaces has dominated the user interfaces for several years now. These use icons, texts, and other types of widgets to support communication with users. Smart materials can introduce new semantics to the human-computer interaction, which focuses on change of shapes, colors, size or positioning. Of course, the potential and semantic value of such a type of communication have to be explored and experimented further. However, the use of smart materials can be seen as a radical shift in the way we see our user interfaces.

2 The Vision

The basic idea behind the SMIs vision is that it attempts to sensibly utilizes readily-available, engineered materials as physical properties of an interface to convey information to its users. Additionally, following the ubicomp vision, SMIs attempts to close the gap between the computation and the physical medium

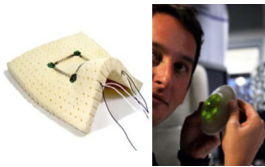
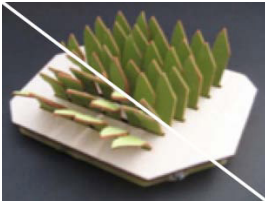

– where the physical medium itself is capable of making changes. Computation and other external stimuli could help in this but it is not a necessity. This way our everyday used objects can convey informations by means of their physical properties and use the material itself as a medium of physical representation. SMIs emphasis on the medium used for the interaction, the object itself, instead of having a simulacrum giving the idea of interaction of another object augmented as input system.

To make the SMIs vision clearer, we will first provide a brief overview of the type of smart materials that are currently available and how they are used in designing interfaces and products. Next, we will provide an informal comparison of SMIs with TUIs – that have been around for some time.

2.1 Smart Materials

Before going further, we would like to explain what we mean by “smart materials”. A smart material has at least one or more properties that can be dynamically altered in certain conditions that can be controlled from outside (external stimuli). Each individual type of smart material has specific properties which can be altered, such as shape, volume, color and conductivity. These properties can influence the types of applications the smart material can be used for. The most common smart materials can be in the form of polymers, ceramics, memory metals or hydro gels. These materials are engineered within the fields of

Table 1. Examples of existing smart material interfaces

Concept	Description	Material
	<p><i>SpeakCup</i> is a voice recorder in the form of a soft silicone disk with embedded sensors and actuators, which can acquire different functionalities when physically deformed by a user [3].</p>	<p>Composition of disk of platinum cure silicone rubber (passive shape memory)</p>
	<p><i>Sprout I/O</i> is a textural interface for tactile and visual communication composed of an array of soft and kinetic textile strands, which can sense touch and move to display images and animations. [3]</p>	<p>Shape memory alloy used as electrode for capacitive sensing and actuation soft mechanism.</p>
	<p>Concept that displays different information about safety and risk relative to the temperature of the content of the bottle. Designers: Hung Cheng, Tzu-Yu Huang, Tzu-Wei Wang and Yu-Wei Xiang</p>	<p>Thermochromic liquid crystals</p>

chemistry, polymer sciences and nano technology. Importantly, these fields can offer specific kind of smart material that can be operated using specific external stimuli. For example, polymers can be activated through light, magnetism, thermally or electrically. Other smart materials: NiTinol [10] (memory shape alloy, used for internal surgery); phase change materials [13] (heat is absorbed or released when the material changes state, used for mugs and clothes); chromogenic material [1] (changes color in response to electrical, optical or thermal changes, used in sunglasses and lcd); ferrofluid liquid [12] (becomes strongly magnetized in the presence of a magnetic field, used for Hard Disk and Magnetic resonance).

In the Table 1, we provide some examples of interfaces built using smart materials.

2.2 SMI vs TUI

Figure 1 shows an architectural comparison between SMIs and TUIs and Table 2 summarizes their differences.

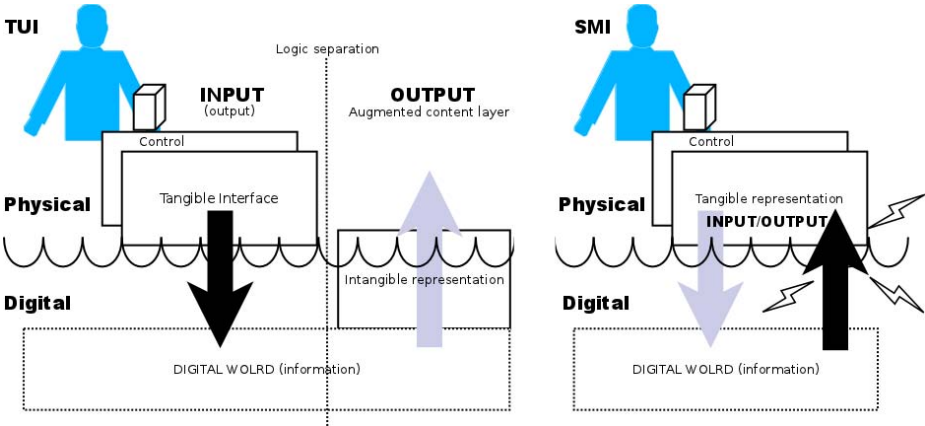


Fig. 1. Making a comparison between TUI (left) and SMI (right), we want to stress the tight coupling of information and tangible interface, and especially the use of tangible elements as output of the system. This will take advantage of the smart properties that can be carried by the object itself as interface. The black arrow emphasize the focus of interest for the interface (as input in TUI, as output in SMI).

TUI. As mentioned in [5], “the tangible representation allows the physical embodiment to be directly coupled to digital information”, but the “limited ability to represent change in many material or physical properties” has been a drawback. As can be seen in figure 1 (left side), the user interacts with a tangible form of information (the object itself) to control the underneath mechanism – the object translates movements into commands and data in a digital form for the system (digital world). Once the computation has been done a different output is prompted to the user. The information returned (augmented content

Table 2. Advantages of *SMI* in comparison with *TUI*

TUI vs SMI	
sometimes incoherent in the relevant ambience (physical - digital)	coherent space of information (physical - physical)
information is represented as an augmented overlay on the object	information is part of the material/object itself
tends to separate input and output (distinction by physical - digital)	promotes a more tight coupling input/output
users can feel the difference from the “real” and augmented information	information added in a completely transparent way
output is felt non-continuous non-persistent	output physically present (not a digital representation), continuous and persistent
balances coupling the tangible and intangible representations	uses physicality of the object as way to deliver information
makes use of electronics and controllers	uses properties of smart materials

layer) can be presented over the tangible interface itself. The user can interact with the augmented layer by moving the physical interface. In TUI, we need to balance the intangible digital information (inside the augmented content layer) and the tangible representation (represented by the object itself) in such a way as to create a perceptual coupling between the physical and digital [5].

SMI. With the use of smart materials, SMI attempts to overcome the limitation of TUI. SMI focuses on changing the physical reality around the user as the output of interaction and/or computation as well as being used as input device. SMI promotes a much tighter coupling between the information layer and the display by using the tangible interface as the control and display at the same time – embedding the augmented information layer directly inside the physical object. It uses the physicality of the object as a way to deliver information. Utilizing smart materials’ properties, SMI can support cohesive interaction by maintaining both channels (input and output) on the same object of interaction. The interaction constructed in this way will grant the user a continuous perception of the object and of the output with a persistent physicality coherent with the space.

3 Conclusions: Applications and Future Possibilities

We believe that SMIs could have a wide range of applications, not limited to the field of computing. In fact, literature has shown how smart materials are used in surgery [9], architecture, art and engineering. SMIs do not need any kind of display, with materials being both the interface and input-output stimuli. Their physical characteristics may be enough to carry and convey information. In this way, SMIs propose a radical change in the way we see and understand common

user interfaces as well as the way we interact with things, introducing a new space for research and development. We believe that in the future we will have a more seamless interaction between the real world and the digital world. This will provide a new meaning to augmented reality interaction that will have a more continuous, persistent and coherent feedback in relevant contexts.

References

1. Barry, I.G.: The commercialization of plastic photochromic lenses : A tribute to john crano. *Molecular Crystals and Liquid Crystals*, 57–62 (2000)
2. Buechley, L., Coelho, M.: Special issue on material computing. *Personal Ubiquitous Comput.* 15, 113–114 (2011)
3. Coelho, M., Zigelbaum, J.: Shape-changing interfaces. *Personal Ubiquitous Comput.* 15, 161–173 (2011)
4. Ishii, H., Ratti, C., Piper, B., Wang, Y., Biderman, A., Ben-Joseph, E.: Bringing clay and sand into digital design - continuous tangible user interfaces. *BT Technology Journal* 22, 287–299 (2004)
5. Ishii, H.: Tangible bits: beyond pixels. In: *Proc. of TEI 2008*, pp. xv–xxv. ACM, New York (2008)
6. Ishii, H., Ullmer, B.: Tangible bits: Towards seamless interfaces between people, bits and atoms. In: *CHI 1997*, pp. 234–241 (1997)
7. Jacob, R.J., et al.: Reality-based interaction: a framework for post-wimp interfaces. In: *Proc. of CHI 2008*, pp. 201–210. ACM, New York (2008)
8. Merrill, D., Kalanithi, J., Maes, P.: Siftables: towards sensor network user interfaces. In: *Proc. of the 1st International Conf. on TEI 2007*, pp. 75–78. ACM, New York (2007)
9. Morgan, N.B.: Medical shape memory alloy applications—the market and its products. *Materials Science and Engineering A* 378(1-2), 16–23 (2004), European Symposium on Martensitic Transformation and Shape-Memory
10. Nitinoldraht, http://www.nitinoldraht.de/shop_content.php?language=en
11. Pangaro, G., Maynes-Aminzade, D., Ishii, H.: The actuated workbench: computer-controlled actuation in tabletop tangible interfaces. In: *Proc. of the 15th Annual ACM Symp. on UIST 2002*, pp. 181–190. ACM, New York (2002)
12. Scherer, C., Figueiredo Neto, A.M.: Ferrofluids: properties and applications. *Brazilian Journal of Physics* 35, 718–727 (2005)
13. Simone, R.: Phase change materials. *Annual review of materials research* (2009)
14. Ullmer, B., Ishii, H.: Mediablocks: tangible interfaces for online media. In: *CHI 1999 Extended Abstracts on Human Factors in Computing Systems 1999*, pp. 31–32. ACM, New York (1999)
15. Underkoffler, J., Ishii, H.: Illuminating light: a casual optics workbench. In: *CHI 1999 Extended Abstracts on Human Factors in Computing Systems*, pp. 5–6. ACM, New York (1999)
16. Underkoffler, J., Ishii, H.: Urp: a luminous-tangible workbench for urban planning and design. In: *Proc. of CHI 1999*, pp. 386–393. ACM, New York (1999)
17. Vallgård, A., Redström, J.: Computational composites. In: *Proc. of CHI 2007*, pp. 513–522. ACM, New York (2007)
18. Weiser, M.: The computer for the 21st century. *Scientific American* 265(3), 94–104 (1991)

User-Centered Evaluation of the Virtual Binocular Interface

Donald Glowinski¹, Maurizio Mancini¹, Paolo Coletta¹, Simone Ghisio¹,
Carlo Chiorri², Antonio Camurri¹, and Gualtiero Volpe¹

¹ InfoMus Lab, DIST, University of Genova, Italy
{donald,maurizio,ghisio}@infomus.org,
{paolo.coletta,antonio.camurri,gualtiero.volpe}@unige.it

² DiSA - Department of Anthropological Sciences
Psychology Unit, University of Genova, Italy
carlo.chiorri@unige.it

Abstract. This paper describes a full-body pointing interface based on the mimicking of the use of binoculars, the Virtual Binocular Interface. This interface is a component of the interactive installation “Viaggiatori di Sguardo”, located at Palazzo Ducale, Genova, Italy, and visited by more than 5,000 visitors so far. This paper focuses on the evaluation of such an interface.

Keywords: Virtual Reality, Interactive Museum Applications and Guides, Novel Interaction Technologies.

1 Introduction

In this paper we propose the evaluation of a full-body pointing interface: the *Virtual Binocular Interface*. The permanent interactive museum installation “Viaggiatori di Sguardo”, opened on December 2009 at Palazzo Ducale, Genova, Italy, allows visitors to explore and interact with an audiovisual content by mimicking the use of binoculars. Via this interface, users can start a virtual journey to discover the monumental buildings “Palazzi dei Rolli” (UNESCO Treasure) of the Italian city of Genova.

In such context, the role of the user can resemble that of an *explorer* or a *traveller*, *viaggiatore* in Italian. The metaphors of *journey*, *exploration* and *travelling* led us to conceive, design, and implement the Virtual Binocular Interface to explore, in an ecological way, audiovisual content.

2 “Viaggiatori di Sguardo”: Overview and Background

The installation “Viaggiatori di sguardo”, designed and developed by Casa Paganini - InfoMus Research Centre, University of Genova, in collaboration with Palazzo Ducale, Fondazione per la Cultura of Genova, is a sensitive environment located in a room available to the public of tourists and visitors of Palazzo

Ducale in Genova, Italy. The goal of the installation is to present to visitors the UNESCO treasure of “Palazzi dei Rolli”, a number of magnificent monumental buildings located in the city.



Fig. 1. The “Viaggiatori di Sguardo” permanent installation at Palazzo Ducale

Figure 1 shows the installation paradigm: a large screen (7 meters) that can be explored by several users by via the *Virtual Binocular Interface*. Below the large screen, smaller LCD screens are installed to provide brief textual information on the artwork that can be explored. The user, being a traveler on such cultural heritage, has at her disposal the Virtual Binocular: when she mimics with both hands the gesture of raising and pointing a binocular, she is enabled to zoom in the available cultural heritage audiovisual content, as shown on the left side of Figure 2.

Such gesture is functional and ecological in the context of the exploration of unknown places, since the user feels at her ease in such interaction. In several other cases of full body interfaces designed for interaction with content, it has been reported by users a varying extent of shame or embarrassment, mostly due to the need to perform unnatural movements in presence of others [9]. The proposed interface may represent a solution to this and other problems.

On the right side of Figure 2 one of the mechanisms available to navigate the monumental buildings is shown: in some spots of the image explored by the user, the contour of the binocular shape projected on the screen changes colour: this means that if the user continues pointing to that spot for a few seconds then she will enter deeper in the selected location (e.g., in a room or in a painting inside the monumental building).

3 Evaluation

This paper focuses on the evaluation of the Virtual Binocular Interface. A first pilot evaluation was conducted by adapting the cyclical multi-direction pointing



Fig. 2. Illustration of the Virtual Binocular Interface interaction paradigm. On the left: the user mimics the gesture of pointing with a binocular; the shape of a binocular view is projected on the screen allowing the user to zoom on the displayed picture. On the right: the binocular shape changes colour, that is, if the user continues pointing to that spot for a few seconds then she will enter deeper in the selected location.

task paradigm of ISO 9241-9 [12,11]. By mimicking the use of a binocular, the user had to reach, one at a time, 8 visual points displayed in a circle. The binocular posture and movement were captured by infrared video cameras and the EyesWeb XMI open software platform [2] was used to develop the binocular interface. Preliminary results showed a relatively high difference between the movement times to reach targets in the vertical direction with respect to the horizontal one [1]. We hypothesized that the difference may result from the application of different algorithms to track user movements on the vertical and horizontal planes: the former one is based on optical flow, whereas the latter relies on geometric features.

We decided to investigate such difference in detail by adapting the serial Fitts' paradigm for pointing task in a new experiment [5]. The serial paradigm was preferred to assess interface usability by focusing on performance efficiency and comfort in repetitive tasks. Specifically, we aimed at assessing the effects of target orientation, type of algorithm and difficulty of the task on movement time, controlling for task error rate. The evaluation of participants' performance was then refined by considering two original features (*geometry entropy* and *directness indexes*) to obtain a more qualitative description of the task.

3.1 Set-Up

The Virtual Binocular Interface is implemented by analyzing in real-time the movements of the participants in a room through infrared cameras. The EyesWeb platform [2] is used to analyze in real-time the video signal and identify the relevant gestures of the participant.

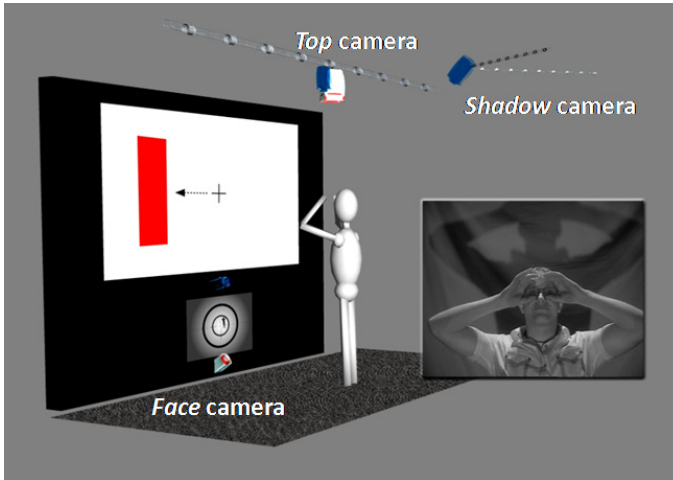


Fig. 3. Setup of the experiment

The participant is monitored by three infrared video cameras, the *top*, the *shadow*, and the *face* cameras (see Figure 3):

1. the top camera is installed above the user, to map from the vertical perspective the position on the floor of the head of the user.
2. the shadow camera is placed above the projection screen, looking toward the wall behind and above the user, where the shadow of the visitor upper-body part is projected by an IR light placed on the floor in front of the user. The shadow is analyzed to identify the binocular posture.
3. the face camera is installed below the projection screen, near the IR light, toward the user face, with an elevation of about 45 degrees. It is used to analyze the vertical movement of the user's face and control the up/down movement of the binocular.

The analysis of up-down movements is based on optical flow techniques applied to the face. Starting from the face camera, a region of interest corresponding to the user face is segmented and tracked. Then two optical flow algorithms have been selected and evaluated: the HornSchunck and the LucasKanade [6,7]. Left-right movements are estimated using visual information from the face camera or the top camera. In the first case, blob tracking techniques are used to detect the user rotation around the vertical axis. In the second case, the Lucas-Kanade optical flow is again used for tracking displacement along the horizontal axis. This leads to four different solutions to implement the Virtual Binocular Interface: Horn-Schunck (Algo1) and Lucas Kanade (Algo2) for up-down movements, blob tracking (Algo3) and Lucas Kanade optical flow (Algo4) for left-right movements. We tested all four solutions using a standard experimental procedure in order to select the best one, which has been chosen for the public permanent installation “Viaggiatori di Sguardo”.

3.2 Procedure

Participants. Twenty subjects (10 male, 10 female, age 23-50) were recruited and participated on a voluntary basis. All subjects were healthy and had normal or corrected-to-normal vision.

Task and Design. The serial tapping task developed by Fitts [5,11] was adapted to perform our test. Participants were instructed to alternately reach as quickly and accurately as possible between two targets shown on a screen in front of them of width W at a distance D both in the horizontal and in the vertical axis. The target was presented as a red rectangle on a screen of 240 x 180 cm with a resolution of 720 x 576 pixel (see Figure 3). Standing position in front of the screen was fixed by the area delimited by the top camera view (around 1.5 m from the screen). Following classic Fitts' law studies [5], movement difficulty was manipulated by varying target width (W) and distance (D). W ranged from 25 to 40 pixels. D ranged from 216 to 576 pixels. Four W/D combinations were generated to cover four indexes of difficulty (ID) that represent typical tasks performed by users in the Viaggiatori di Sguardo installation (e.g., performing short distance between large targets or performing long distance between small targets). For each ID, the participant had to go back and forth between the two targets for 10 trials. To control for order and sequence effects, the order of ID differed for each subject according to a balanced Latin Square. The number of trials per participant for testing the Virtual Binocular was: Orientation (2) x Algorithm (2) x ID (4) x Trials (10) = 160 trials.

Procedure and Instructions. The experiment started when the participant began mimicking the handling and use of the Virtual Binocular (i.e. raise both hands at the level of her/his eyes). The participants could then move the crosshair displayed in the center of the screen through their upper-body part movements and reached the targets one at a time, dwelling over each of them for at least 0.75s. A maximum time of 3 seconds was allowed to reach each target. If a target was missed, participants were instructed not to try to correct the error, but to continue to the next target. Before starting the experiment, a sequence of warm-up trials, which consisted in reaching 4 targets twice without error, was performed. During the experimental session, resting duration was freely decided by participants according to the level of their muscular/mental workload that resulted from the task demands. After the experiment, information about the subject background and verbal accounts of pointing strategies were collected through questionnaires developed by [3]. The overall duration of the experiment was about 20 minutes. This relatively brief duration with respect to usual experiment in this field [11] was motivated by effort required to stand in binocular posture.

3.3 Results

The experiment yielded 3200 data points (20 subjects x 2 orientations x 2 algorithms x 4 ID x 10 trials). In each ID condition, outliers of more than 3 SD

from the mean were not included in the movement time (MT) analysis. These removals left 3196 data points.

Mean Movement Time and Error Rate. A mixed three-way ANCOVA was performed on the Movement Time (MT) of participants with algorithm (Algo), Index of Difficulty (ID) and Orientation as factors and error rate as covariate. Main significant effects were found for ID ($F_{3, 2527} = 51.605, p < 0.001$), Orientation ($F_{2, 2527} = 3.876, p < 0.05$) and ID x Algo interaction ($F_{6, 2527} = 4.061, p < 0.001$). Bonferroni-corrected post-hoc comparisons revealed that Movement Time (MT) values for the targets located on vertical axis are significant lower than the targets located on horizontal axis confirming preliminary results. The MT values corresponding to higher Index of Difficulties were also significantly higher. In addition, a polynomial contrast of order 1 (linear) identified a positive trend of MT means for the four IDs. Post-Hoc comparisons of ID x Algo interaction effects showed that when the Lucas-Kanade algorithm is used for monitoring the participants along the horizontal axis (Algo 4), levels of difficulty (IDs) do not significantly affect Movement Time (MT). On the contrary, Algo 1,2 and 3 (respectively HornSchunck, LucasKanade for up-down movements and blob tracking for left-right ones), levels of difficulty (IDs) significantly affect Movement Time (MT) (see Figure 3). However, differences are more often significant when ordinal distance between Index of Difficulty is greater than 1 (e.g., Movement Time values (MT) in ID=4 are significantly only higher than MT values in ID=1 and ID=2 cases).

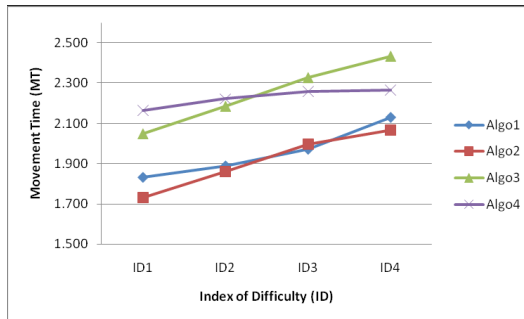


Fig. 4. Interaction plot of Algorithm (Algo) by Index of Difficulty (ID) effects. Algo1 and Algo2 respectively refer to the HornSchunck, LucasKanade solutions for monitoring up-down movements from the face camera; Algo3 and Algo4 respectively refer to blob tracking and Lucas-Kanade solutions for monitoring left-right movement from the top and face cameras.

Accuracy Measures. Following [8], a set of new accuracy measures was developed to supplement traditional ones, such as movement time and error rate, in assessing the Virtual Binocular Interface performances. Current features include directness and geometric entropy indexes to investigate the spread of the trajectories.

- *Directness index* (DI) is computed as the ratio between the length of the straight line connecting the first and last point of a trajectory (in this case, the line between the two targets) and the sum of the lengths of each segment composing the trajectory. It is inspired by the Space dimension of Laban’s Effort Theory [2].
- *Geometry Entropy Index* (GEI) is computed by taking the natural logarithm of twice the length of the trajectory (LP) divided by the perimeter of the convex hull around that path:

$$GEI = \ln\left(\frac{2 * LP}{c}\right) \quad (1)$$

where LP is the path length and c is the perimeter of the convex hull around LP . It is inspired by [10].

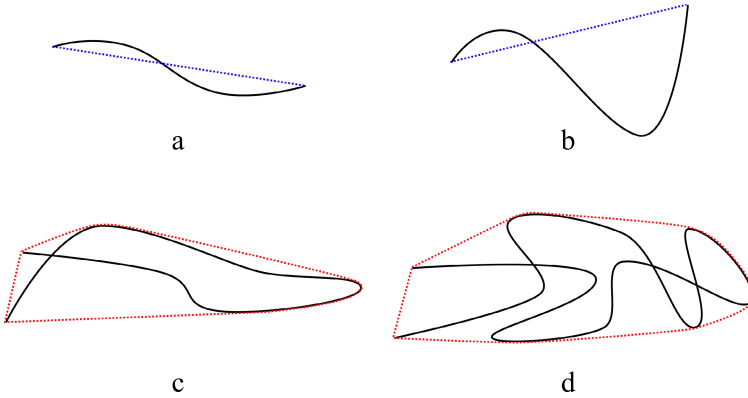


Fig. 5. Four trajectories (continuous black lines) exhibiting, respectively: (a) high DI, (b) low DI, (c) low GEI and (d) high GEI. The blue dotted lines represent the shortest paths between the starting and ending point of trajectories a and b. The red dotted lines represent the convex hulls of trajectories c and d.

In order to assess the validity of these new features, we replicated the statistical analysis conducted on Movement Time. A mixed three-way ANCOVA was performed, first on the Geometry Entropy Index (GEI) and secondly, on the Directness Index (DI) of participants with algorithm (Algo), Index of Difficulty (ID) and Orientation as factors and error rate as covariate.

In the Geometry Entropy Index case (GEI), a main significant effect was found for Algo ($F_{2,2528} = 55.061, p < 0.001$). For the Directness index (DI), main significant effects were found for ID ($F_{3,2528} = 3.866, p < 0.05$), Algo ($F_{2,2528} = 10.055, p < 0.001$) and ID x Algo interaction ($F_{6,2528} = 2.371, p < 0.05$). Post-hoc comparisons revealed that the directness index values (DI) augment with the level of difficulties, and at different intensity according to the considered algorithm. In particular, Algo 1 (Horn-Schunck) showed the highest progression in this respect.

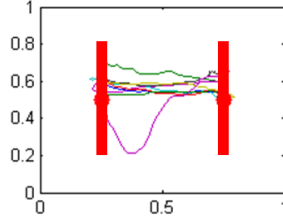


Fig. 6. Sample trajectories from a participant. The participants were instructed to select the red-highlighted target as quickly and accurately as possible. One block consisted of going back and forth between the two targets 10 times.

Questionnaire. The device assessment questionnaire consisted of 12 questions conforming to ISO 9241-9 [12,13]. The questions pertained to the Virtual Binocular. Each response was rated on a seven-point scale, with 7 as the most favorable response and 1 the least favorable response. Results (means and confidence intervals) are shown in Figure 7.

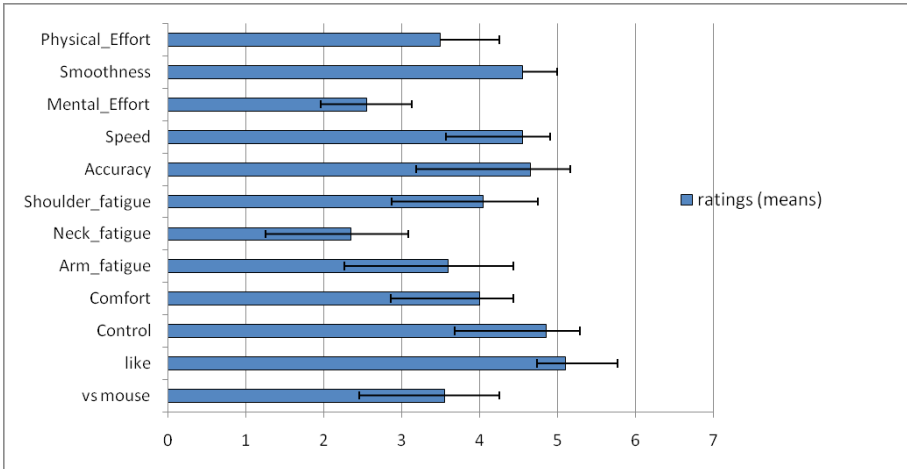


Fig. 7. Participants' ratings of the 12 items questionnaire

These new ratings confirmed that the participants enjoyed using the Virtual Binocular as pointing device (see like item, mean=5.2). The relatively high values obtained for smoothness, speed and accuracy and control (mean respectively of 4.5, 4.5, 4.6 and 4.9) may also suggest that the system efficiently support participants' pointing strategies. A repeated measure ANOVA showed that differences in ratings were significant ($F_{11,209} = 9.92, p < .001$). Bonferroni-corrected post-hoc comparisons revealed that ratings for mental effort and neck fatigue were

significantly lower than smoothness, speed, accuracy, control and like items . These results may ensure that the high performance of the system are enjoyed as they can be achieved through intuitive and natural gestural behavior. A confirmation on the good acceptance of the interface is the very high rating of the installation *Viaggiatori di Sguardo*, where we collected more than 1000 informal reports from visitors, resulting in about 98% of “high satisfaction” and “enjoyment” in the use of the interface in the context of “journey”/“exploration” of the audiovisual content of monumental buildings.

3.4 Discussion

The statistically significant differences between MT with respect to ID values and the observed linear trend may indicate that the modeling of our data can result from the application of Fitts law. In pointing tasks (i.e. for rapid aim movement), the values of IDs is said to predict movement time in a linear way [4]. However, the level of difficulty as assessed by the ID values can be moderated or increased by the technique employed for monitoring the participants movement (see the Lukas Kanade solution for left-right movement tracking). Additional elements may be considered to reach a better understanding of the Virtual Binocular performance. In this perspective, the information obtained through the application of the new accuracy measures (Geometry Entropy and Directness Indexes) confirmed and extend the outputs of the Movement Time analysis. The effect of ID x Algorithm interaction on the participants’ performance is for example confirmed. These results also help considering the way in which each ID and each algorithm solution affect the participant movement. According to the algorithm envisaged solution (for example Algo1, Horn-Schunck), a high level of difficulty can foster the participant in reducing the spread of the cursor trajectory. This qualitative information about the participant reactions help characterizing the usability of the Virtual Binocular in an objective manner. Other aspects of movement may relate to more high-level expressive dimension that may indirectly affect the performance, or at least the feeling of the participant in using the Virtual Binocular.

4 Conclusion

This paper presented an evaluation of the novel Virtual Binocular Interface. The binocular is one of the emerging examples of new active experience paradigms developed at our research centre. A sensitive environment equipped with non-invasive tracking technology and following interaction design principles enables a seamless interaction with the content through natural gesture.

We proposed a set of evaluation tools including qualitative analysis of motor performance. Results showed that the Geometry Entropy Index may be particularly suitable to evaluate the usability of an interface that support active embodied exploration in pointing tasks.

Future work includes enhancements to the interface, testing of further new modalities, metaphors, and paradigm of interaction, for increasing the degrees of freedom and possibilities of users in controlling cultural audiovisual content.

Acknowledgments. This work is partially supported by the EU FP7 ICT Project I-SEARCH (A unified framework for multimodal context Search; n°248296; DG INFSO Networked Media Unit; 2010-2012).

References

1. Camurri, A., Canepa, C., Coletta, P., Cavallero, F., Ghisio, S., Glowinski, D., Volpe, G.: Active experience of audiovisual cultural content: the virtual binocular interface. In: ACM Workshop eHeritage 2010, Firenze, Italy (2010)
2. Camurri, A., Mazzarino, B., Volpe, G.: Analysis of Expressive Gesture: The EyeWeb Expressive Gesture Processing Library. In: Camurri, A., Volpe, G. (eds.) GW 2003. LNCS (LNAI), vol. 2915, pp. 460–467. Springer, Heidelberg (2004)
3. Douglas, S.A., Kirkpatrick, A.E., MacKenzie, I.S.: Testing pointing device performance and user assessment with the iso 9241, part 9 standard. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: The CHI is the Limit. ACM (1999)
4. Drewes, H.: Only one fitts' law formula please! In: Proceedings of the 28th of the International Conference Extended Abstracts on Human Factors in Computing Systems, CHI EA 2010, pp. 2813–2822. ACM, New York (2010)
5. Fitts, P.M.: The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology* 47(6), 381–391 (1954)
6. Horn, B.K.P., Schunck, B.G.: Determining optical flow. *Artificial Intelligence* 17(1-3), 185–203 (1981)
7. Lucas, B.D., Kanade, T.: An iterative image registration technique with an application to stereo vision. In: International Joint Conference on Artificial Intelligence, vol. 3, pp. 674–679. Citeseer (1981)
8. MacKenzie, I.S., Kauppinen, T., Silfverberg, M.: Accuracy measures for evaluating computer pointing devices. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 9–16. ACM (2001)
9. Bianchi-Berthouze, N., Mussio, P.: Context and emotion aware visual interaction. *Journal of Visual Languages and Computing* 16, 383–385 (2005)
10. Cordier, P., Mendes France, M., Pailhous, J., Bolon, P.: Entropy as a global variable of the learning process. *Human Movement Science* 13, 745–763 (1994)
11. Soukoreff, R.W., MacKenzie, I.S.: Towards a standard for pointing device evaluation, perspectives on 27 years of fitts' law research in hci. *International Journal of Human-Computer Studies* 61(6), 751–789 (2004)
12. ISO International Organization for Standardization. 9241 9 ergonomic requirements for office work with visual display terminals (vdts) requirements for non-keyboard input devices (2000)

Does Movement Recognition Precision Affect the Player Experience in Exertion Games?

Jasmir Nijhar¹, Nadia Bianchi-Berthouze¹, and Gemma Boguslawski²

¹ UCLIC, University College London, MPEB Gower Street, London, WC1E6BT, UK

² PlayableGames, 22 Hand Court, Holborn, London, WC1V6JF
n.berthouze@ucl.ac.uk, jasmir.nijhar.09@gmail.com,
gemma.boguslawski@playablegames.net

Abstract. A new generation of exertion game controllers are emerging with a high level of movement recognition precision which can be described as the ability to accurately discriminate between complex movements with regards to gesture recognition and in turn provide better on-screen feedback. These controllers offer the possibility to create a more realistic set of controls but they may require more complex coordination skills. This study examines the effect of increased movement recognition precision on the exertion gaming experience. The results showed that increasing the level of movement recognition precision lead to higher levels of immersion. We argue that the reasons why players are more immersed vary on the basis of their individual motivations for playing (i.e. to ‘relax’ or to ‘achieve’).

Keywords: computer games, control devices, movement recognition precision, exertion games, immersion.

1 Introduction

Exertion games can be described as gaming interactions with technology in which users invest significant physical effort, and are believed to have social, mental and physical benefits [1]. In recent years these games have experienced massive commercial success due to the emergence of control devices that allow for a more natural type of interaction (e.g., Nintendo Wii). A new generation of these controllers such as Nintendo’s Wii Motion Plus, Sony’s PlayStation Move and Microsoft’s Kinect have entered the market. These new systems have a higher level of movement recognition precision than the first generation. These controllers offer the possibility to create a more realistic setting, which are then more likely to enhance the gaming experience by meeting up with players’ expectations.

This study will examine the impact of movement recognition precision on the gaming experience in exertion games. Increased levels of movement recognition precision should lead to a more realistic set of controls through movements being imposed and afforded.

2 Background

The exertion gaming experience can be split into 3 components: ‘motivations’ that players have when approaching the game, ‘strategies’ (i.e., playing styles) that they employ during the game, and ‘levels of immersion’ that the players reach during the game. We briefly review here the literature on these three components.

Lazzaro [7] identified the four motivations people have when playing computer games; 1. ‘Hard fun’ - gamers enjoy the obstacles and challenges presented in the game. 2. ‘Easy fun’ - gamers are driven by the sense of curiosity and adventure. 3. ‘Altered states’ - gamers play to experience sensations of excitement and enjoyment. 4. ‘People factor’ - gamers look for social interaction with others outside or inside the game. Pasch et al. [8] identified two types of motivation that occur in exertion games; 1. ‘Achieving’ - some people play with the motivation to challenge their ability and to find the best way to achieve a high score (i.e., hard fun). 2. ‘Relaxing’ – these people play with the motivation to relax (i.e., mental relaxation) by enjoying their movement skills (i.e., easy fun) without worrying about the scores. The control modality of a game can influence the motivation of the player [15], but it is unclear whether this applies to varying levels of movement recognition precision.

Pasch et al. [8] investigated the relationship between motivation and whole-body playing strategies. They showed that different motivations can lead to different strategies in whole-body sports games. Those whose motivation is to ‘achieve’ will optimize their strategy to obtain the most points by using the minimal movements required. While those whose motivation is to ‘relax’ will try to recreate movements from the actual sport. It is unclear whether this holds for different genres of exertion games and with players of different experience levels. It could be argued that controllers with an increased level of movement recognition precision will lead to a more realistic set of controls through movements being afforded and imposed. This in turn could influence a player’s choice of strategy, regardless of their motivation.

With respect to immersion, several studies [2, 8] have claimed that a player’s motivation or playing style can have an impact on the overall immersion level and/or different factors of immersion. However, this has yet to be explored in detail. An experiment by Bianchi-Berthouze et al. [2] compared a traditional control pad to a motion-sensed guitar shaped controller. Results suggested that an increase in body movement imposed, or allowed, by the game controller results in an increase in the player’s engagement level. The authors argue also that the increased involvement of the body can afford the player a stronger affective experience. Another study from Lindley et al [9] compared a traditional control pad to a set of Bongo drums which afforded natural movements. They showed that an increase in movement afforded by the input device made for a more engaging experience, and that this was not compromised by the increase in social interaction. All these results suggest that by imposing or allowing more movement in the game control can lead to an increased level of immersion. However, it is unclear what factors of immersion are being affected and it is still not clear if this would apply to controllers with better movement recognition precision. It is also unclear what type of imposed movement would facilitate these mechanisms [15, 17].

3 Research Focus and Experimental Design

This study will examine the impact of movement recognition precision on the gaming experience in exertion games by taking into account the motivation of the player. This study will investigate if an increased level of movement recognition precision leads to a more realistic strategy and to a higher immersion level, and also explore how the different motivation groups adapt their strategies.

The Nintendo Wii was selected for use in this study, as it supports many exertion games and also supports two movement based controllers with two levels of movement recognition precision, i.e., the Standard Wii Controller (called SC hereafter) and the Motion Plus Controller (called MPC hereafter) which has an increased level of movement recognition precision. The two levels of exertion game chosen were Tennis and Golf. EA Sports Grand Slam Tennis was chosen as it supports both controllers. Two different Golf games were used; Wii Sports (Golf) which supports the standard controller, and Wii Sports Resort (Golf), which is a sequel to Wii Sports that supports the motion plus controller. It is worth noting that both golf games have almost identical interfaces, the exact same courses, choice of clubs and characters, i.e. the only differences are the ones bought on by the motion plus. The user manuals along with the games advertising [12, 19] heavily imply that the motion plus games for both Tennis and Golf are a simulation of the real sport. This may have implications in the expectations raised in the players as further discussed in the conclusions.

From looking at description of the control systems detailed in the user manuals, and also from playing the game/tutorials with both controllers (standard and motion plus), we were able to list the differences (Table 1) that the increased movement recognition precision brings to these games.

Table 1. Differences between the Standard (SC) and Motion Plus Controllers (MPC)

Differences between SC and MPC
Accuracy and Responsiveness – With MPC, the swing trajectory is more accurately detected and replicated onscreen.
Swing Amplitude – MPC requires a larger swinging arm movement to initiate a swing. SC requires only a small movement in golf and just a small wrist movement in tennis.
Aiming System – With SC, aiming in tennis is determined by how early a player swings, whereas with MPC, the swing follow through determines the direction of the ball.
Power – MPC can detect the swing velocity in tennis.
Spin Shots – With MPC, a tennis player can add spin to shots by wrist rotations
Wrist Control – With MPC, a golf player must control wrist movements to perform a successful swing, e.g., twisting the wrist when swinging causes the ball to go off target.

The participants were split into 3 levels of experience: Beginner, Experienced with Wii, Experienced with Wii and Motion Plus. Each participant would experience both genres of game using both controllers. A counterbalancing table was made to minimise order and practice effects. To ascertain players' motivations for engagement, they were

interviewed straight after game play. The reason for interviewing after game play was because they might have not known what their motivation was before playing and it could have also changed during game play. To measure player's strategies they were video recorded during game play and the data was analysed by two evaluators who had experience in playing both Golf and Tennis. The evaluators were shown a series of video clips (44 clips in total) in a random order and then asked to rate both players in the clip on a scale of 1-5 (1 being unrealistic, 5 being realistic) based on their expert knowledge of the sports. Coding sessions were split up in to smaller sessions, to ensure the evaluators did not become too tired. In order to check the inter-rater reliability of the two evaluators, the intra-class correlation coefficient [20] was computed (0.9327) and the scores of the two evaluators were averaged to give each participant an average realism rating for each of the four conditions (2 games x 2 controllers). A motion capture system¹ was also used to obtain a more objective measure of swing and wrist movements (Table 2). Due to time consideration, the metrics were applied to 20 seconds of motion capture data (1200 frames) in order to capture a section of continuous game play, taken randomly from the middle of the game session. Whereas, analysing the full motion capture data would have provided a more accurate response, the fact that the 20 second windows were chosen randomly should provide sufficient accuracy.

Table 2. Motion Capture Movement Metrics

Tennis	Golf
<p>Swing Amplitude - This refers to how wide/open a player's tennis forehand swings are, i.e. the maximum range of swings, calculated from the rotation of the player's shoulder. The higher the amplitude, the more realism.</p>	<p>Swing Amplitude – This refers to how wide a player's golf swings are, i.e. the maximum range of swings, calculated from the rotation of the player's shoulder. The higher the amplitude, the more realism.</p>
<p>Max Speed (Power) – This refers to how fast a player swings. This metric is interesting because adding power to shots (i.e. by swinging faster) is a new feature in the motion plus condition.</p>	<p>Straightness of Swinging arm– This refers to how straight a player's swinging arm is i.e. the angular displacement of the swinging arm elbow. The straighter the swinging arm, the more realism.</p>
<p>Amount of Wrist Rotation – It refers at the amount of rotation perform in a spin shot, a new feature in the motion plus condition.</p>	<p>Amount of Wrist Rotation – This refers to whether a player is keeping a firm wrist (i.e. by not rotating it). This metric is interesting because the wrist control aspect is a new feature in the motion plus condition.</p>

To measure immersion, the immersion questionnaire developed by Jennett et al. [10] was chosen as it breaks down immersion into different factors, i.e., person factors (cognitive involvement, real world dissociation, emotional involvement) and game factors (challenge and control). Semi-structured interviews were also conducted as they allow participants to re-tell their game play experience which can reveal further experiential aspects [11]. The final game score was also recorded to measure performance as this was a possible confound that could affect immersion.

¹ Motion capture system: IGS-190-M with 18 gyroscopes. (<http://www.animazoo.com/>)

3.1 Participants and Procedure

A total of 22 participants were recruited (16 Male, 6 Female) ranging from 22 to 34 years old (average age = 27; st. dev = 3.2). This age range was chosen on the basis of a recent advert for EA Sports Grand Slam Tennis showing that it was marketed at 25-54 year olds [12]. Participants were paired to play the games by experience level, i.e., Beginners (4 pairs), Experienced with Wii (4 pairs), and Experienced in Wii and Motion Plus (3 pairs). Also, the members of each pair were friends. From a pre-trial questionnaire administered during recruitment, we were able to establish that all participants played video games at least once a month, exercised at least once a month and had either played tennis/golf or had an interest in tennis/golf.

On arrival, each pair of participants were asked to read an information sheet, health and safety form and sign a consent form. The experimenter would first load up the first game genre and then explain the first controller condition. As only one motion capture system was available at the time of this study, a member of the pair chosen randomly would wear the motion capture suit. The chosen participant was told that the suit would be capturing all of their movements during the actual experiment. Both participants would then participate in a practice session, where they were given an instructional sheet explaining the controls of the game, which they would be asked to read before the experimenter gave a demonstration of the control system. The participants would then have 5 minutes to get familiar with the controls, before starting the actual experiment which would also last 5 minutes. The participants would then be asked to fill in the immersion and answer questions about their motivations for playing the game. The above procedure would then be repeated for the second controller condition. After the participants had completed both controller conditions for the first game genre, the experimenter would then conduct a semi-structured interview. Finally, the whole procedure would be repeated for the second game genre. Each session lasted approximately 1 hour 30 minutes.

4 Controllers, Motivation and Strategy: Results

After each condition, participants were asked what their motivation was whilst playing. Responses were grouped into two categories – ‘Achieving’ and ‘Relaxing’. Responses such as ‘to challenge myself’, ‘to learn and improve’, ‘to beat my opponent’ and ‘to obtain a high score’, were attributed to the ‘Achieving’ group. Responses such as ‘to enjoy myself’, ‘have fun’, and ‘to experience real tennis/golf’ were attributed to the ‘Relaxing’ group. The results showed that each player’s motivations for playing were not affected by the type of the controller used.

A Multivariate Analysis of Variance (MANOVA) [21] using SPSS-18 software was then conducted to see if there was any relationship between player’s motivations and their strategy, i.e. average realism ratings. Even though the strategy realism rating data did not follow a normal distribution, the MANOVA analysis was conducted as it is robust over non-normality. The results showed that players whose motivation is to ‘relax’ use a significantly more realistic strategy than players whose motivation is to ‘achieve’, and this holds across genres and conditions (see Figure 1, first two graphs):

i.e. holds for Tennis in both the standard condition ($F=31.347$, $p=0.000$) and motion plus condition ($F=27.631$, $p=0.000$), and holds for Golf in both the SC condition ($F=5.69$, $p=0.30$) and MPC condition ($F=11.74$, $p=0.03$).

We also explored if the level of realism changed within the sample between the two controller conditions. The non-parametric Related-Samples Wilcoxon Signed Rank Test [22] was conducted, as the data did not follow a normal distribution. Players motivated to ‘achieve’ have a significantly higher realism rating when the level of movement recognition precision increases for both Tennis ($W=2.546$, $p=0.011$) and Golf ($W= 2.536$, $p=0.011$) (Figure 1, first graph). The players also have a significantly higher swing amplitude (measured with the motion capture) in both Tennis ($W=2.023$, $p=0.043$) and Golf ($W= 2.023$, $p=0.043$) in the MPC condition. Players motivated to ‘relax’ have a significantly higher realism rating when the level of movement recognition precision increases for Tennis ($W=2.232$, $p=0.026$) and for Golf ($W= 2.53$, $p=0.011$) (Figure 1, second graph). These players also have a significantly higher swing amplitude for Golf ($W= 1.782$, $p=0.075$) in the MPC condition. However, there was no significant difference for swing amplitude in Tennis for this motivation-group.

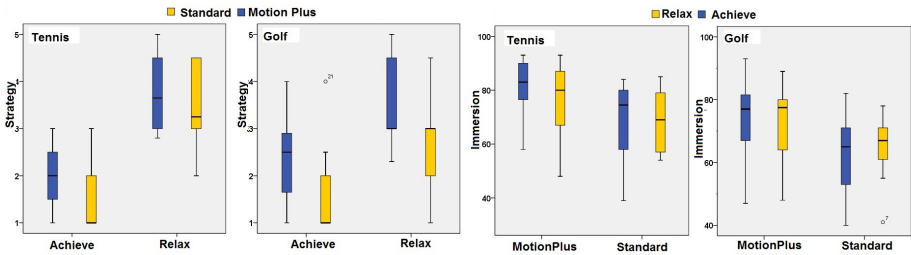


Fig. 1. Box-plots showing the effect of the increased Movement Recognition Precision on Strategy (Realism Rating) and Immersion for both motivation groups and games

Players motivated to ‘achieve’ will also use additional realistic movements (measured by the motion capture), if it will help them to achieve a higher score. For Tennis, these players will use significantly more wrist rotation, i.e. spin shots, when there is an increase in movement recognition precision ($W=2.023$, $p=0.043$), even though this movement is not required. This is due to the fact that ‘spin shots’ are more difficult for their opponent to return, i.e. they maximize all efforts towards achieving a higher score. For Golf, these players will also have significantly less wrist rotation ($W=-2.023$, $p=0.043$) and a straighter swinging arm ($W=2.023$, $p=0.043$) when there is an increase in movement recognition precision, even though these realistic movements are not required by the game. Keeping the wrist firm and having a straighter swinging arm are important for producing a good swing in the motion plus condition, as well as in real golf. So, these players have adapted their strategy in order to perform better.

Players motivated to ‘relax’ will overlook additional realistic movements. For Tennis, there was also no significant difference in maximum velocity and amount of wrist rotation between the standard condition and motion plus condition. For Golf, there was no significant difference in the amount of wrist rotation between the

conditions. These additional realistic movements all contribute to achieving a higher score in motion plus, but these players are not motivated to achieve, which could explain why there was no significant difference for these metrics.

5 Movement Recognition Precision Affects Immersion: Results

Given the importance of motivation, we investigated our hypothesis on each motivation group separately. Since the distribution of the overall immersion scores and factor scores did not follow a normal distribution, the non-parametric Related-Samples Wilcoxon Signed Rank Test was used.

Players motivated to 'achieve' have a significantly higher level of immersion in the MPC condition for both Tennis ($W=2.869$, $p=0.004$) and Golf ($W=2.938$, $p=0.003$) (Figure 1, last two graphs). For Tennis there was a significant increase in the level of Challenge ($W=2.966$, $p=0.003$), Control ($W=2.732$, $p=0.006$), Real World Disassociation ($W=2.764$, $p=0.006$), Emotional ($W=2.298$, $p=0.022$) and Cognitive Involvement ($W=1.836$, $p=0.066$) when the level movement recognition precision increased. For Golf there was a significant increase in the level of Challenge ($W=1.93$, $p=0.054$), Emotional ($W=2.673$, $p=0.008$) and Cognitive Involvement ($W=2.236$, $p=0.025$), when the level movement recognition precision increased.

Players motivated to 'relax' have a significantly higher level of immersion when the level of movement recognition precision increases and this holds for both genres i.e. for Tennis ($W=1.887$, $p=0.059$) and for Golf ($W=2.192$, $p=0.028$) (Figure 1, last two graphs). However it is worth noting that these increases are not as significant as the increases shown by players motivated to 'achieve', probably because the effects of the motion plus are not as apparent to players motivated to 'achieve'. For Tennis there was a significant increase in the level of Challenge ($W=1.851$, $p=0.064$) when the level of movement recognition precision increased. For Golf there was a significant increase in the level of Challenge ($W=2.021$, $p=0.043$), Emotional ($W=2.565$, $p=0.01$) and Cognitive Involvement ($W=2.259$, $p=0.024$), when the level movement recognition precision increased.

The non-parametric Related-Samples Wilcoxon Signed Rank tests were computed to exclude a possible effect of performance over immersion. The tests showed that for Tennis there were no significance differences in performances between controller conditions ($p\text{-value} = 1.0$), whereas for Golf there was a significance difference ($p\text{-value} = 0.04$). However, the correlation coefficients between these two sets of scores in the Golf condition showed very low correlation for both controllers (SC: person = 0.07; MPC: person = -0.1) indicating that the effect of performances on immersion was negligible.

From the analysis of the interviews it is clear that the increase level of movement recognition precision contributes to the level of immersion for both type of players, the ones that want to 'achieve' and the ones that want to 'relax'. In both case, the reason is that the controller fits better their expectations. For the players that are motivated to 'achieve' this means that the controller offers a more complex game (i.e., a large set of shots to make points) and, at the same time, the controller is not a barrier to immersion as it is more intuitive. As a result, players feel more challenged, cognitive and emotionally involved and more dissociated with the real world.

For the players motivated to ‘relax’, higher recognition precision means less frustration and the possibility to engage with the pleasure of moving. With low movement recognition precision these players reported to become frustrated by the poor accuracy and responsiveness of the controller. Instead, the increased movement recognition precision offers better ‘one-to-one’ response time between their actions and on-screen feedback. This allows the players to better enjoying their movement by playing more realistically, i.e., creating a better simulation as the controller meets better the players expectations. This decreased control barrier may have eventually brought them to feel more emotionally and cognitively involved in experiencing their movement skills.

6 Conclusions

The link between motivation and strategy in exertion games was clear – players whose motivation is to ‘relax’ will use a more realistic strategy than players whose motivation is to ‘achieve’, and this holds across genres and different levels of movement recognition precision. These results follow the Pasch et al. [8] study which showed that different motivations can lead to different strategies. Those whose motivation is to ‘achieve’ are looking to challenge themselves (hard fun), thus they will optimize their strategy to obtain the most points, i.e., an unrealistic ‘game’ strategy – using the minimal movements required. Instead, those whose motivation is to ‘relax’ are looking for mental relaxation (easy fun), thus they will try to recreate movements from the actual sport, i.e., a realistic ‘simulation’ strategy.

We first explored the effect of an increased movement recognition precision on players’ strategy. The results showed that players use a more realistic strategy as the level of movement recognition precision increases, and this holds across genres and motivation groups. However, the reason why players use more realistic strategies differs between motivation groups. Those motivated to ‘achieve’ use a more realistic strategy because the improvement in movement recognition requires them to, but only to a certain extent, i.e., the improved controller does not yet offer a fully accurate simulation. Those motivated to ‘relax’ will use a more realistic strategy possibly to reach a better simulation of the sport. However, these players overlook additional realistic movements that contribute to them achieving a higher score, as they are not motivated to achieve. Players motivated to ‘achieve’ become more immersed when the level of movement recognition precision increases. A possible reason for this is that an increased movement recognition precision allows for additional realistic movements which can help the player to ‘achieve’, i.e., their motivation for engagement. A second reason could be that these additional movements allow for a more exciting game play which allows the player to become more emotionally involved in the game [2, 9, 15]. These additional movements also make the game more challenging and require the player to think more, thus allowing them to become more cognitively involved [13].

Players motivated to ‘relax’ also become more immersed when the level of movement recognition precision increases, however the increase is not as significant as those motivated to ‘achieve’. The reason for higher immersion in this case could be due to the fact that the controller with increased movement recognition is more responsive and accurate at replicating movements allowing the player to focus and

enjoying movement per se. The players can play more realistically, thus meeting the player's expectation [4]. Also, from creating a better simulation, the players will use more body movement, which according to various studies [e.g., 11, 15, 16, 18] will afford the player a stronger affective experience. Their movements are also more accurately replicated and it is easier for them to anticipate what will happen next in response to their actions. This is backed up by Slater et al. [14] who state that presence in virtual environments may be enhanced the stronger the match between proprioceptive information from human body movements and sensory feedback from computer generated displays. This in turn facilitates the player's empathy for the character they are playing [6], and increases the emotional involvement. Further support to our conclusion could be obtained by running a longitudinal study to understand the effect of prolonged exposure to the controllers. Also, a more thorough analysis of motion capture data (e.g. segmentation between gestures) could provide further insights on the metrics to evaluate the players' experience in exertion games.

Our study has successfully shown that increasing the level of movement recognition precision will lead to a richer gaming experience with higher levels of immersion. The reason why players are more immersed differs between their individual motivations for engagement, but the underlying core reason is the same. The increased movement recognition precision makes the control system more realistic and therefore is better at meeting up with the players' expectations that they build from the real world. Controllers that match the user's expectation can enhance the gaming experience [3], while inappropriate controllers can create a breakdown in the gaming experience [5, 6]. However, movement recognition precision is still not at a level where it can create an exact simulation of a sport or activity, i.e. completely meeting up to a player's expectation. Further developments in movement recognition precision could facilitate this.

To conclude, this study has further advanced theory which showed that increasing the body movement imposed or afforded by a game controller results in an increase in the player's immersion level [2, 9]. This study also has relevance to exertion game designers and developers, by highlighting how important the design of controls are in shaping the gaming experience, i.e., the set of movement controls need to replicate movements from the actual sport or activity in order to meet players' expectations. Additional movements can be added to create more exciting and challenging game play; however designers need to be aware that these movements will not be used by all players, i.e., those motivated to 'relax'.

References

1. Mueller, F., Gibbs, M.R., Vetere, F.: Taxonomy of exertion games. In: Proceedings of the 20th Australasian Conference on Computer-Human interaction: Designing For Habitus and Habitat, vol. 287, pp. 263–266. ACM, New York (2008)
2. Bianchi-Berthouze, N., Kim, W.W., Patel, D.: Does Body Movement Engage You More in Digital Game Play? and Why? In: Paiva, A.C.R., Prada, R., Picard, R.W. (eds.) ACII 2007. LNCS, vol. 4738, pp. 102–113. Springer, Heidelberg (2007)
3. Li, N., Moraveji, N., Kimura, H., Ofek, E.: Improving the Experience of Controlling Avatars in Camera-Based Games Using Physical Input. In: Proceedings of the 14th Annual ACM International Conference on Multimedia (2006)

4. Höysniemi, J., Hämäläinen, P., Turkki, L.: Wizard of Oz prototyping of computer vision based action games for children. In: Conference on Interaction Design and Children: Building a Community (2004)
5. Dourish, P.: Where the Action Is - The Foundations of Embodied Interaction. The MIT Press, Cambridge (2001)
6. Rambusch, J.: The embodied and situated nature of computer game play. In: Proceedings of the Workshop on Cognitive Science of Games and Gameplay, Cogsci. (2006)
7. Lazzaro, N.: Why We Play Games: Four Keys to More Emotion Without Story, http://www.xeodesign.com/whyweplaygames/xeodesign_whyweplaygames.pdf (accessed)
8. Pasch, M., Bianchi-Berthouze, N., van Dijk, B., Nijholt, A.: Movement-based Sports Video Games: Investigating Motivation and Gaming Experience. *Entertainment Computing* 9(2), 169–180 (2009)
9. Lindley, S., Le Couteur, J., Bianchi-Berthouze, N.: Stirring up Experience through Movement in Game Play: Effects on Engagement and Social Behaviour. In: SIGCHI Conference on Human Factors in Computing Systems, pp. 511–514 (2008)
10. Jennett, C.I., Cox, A.L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., Walton, A.: Measuring and Defining the Experience of Immersion in Games. *International Journal of Human Computer Studies* (2008)
11. Mueller, F., Gibbs, M.R.: Evaluating a distributed physical leisure game for three players. In: Proceedings of the 19th Australasian Conference on Computer-Human Interaction: Entertaining User Interfaces, vol. 251, pp. 143–150. ACM, New York (2007)
12. MCN.com : Case Study – EA Sports Grand Slam Tennis, <http://www.mcn.com.au/Resource/CaseStudyDetail.aspx?IdDataSource=102> (accessed)
13. Bereiter, C., Scardamalia, M.: Surpassing ourselves: An inquiry into the nature and implications of expertise. Open Court, Chicago (1993)
14. Slater, M., Usoh, M., Steed, A.: Taking steps: the influence of a walking technique on presence in virtual reality. *ACM Trans. Comput.-Hum. Interact.* 2(3), 201–219 (1995)
15. Bianchi-Berthouze, N.: Does body movement affect the player engagement experience? In: Conference on Kansei Engineering and Emotion Research, pp. 1953–1963 (2010)
16. Mueller, F., Agamanolis, S., Picard, R.: Exertion interfaces: Sports over a distance for social bonding and fun. In: International Conference on Human Factors in Computing Systems CHI 2003, pp. 561–568. ACM Press (2003)
17. Muller, F., Bianchi-Berthouze, N.: Evaluating Exertion Games Experiences from Investigating Movement Based. *Human-Computer Interaction Series*, Part 4, pp. 187–207. Springer, Heidelberg (2010)
18. Hoysniemi, J.: International survey on the Dance Dance Revolution game. *Computer Entertainment. Section: Games, user interface and performing arts* 4(2), 8 (2006)
19. Wired.com: Wii Sports Resort Makes Golfing Real Again (2009), <http://www.wired.com/geekdad/2009/08/wii-sports-resort-makes-golfing-real-again/> (accessed)
20. ShROUT, P.E., FLEISS, J.L.: Intra-class Correlations: Uses in Assessing Rater Reliability. *Psychological Bulletin* 86(2) (1979)
21. Stevens, J.P.: Applied multivariate statistics for the social sciences. Lawrence Erlbaum, Mahwah (2002)
22. Wilcoxon, F.: Individual comparisons by ranking methods. *Biometrics Bulletin* 1(6), 80–83 (1945)

Elckerlyc in Practice – On the Integration of a BML Realizer in Real Applications

Dennis Reidsma and Herwin van Welbergen*

Human Media Interaction, University of Twente, The Netherlands

d.reidsma@utwente.nl

<http://hmi.ewi.utwente.nl>

Abstract. Building a complete virtual human application from scratch is a daunting task, and it makes sense to rely on existing platforms for behavior generation. When building such an interactive application, one needs to be able to adapt and extend the capabilities of the virtual human offered by the platform, without having to make invasive modifications to the platform itself. This paper describes how Elckerlyc, a novel platform for controlling a virtual human, offers these possibilities.

Keywords: Virtual Humans, Embodied Conversational Agents, Architecture, System Integration, Customization.

1 Introduction

Virtual Humans (VHs) are used in many educational and entertainment settings: serious gaming, interactive information kiosks, kinetic and social training, tour guides, storytelling entertainment, tutoring, interactive virtual dancers, entertaining games, motivational coaches, and many more. Building a complete VH from scratch is a daunting task, and it makes sense to rely on existing platforms. However, when one builds a novel interactive VH application, one needs to be able to adapt and extend the capabilities of the VH offered by the platform, without having to make invasive modifications to the platform itself.

The SAIBA framework [1] provides a good starting point for designing interactive VHs. Its emerging Behavior Markup Language (BML) defines a specification of the form and relative timing of the behavior (e.g. speech, facial expression, gesture) that a BML Realizer should display on the embodiment of a VH.

Elckerlyc is a state-of-the-art BML Realizer. Elsewhere, we described its mixed dynamics capabilities, that allow one to combine physics simulation with other types of animation, and its focus on continuous interaction, which allows it to monitor its own performance and allows for last moment modification of behavior plans with respect to content and timing, which makes it very suitable for VH applications requiring high responsiveness to the behavior of the user [2]. Here, we will focus on its role as a component in a larger application.

* This research has been supported by the GATE project, funded by the Dutch Organization for Scientific Research (NWO) and the Dutch ICT Regie.

2 Requirements for a Modular and Extensible Realizer

An application that uses a VH as one of its components might have several requirements for the BML Realizer. Specific additional gestures and face expressions might be needed; the application might need to run distributed over several machines; the experimenter might need detailed logs of everything that the VH does; one might want to replace the graphical embodiment of the VH, or its voice; the embodiment of the VH might need to reside in a custom game engine instead of Elckerlyc’s default renderer; and one might need to plug in completely new custom behaviors and modalities for a specific usage context.

Developing extensions or alternative configurations of Elckerlyc should be possible without requiring changes to the core Elckerlyc system (that is, extensions should not require recompilation of the Elckerlyc source). After all, if Elckerlyc extensions lead to a modification of Elckerlyc itself, then this would essentially lead to a separate Elckerlyc fork for every application using Elckerlyc. This would make it difficult to share new extensions with the community. Also, once Elckerlyc has been forked to accommodate a new modality engine or behavior type, it becomes difficult to take advantage of improvements in the ‘core’ Elckerlyc source: they need to be painstakingly merged into the fork.

Below follows a number of extensibility requirements for Elckerlyc, that should be implemented as *non-invasive modifications*: they may entail the implementation of new *run-time libraries*, or the addition of new *resources*; but should not require *compile time* dependencies for Elckerlyc on new code.

- Integration with new renderers, speech synthesizers, physics simulators, ...
- Flexible ways to send BML to the realizer, and to adapt the BML stream with capabilities for filtering and logging.
- Provide a transparent mapping from input (BML behavior elements) to output (control of the VH’s embodiment).
- Provide possibilities to add new behavior types or output modalities.
- Provide easy ways to integrate a BML Realizer as a component in an application, independent of variables such as the OS and programming language on which the application is developed.

3 Related Work

Like Elckerlyc, the BML Realizers Smartbody [3], EMBR [4] and Greta [5] were specifically designed for integration with existing renderers, to allow a wide range of behavior types, and/or to facilitate integration in different applications. Elckerlyc additionally contributes a transparent and adjustable mapping from BML to output behaviors (rather than the mostly hardcoded mappings in other realizers), and allows for easy integration of new modalities and embodiments, for example to control robotic embodiments. In this section, we discuss how various requirements were solved for the three realizers mentioned above, and shortly indicate the differences with our solutions. In the next section, we will go deeper into the solutions used in Elckerlyc, also showing how they impact actual use.

Integration with Existing Renderers. Smartbody provides the BoneBus library to connect the Smartbody realizer to a renderer. BoneBus uses UDP to transport (facial) bone positions and rotations from the realizer to the renderer. BoneBus is designed to hide the details of the exact communication protocol used, so that its exact implementation can be changed at a later stage without changing realizers or renderers that use the library. As the data transport protocol is non-trivial and due to change, reimplementing BoneBus in programming languages other than C++ or using the BoneBus interface with other transport mechanisms (TCP/IP, shared memory, etc.) is infeasible. Currently, SmartBody has been integrated with a number of renderers. The output of Greta contains MPEG-4 facial and body action parameters. By using the MPEG-4 standard, Greta can potentially be used with any renderer that supports MPEG-4. However, MPEG-4 –especially for body animation– is not widely supported.

Elckerlyc currently uses the Thrift remote procedure call (RPC) framework [6] to handle its communication with the renderer. Unlike the BoneBus library, this allows us to set up a communication channel that is agnostic to the programming language used on either side and that allows one to configure and change the mode of transport (e.g. TCP/IP, shared memory, pipes).

Available Behavior Types and Extensibility. Smartbody use keyframe animation and a fixed set of biologically motivated motion controllers (e.g. for gaze) to achieve facial and body motion. EMBR uses keyframe animation, procedural animation with a fixed set of expressive parameters, autonomous motion (such as eyeblink and balancing), morph targets for facial animation, and controllable shaders (e.g. for blushing). Greta uses procedural body animation with a fixed set of expressivity parameters, and Ekman’s action units [7] for facial animation.

Elckerlyc allows all of the above, and adds physically simulated animation and audio (sound effect) behaviors. More importantly, we contribute the ability to add custom behavior types and new output modalities *without* requiring modifications to Elckerlyc’s source code, described in Sections [4.3] and [4.4].

Integrating the Realizer as a Component in an Application. SmartBody offers integration with the Active MQ messaging system to provide independency of platforms and programming language, and to allow distributed setups. EMBR and Greta offer integration with the SEMAINE/Active MQ [8] messaging frameworks to achieve this; Greta additionally offers integration with Psyclone.

Elckerlyc uses Ports and Adapters to facilitate quick development of support for new types of integration; current implementations include support for the SEMAINE/Active MQ system and a simple direct TCP/IP connection. In Section [4.1] we discuss this in detail, and also touch upon several other things made possible by this architectural feature.

4 Design of a Flexible and Extensible BML Realizer

In this section we discuss the elements in Elckerlyc’s architecture that facilitate configuration, extension, and adaptation of the system. We start with a global

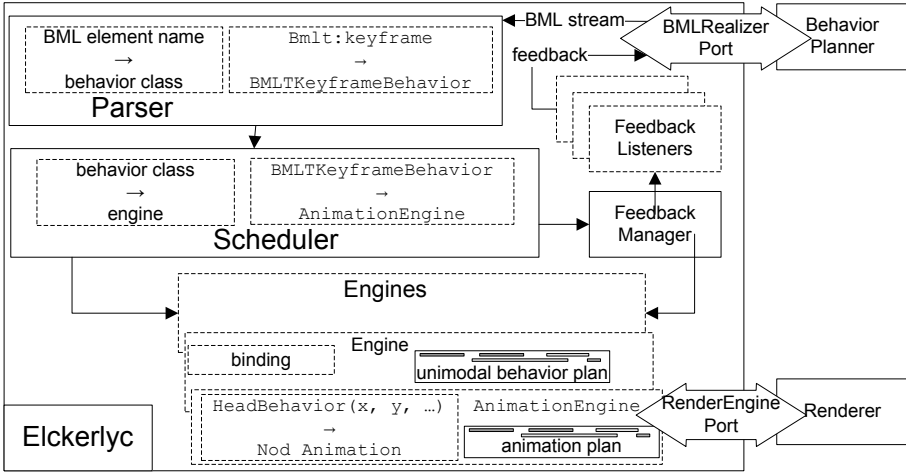


Fig. 1. Elckerlyc's architecture

overview. After that, we discuss the main possibilities in detail. For each topic we first sketch a ‘user need’; subsequently, we show which elements of Elckerlyc are designed to meet that user need, and how one uses them.

Fig. 1 shows the relevant parts of the architecture. Dashed boxes indicate components that can be changed at initialization, black boxes indicate unchangeable components. The Behavior Planner controls the VH by sending a stream of BML Blocks to Elckerlyc through a BML Realizer Port. Section 4.1 discusses how Ports can be used, e.g., to integrate Elckerlyc with various distributed messaging systems. The Parser parses the BML stream, and provides the Scheduler with a list of BML behavior elements and time constraints between these elements. Section 4.3 discusses how to add custom BML behavior elements. The Scheduler generates an execution plan, based on these elements and constraints. Different Engines (e.g., a speech engine, an animation engine, a face engine) keep track of, and manage, unimodal plans for their specific modality. Section 4.4 discusses how to add new Engines. Engines are also responsible for translating behavior elements to a form that is actually displayed on the embodiment of the VH. Section 4.2 discusses how this mapping from abstract behavior element to concrete forms can be reconfigured. The final resulting animation is sent to the Renderer. Section 4.5 shows how new Renderers can be integrated with Elckerlyc.

4.1 Ports, Pipes, and Adapters

User need 1: Integrating Elckerlyc as component in an application

Elckerlyc is designed to be used as component in a larger application context. The application may need to run distributed over several machines, platforms, and programming languages. The developer may want to log all interactions for post-hoc analysis. Nevertheless, the interface between Elckerlyc and application should remain as simple as possible: BML goes in; feedback comes out.

A minimal interface to a BML Realizer has functionality to (1) send a BML string to the Realizer and (2) register a listener for Realizer feedback. This is the BMLRealizerPort in Fig. 1. Both the Behavior Planner and the BML Realizer are connected to such a BMLRealizerPort. The adapter pattern [9] allows us to change the exact transport of BML and feedback to and from a BML Realizer, with no impact on the Behavior Planner and BML Realizer.

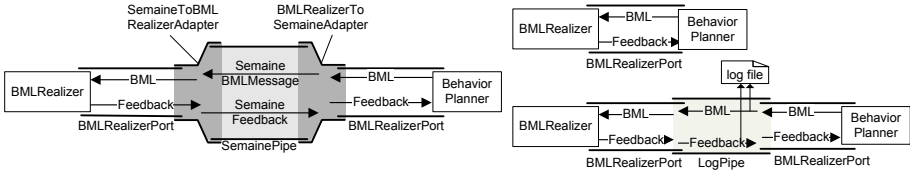


Fig. 2. Top right: the Realizer and BehaviorPlanner are connected directly on a RealizerPort. Left: the Realizer and BehaviorPlanner are connected through the Semaine API; they are unaware of this plumbing, they still communicate through RealizerPorts. Bottom right: a LogPipe logs the messages that pass through it to a file.

Elckerlyc implements the BMLRealizerPort interface. We have implemented Adapters that plug into BMLRealizerPorts and transport their messages over various messaging frameworks. Pipes are used to intercept BML and feedback, allowing one to measure it, let it go through slightly modified, or at a different rate. We have developed a pipe that logs the BML and feedback passing through, and one that buffers BML messages for a BMLRealizerPort that can only handle one BML message at a time. Fig. 2 shows some examples.

4.2 Gesture Binding and Other Bindings

User need 2: Transparently Mapping BML to Output Behaviors

BML provides abstract behavior elements to steer the behavior of a VH. A specific BML Realizer is free to make its own choices concerning how these abstract behaviors will be displayed on the VH's embodiment. For example, in Elckerlyc, an abstract 'beat gesture' is by default mapped to a procedural animation from the Greta repertoire. The developer may want to map the same abstract behavior to a different form, e.g., to a high quality motion captured gesture.

Elckerlyc's AnimationEngine uses a XML description, called the GestureBinding, to achieve a mapping from abstract BML behaviors to Plan Units that determine how the behavior will be displayed in the embodiment. The GestureBinding, clearly illustrated in Fig. 3, can be customized by the application developer; other Engines provide similar bindings.

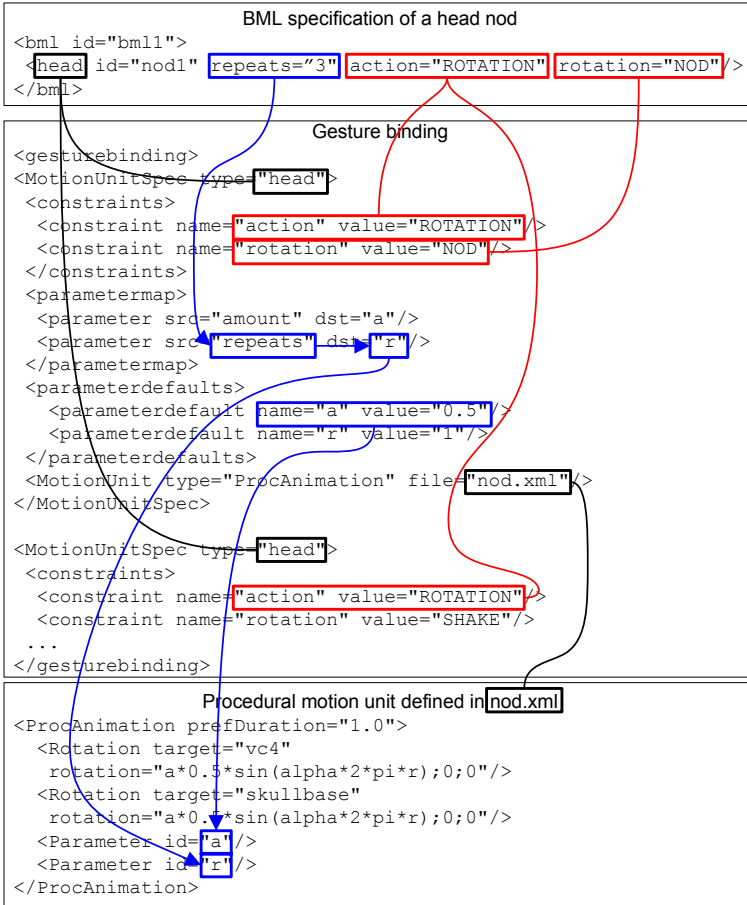


Fig. 3. Gesture Binding fragment binding the `head` element to the `nod` plan unit. Both the `nod` and `shake` motion units execute behaviors of type `head`. They both satisfy the constraint `action="ROTATION"`, but only the `nod` motion unit satisfies the constraint `rotation="NOD"` and is therefore selected to execute the head nod. The Gesture Binding maps the `repeats` parameter value in the BML behavior to the value of parameter `r` specified in the procedural motion unit. The value of parameter `a` is not defined in the BML head behavior, therefore the default value of `a`, as defined in the Gesture Binding, is used in the procedural animation.

4.3 BML Elements and Plan Units

User need 3: Adding new behavior types

Elckerlyc offers a large repertoire of Plan Unit types, in various Engines, that can be mapped in a Binding to give form to the abstract BML behaviors: physical simulation, procedural animation, morph target and MPEG-4 face control, Speech Units, etcetera. Still, a developer may need completely new Plan Unit types. For

example, to make the VH more lively, one may want to add a PerlinNoise Plan Unit that applies random noise to certain joints of the VH, as a kind of ‘idle motion’. Such new Plan Units need to become available in the GestureBinding (see previous section); furthermore, one might want to extend the XML format of BML with `<PerlinNoiseBehavior>` to allow direct specification of this idle motion by the Behavior Planner.

New BML behaviors are created by subclassing the abstract class `BMLBehaviorElement`; they can be registered with the Parser using a static call. At initialization of Elckerlyc, the new BML behavior type are coupled to a single Engine by adding it to the behavior class \rightarrow engine mapping (note that multiple behavior types can be coupled to the same Engine).

New PlanUnits implement the PlanUnit interface (for the AnimationEngine: rotate joints on the basis of time and animation parameters [2]). Such plan units are initialized from the GestureBinding through their class name (as a string), using Java’s reflection mechanism (that is, the ability to construct a new object from its class name). This ensures that any Plan Unit implementing the right interface for an Engine can be used in the Binding for that Engine without requiring additional compile time dependencies.

4.4 New Modality Engines

User need 4: Adding new modality Engines

The Nabaztag is a robot rabbit with ears that are controlled by servo motors and a body on which colored led lights are displayed. We needed to control this rabbit using BML, without encumbering Elckerlyc itself with Nabaztag specific code and libraries. To achieve this, we built a new Nabaztag Engine that was registered for handling all non-speech behaviors. For example, head nods were mapped in the Nabaztag Engine to a NabaztagPlanUnit that would move the ears shortly forward and back again; a sad face expression was mapped to a NabaztagPlanUnit that let the ears droop; etcetera.

Each Engine must implement the Engine interface (indicated by the lollipops in Fig. 4.1 top). All our current Engines are implemented on the basis of the `DefaultEngine`, a skeleton implementation of the interface. The `DefaultEngine` uses a `Planner`, `PlanManager`, `Player` and `PlanPlayer` and manages and plays a unimodal plan containing Plan Units (e.g. a gesture, a speech clause, etc.). The `Planner` resolves and constructs the unimodal plan on the basis of provided behavior elements and the constraints acting upon them. The `PlanManager` manages the unimodal plan and provides several functions to query its state or modify it. The `Player` plays back the units in the unimodal animation plan. In the `DefaultPlayer`, this functionality is fully delegated to a `PlanPlayer`. The `Animation Engine` and `Face Engine` require specialized `Players` that manage the combination of plan units that act simultaneously on the VH (e.g. physical simulation and keyframe animation), but can still delegate most of their playback

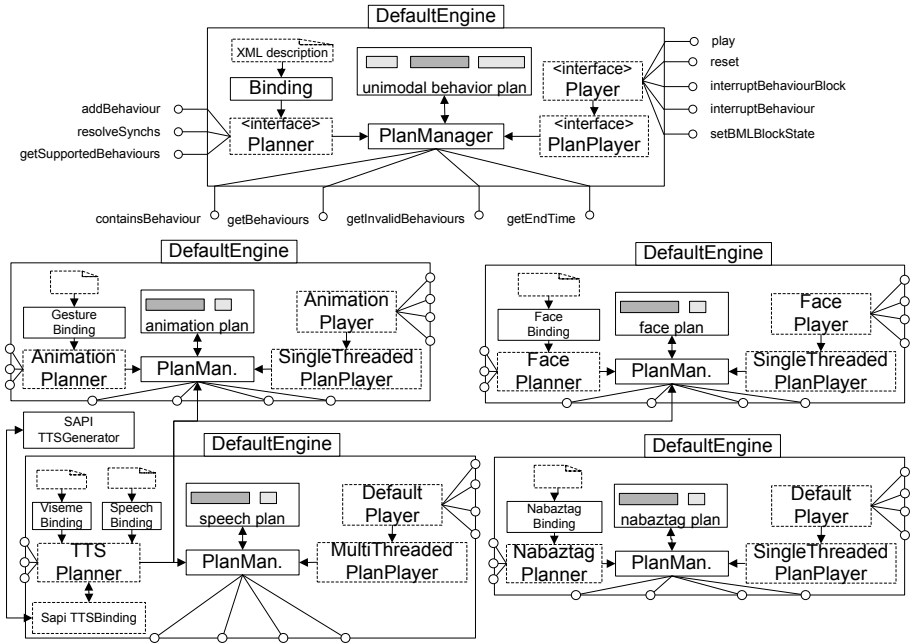


Fig. 4. Elckerlyc’s Default Engine setup (top), and the internals of (from left to right, top to bottom) the Animation Engine, Face Engine, Speech Engine and Nabaztag Engine. Dashed blocks are changeable at initialization. Note that the Speech Engine requires access to the PlanManagers that handle the Animation and Face Plans, to set up the facial movement co-occurring with speech. The Nabaztag Engine, like most other Engines, mostly uses the default Engine components.

functionality to a PlanPlayer. A MultiThreadedPlanPlayer plays its plan units in a separate thread. This is beneficial for plan units whose playback would otherwise block the playing thread.

The Nabaztag Engine. Building the new Nabaztag Engine involves developing the Plan Units that implement the basic control for the modality. A Plan Unit defines a way to control the robot – using one of its control primitives, see below – over the duration from the start time till the end of the Plan Unit. The control primitives for the Nabaztag robot are (1) move the ears of the robot to a specified position, (2) move the ears forward or backward by a specified amount, and (3) set one of the LEDs to a certain color. We implemented two Plan Unit types. The “MoveEarTo” Plan Unit moves the ears to a specified position by linear interpolation during the duration of the Plan Unit. The “WiggleEarTo” Plan Unit interpolates the ear from its current position to the specified target position and back to the starting point, during the duration of the Plan Unit, using a sinoid interpolation. Given these Plan Units, and a NabaztagBinding for mapping BML behaviors to Nabaztag PlanUnits, the Nabaztag Engine is

constructed using the standard available Engine components (see Fig. 4). A completely new modality Engine has been added by implementing two basic control Plan Units and an XML Binding. Due to the setup of Scheduler and Engines, synchronisation between the new Nabaztag Units and other modalities –e.g., speech– is automatically handled by Elckerlyc and requires no further implementation effort.

4.5 Integration with Renderers

User need 5: Integration with other rendering environments

By default, Elckerlyc renders the VH in its own OpenGL based rendering environment. One might, however, want to use Elckerlyc to animate an embodiment in another rendering environment such as Half Life, Ogre, or Blender.

To separate the renderer from Elckerlyc, we follow a design similar to that proposed by Russel and Blumberg [10]. The Animation Engine animates a local copy of the joint setup of the VH. The joint rotations set by Elckerlyc are copied to the renderer regularly (typically each frame).

The renderer therefore needs to support functionality to (1) provide Elckerlyc with the joint structure of the VH at its initialization, and (2) provide Elckerlyc with means to copy joint rotations to the virtual human in the renderer. Both requirements should be satisfied in a manner independent of renderer and transport (e.g. through TCP/IP, function call, shared memory). We use the remote procedural call framework Thrift [6] to achieve this. We have designed a language independent interface (using Thrift’s interface definition language) that a renderer should implement to achieve connectivity with Elckerlyc. This interface is automatically compiled to an interface in the target language of the renderer. The transport mode is chosen at initialization time. We have made a proof-of-concept implementation for the Ogre rendering environment.

5 Discussion

We have discussed how Elckerlyc can be tailored to the needs of specific applications, without requiring invasive modifications to Elckerlyc itself. Elckerlyc’s flexibility has allowed us to connect it to a behavior planner using either the SEMAINE framework or simple function calls, and to switch between such connections with a simple configuration option. The logging port allowed us to easily record all communication with Elckerlyc for user experiments, by simply changing the wiring between the behavior planner and Elckerlyc. The BMLRealizerPort also allowed us to exchange both the realizer and the behavior planner very easily. We have designed several behavior planners that implements behavior planning of a VH and one that replaces the VH behavior planning by a generic Wizard of Oz interface. The ability to easily replace the BML Realizer and behavior planner is also valuable for testing. We have designed a mockup BML Realizer that allows us to test behavior planners rapidly. This mockup

BML Realizer does not actually execute the BML behavior, but does provide the behavior planner with appropriate BML feedback. We have also designed a behavior planner that tests realizer implementations. This behavior planner executes test BML scripts on the realizer and inspects if the realizer provides the appropriate feedback. Since this test behavior planner communicates with the realizer through the generic BMLRealizerPort, it can not only test any configuration of Elckerlyc, but also potentially test Realizers designed by other research groups (by writing an adaptor from the BMLRealizerPort to their input and output channels). Elckerlyc's ability to add new modalities has allowed us to hook it up with the Nabaztag rabbit (see also Section 4.4) and to steer this rabbit with generic BML commands. The Nabaztag extension was achieved in a matter of days and did not require any changes in the Elckerlyc's source code.¹

Elckerlyc's extensibility is mainly achieved by a very flexible initialization stage. In this initialization stage, a desired setup of the Elckerlyc Realizer is constructed by combining and configuring different components that are provided by Elckerlyc's code base or by custom extensions. We have designed an XML configuration file format that describes such a configuration. Several default configurations are available, and new configurations are typically easily achieved by slight modifications of an existing configuration.

References

1. Kopp, S., Krenn, B., Marsella, S., Marshall, A.N., Pelachaud, C., Pirker, H., Thórisson, K.R., Vilhjálmsson, H.H.: Towards a Common Framework for Multimodal Generation: The Behavior Markup Language. In: Gratch, J., Young, M., Aylett, R.S., Ballin, D., Olivier, P. (eds.) IVA 2006. LNCS (LNAI), vol. 4133, pp. 205–217. Springer, Heidelberg (2006)
2. van Welbergen, H., Reidsma, D., Ruttkay, Z.M., Zwiers, J.: Elckerlyc: A BML realizer for continuous, multimodal interaction with a virtual human. *Journal on Multimodal User Interfaces* 3(4), 271–284 (2010)
3. Thiebaut, M., Marshall, A.N., Marsella, S., Kallmann, M.: Smartbody: Behavior realization for embodied conversational agents. In: AAMAS, pp. 151–158 (2008)
4. Heloir, A., Kipp, M.: Real-time animation of interactive agents: Specification and realization. *Applied Artificial Intelligence* 24(6), 510–529 (2010)
5. Mancini, M., Niewiadomski, R., Bevacqua, E., Pelachaud, C.: Greta: a SAIBA compliant ECA system. In: *Agents Conversationnels Animés* (2008)
6. Slee, M., Agarwal, A., Kwiatkowski, M.: Thrift: Scalable cross-language services implementation (2007)
7. Ekman, P., Friesen, W.: *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Consulting Psychologists Press, Palo Alto (1978)
8. Schröder, M.: The SEMAINE API: Towards a standards-based framework for building emotion-oriented systems. In: *Advances in Human-Computer Interaction* (319406) (2010)
9. Gamma, E., Helm, R., Johnson, R., Vlissides, J.: *Design Patterns: Elements of Reusable Object-Oriented Software*. Addison-Wesley (1995)
10. Russell, K.B., Blumberg, B.M.: Behavior-friendly graphics. In: *Computer Graphics International*, pp. 44–50. IEEE Computer Society (1999)

¹ See <http://hmi.ewi.utwente.nl/showcase/elckerlyc> for screenshots and movies.

Evaluation of the Mobile Orchestra Explorer Paradigm

Donald Glowinski, Maurizio Mancini, and Alberto Massari

InfoMus Lab, DIST, University of Genova, Italy
{donald,maurizio}@infomus.org, alby@infomus.dist.unige.it
<http://www.infomus.org>

Abstract. The Mobile Orchestra Explorer paradigm enables active experience of prerecorded music: users can navigate and express themselves in a shared (physical or virtual) orchestra space, populated by the sections of a prerecorded music. The user moves in a room with his/her mobile phone in his/her hand: the music performed by the orchestra sections is rendered according to the user position and movement. In this paper we present an evaluation study conducted during the Festival of Science 2010 in Genova, Italy. Forty participants interacted with the Mobile Orchestra Explorer and filled questionnaires about their active music listening experience.

Keywords: mobile, orchestra, paradigm, evaluation, explore, active listening.

1 Introduction

Active listening is a new concept in Human-Computer Interaction in which novel paradigms for expressive multimodal interfaces have been developed [2], empowering users to interact with and shape the audio content by intervening actively into the experience. Active listening applications are implemented using non-invasive technology and are based on natural gesture interaction [4].

The goal of this paper is to present the implementation and evaluation results of the Mobile Orchestra Explorer paradigm, developed in the framework of the EU Projects SAME [10] and MIROR [8].

The Mobile Orchestra Explorer paradigm entails the user to set up the position of a virtual orchestra instruments/sections and then to explore the resulting virtual ensemble by walking through the orchestra space.

This paradigm was tested during during the Festival of Science, a public event hold annually in Genova, Italy. Evaluation was carried out to study system usability and to produce an in-depth description of the user experience.

Section 2 introduces related work; in Section 3 we illustrate the Mobile Orchestra paradigm; finally in Section 4 we resume our evaluation study.

2 Related Work

Previous work *Orchestra Explorer* by Camurri et al. [3] allows users to physically navigate inside a virtual orchestra space, to actively explore the music piece the orchestra is playing, to modify and mold in real-time the music performance through expressive full-body movement and gesture. By walking and moving on a surface, the user discovers each single instrument and can operate through his/her expressive gestures on the music piece the instrument is playing.

Camurri et al. also propose a more sophisticated active listening concept, called *Mappe per Affetti Erranti* [2], where multiple users can physically navigate a polyphonic music piece, actively exploring it; further, they can intervene on the music performance modifying and molding its expressive content in real-time through non verbal full-body movement and expressive gesture.

Goto proposed a GUI-based system for intervening on prerecorded music with some original real-time signal processing techniques to select, skip and navigate sections of the recording [7].

Some projects addressed mobile music performances: *SonicCity* uses multi-modal sensors allowing a single user at creating music and manipulate sounds by using the physical urban environment as interface; information about the environment and the user's actions are captured and mapped onto real-time processing of urban sounds [6]. *SoundPryer* is a peer-to-peer application of mobile wireless ad hoc networking for PDAs, enabling users to share and listen to the music of people in vehicles in the immediate surrounding [9]. The *SonicPulse* system is an application designed to discover in a physical environment other mobile music users and engage with them sharing and co-listening to music [1].

3 Mobile Orchestra Explorer Evaluation Scenario

We present now the Mobile Orchestra Explorer scenario we evaluated during the Festival of Science in Genova, Italy, November 2010. In the scenario the user interacts with the system in two consecutive phases: (i) on the first phase, he/she walks in a sensitive empty space (a theater stage) holding he/she mobile phone in his/her hand and selects orchestra instruments/sections name on the mobile phone screen; when he/she reaches the point of the space in which he/she wants to place an instrument/section he/she press a button on the mobile phone to record its position; (ii) on the second phase he/she is allowed to move in the sensitive space and, as soon as he/she approaches an instrument/section position the corresponding pre-recorded audio track is played back. During both phases the user position is tracked by a fixed infrared camera.

The scenario architecture is represented in Figure 1. A fixed camera grabs frames of the theater stage at 25 frames per second and sends them to the SAME platform on which EyesWeb XMI is running. EyesWeb extracts the user silhouette from the frame background and computes the user barycenter position,

relative to the orchestra space. The user, by touching buttons on the screen of his/her mobile phone, sends the following commands to SAME platform:

command name	possible value	description
<i>mode</i>	<i>0,1</i>	indicates whether the interaction mode is either <i>setup</i> (0) or <i>explore</i> (1): the user either moves in the orchestra space to arrange the position of instruments/sections or is exploring the orchestra space and listens to the instrument/sections that are close to his/her current position.
<i>instrument</i>	<i>name</i>	(works only in setup mode): the user selects the instrument (or section) indicated by the parameter <i>name</i> .
<i>set</i>	<i>x,y</i>	the user sets the currently selected instrument (or section) position to the current user's position obtained by the camera frame.

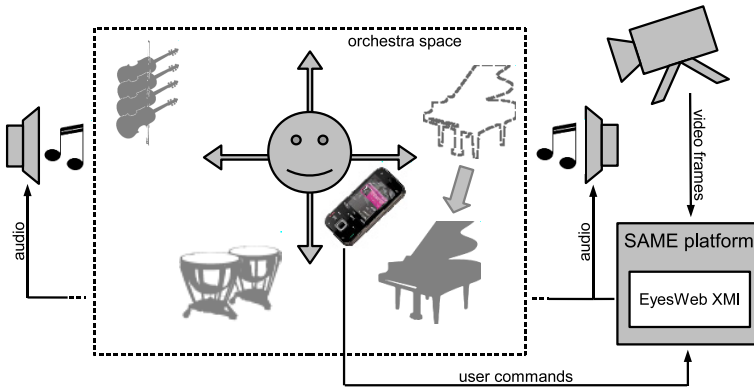


Fig. 1. The Mobile Orchestra Explorer evaluation scenario architecture

4 Evaluation

The evaluation of the scenario presented in Section 3 was conducted during the Festival of Science at Casa Paganini (Genova, Italy) in November 2010. Questionnaires were submitted to visitors. The first part of the questionnaire was the mobile user profile (i): it addresses how the visitor uses his/her mobile in daily life (see Table I); it also investigates his/her musical background (expert, music lover, etc.) and identifies the type and frequency of his/her physical activities (sport, dance, etc.). The second part focused on the evaluation of the Mobile Orchestra Explorer app (ii).

Table 1. Items related to the use of the Mobile Phone including standard call and sms functions plus multimedia applications (music listening, picture and video recording). Items were defined following the most recent standards [5][11].

	Never	Less than once a month	Once a month	Once a week	Several times a week	Once a day	Several times a day
1.1 MAKING CALL							
1.2 SENDING/RECEIVING SMS MESSAGES							
1.3 TAKING PICTURES <input type="checkbox"/> don't have this function							
1.4 RECORDING VIDEOS <input type="checkbox"/> don't have this function							
1.5 LISTENING TO MUSIC <input type="checkbox"/> don't have this function							
1.6 PLAYING GAMES <input type="checkbox"/> don't have this function							
1.7 OTHER MUSICAL APPLICATIONS Please specify:							

4.1 Mobile User Profile

Participants. 40 participants tested the applications (m=30.5, std 19.9), age ranged between 8 and 84 years old. It is worth mentioning that 50% of the population had less than 20 years old. All but one participant had a mobile (the 8-year-old child did not have one). Authorization to consent for the evaluation of the data was requested for any minor.

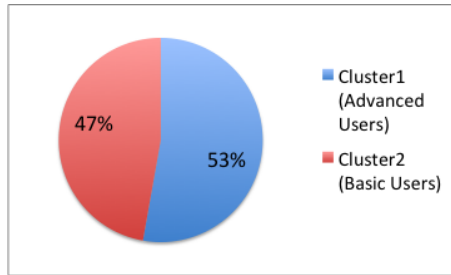


Fig. 2. Cluster Sizes

Mobile User Profile of the Participants: Advanced vs Basic Users. To identify user profiles among the visitors, the two-steps clustering technique, an unsupervised learning method implemented in spss (www.spss.com), were applied on the ratings related to the use of the mobile phone (Table 1). Results showed that two clusters emerged which divided the population of participants

in nearly two half-parts: Cluster1 and Cluster2 contained respectively 53 % and 47 % of the total sum of participants. The analysis of each cluster composition presented in the following sections showed that Cluster1 and Cluster2 refer respectively to group of Advanced and Basic users of mobile phone.

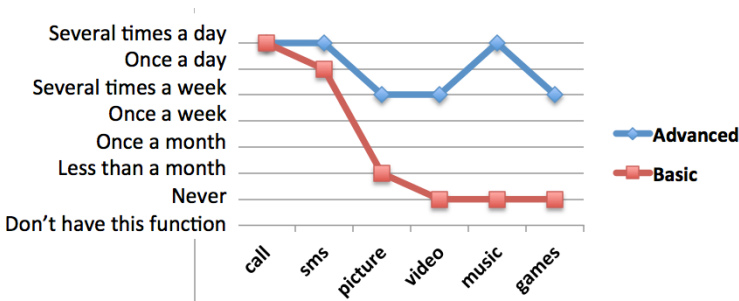


Fig. 3. Median plot of Advanced (Cluster1) vs Basic (Cluster2) Users frequencies by Items.z

Cluster Profile. The original ratings were ranked ordered and a Mann-Whitney U test was used to compare the ranks of the n=19 participants of Cluster1 (Advanced users) and n=17 participants of Cluster2 (Basic users) for all items. To control the inflation of type I error probability due to multiple comparisons, the Bonferroni correction was applied to *p*-values (the levels of statistical significance). The results indicate a significant difference between the ranks obtained by Cluster1’s participants versus the ranks of Cluster2’s participants for all items except for the ones corresponding to call and sms functions (see Table 2). For all other items, the mean rank of Cluster1 are higher than Cluster2’s one. The participants of Cluster1 tended using much more the other multimedia functions, with a particular interest for music listening (see Figure 3).

Table 2. Mean ranks are higher for Cluster1 with respect to Cluster2’s ones for items related to the multimedia functions of the mobile phone

	Picture	Video	Music	Games
z	4.57	4.93	4.76	3.51
Exact Sig.	<i>p</i> < .05	<i>p</i> < .05	<i>p</i> < .05	<i>p</i> < .05

Cluster Composition. To better understand the composition of each cluster, analysis of cross-classifications with respect to age and music expertise characteristics were conducted.

Analysis revealed that young participants (less than 20 years old) are more likely to be part of Cluster1 (Advanced users) (83.3%) than when being older (22.2%), $\eta^2(1, N=36) = 13.486 (p < .001)$, odds ratio = 17.54. The participants

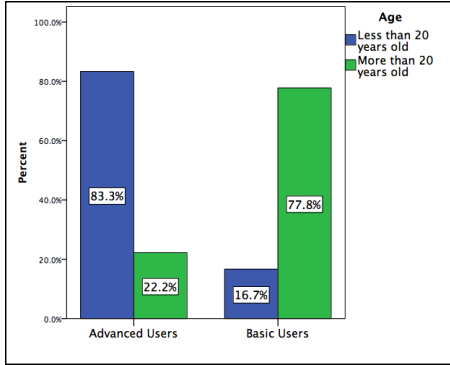


Fig. 4. Age distribution over the two clusters of participants

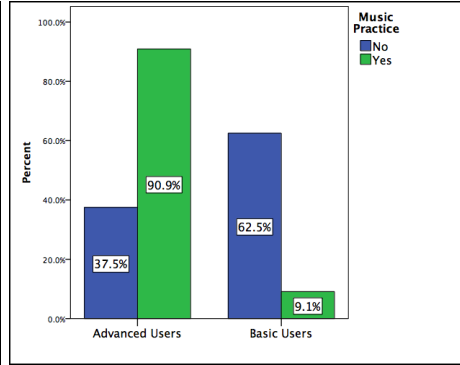


Fig. 5. Music expertise distribution over the two clusters of participants

playing a music instrument also seemed more likely to be part of Cluster1 (Advanced users) (90.9%) than when having no music practice experience (37.5%), $\eta^2(1, N=35) = 8.67$ ($p = .003$), odds ratio = 16.7. We wondered whether Advanced users felt predominantly at ease when testing the application. Point-biserial correlation coefficient between cluster membership and this extravert capacity revealed significantly high $r_{pb} = .414$, ($p < 0.05$).

Briefly said, Cluster1 members, i.e., Advanced users who take the most of the multimedia functionalities of their mobile phone were young, practice music and had no problem at testing applications among other people. However, as a whole, the population of participants have regular physical activities (67.5%), and listen to the music regularly, at least several times a week (94.2%), no significant differences were found between the two clusters for these characteristics.

4.2 Evaluation of the Mobile Orchestra Explorer Paradigm

The assessment questionnaire consisted of 6 items. First items were formulated to gain information about the usability of the application (e.g., level of understanding, of control) and user satisfaction in using it. Last items specifically addressed the music embodiment and active listening experience instantiated in the EU-ICT Project SAME (www.sameproject.eu), e.g., “Were you aware that you action modify the musical content?”. Each response was rated on an eleven-point scale, with 11 as the most favorable response and 1 the least favorable response. Results (median) are shown in Figure 6. Ratings confirmed that overall evaluation was positive ($m = 8.7$, $std = 0.7$). The participants enjoyed testing the Mobile Orchestra explorer ($m = 9.3$), found this application an interesting ($m = 9.4$) and engaging one ($m = 9$). Understanding and playfulness respectively received lower values but remain high (respectively $m = 8.2$ and $m = 8.7$).

A 2 (group) x 6 (item) mixed analysis of variance was run in order to investigate the effects of Group (Advanced vs Basic users), items and their interaction on participants answers. The Greenhouse-Geisser correction was used

when necessary to mitigate violations of the sphericity assumption in repeated measures. To control the inflation of type I error probability due to multiple comparisons, the Bonferroni correction was applied to P-values (the levels of statistical significance). The mixed ANOVA identified a significant main effect of Item, $F(3.165,104.44)=3.703, p < .05, \eta^2 = 0.9$, and of the Item x Group interaction, $F(3.17, 104.44)=3.77, p < .05, \eta^2 = 0.9$. Bonferroni-corrected post-hoc analyses were performed to assess specific difference among the Items and the Items x Group interaction effects. Item 6 (*if the mobile orchestra application was installed on their mobile, would they use it*) received significantly lower ratings than item 3 (*If they enjoyed using the application*) and 4 (*Interest of the application*). These results revealed that whereas this application may be enjoyed, and could capture the participant interest, it may not be considered as an application to use with their mobile phone. This effect is particularly high in the Basic User Group which already tend to under-use the multimedia functions.

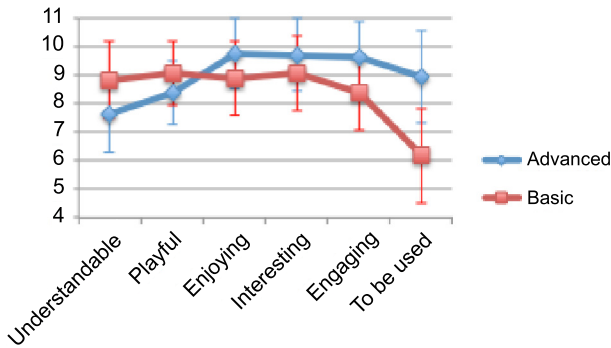


Fig. 6. Interaction plot of Group (Advanced vs Basic Users) by Items

Principal Component Analysis. Data for the 6 items were entered into a principal component analysis. The criterion of examining eigenvalues superior to 1 and elbows in the Screen plot suggested a two factors solution, which cumulatively accounted for 72% of the variance in the data. These factors were subjected to Varimax rotation. Considering the component loadings of the 6 original items (see loading factors in Table 3), and the component plot (see Figure 7), the two rotated factors could be respectively labeled as: Factor 1 (x axis) emotion (e.g., “how engaging and enjoying was the application”) Factor 2 (y axis) cognitive loads (e.g., “how easy the application was to understand and to play”)

Difference between Advanced and Basic Users with Respect to the Two Principal Components Ratings. Two Independent-samples t-tests were conducted to compare (i) the ratings of the first rotated component (emotion) and (ii) the ratings of the second rotated component (cognition) for the Advanced and the Basic users groups respectively. On average, Advanced and

Table 3. Rotated Component Matrix

Rotated Component Matrix^a

	Component	
	1	2
Understanding	.097	.831
Playing difficulty	.061	.536
Satisfaction	.942	.152
Interest	.910	.223
Engagement	.885	.172
Future application	.736	-.413

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 3 iterations.

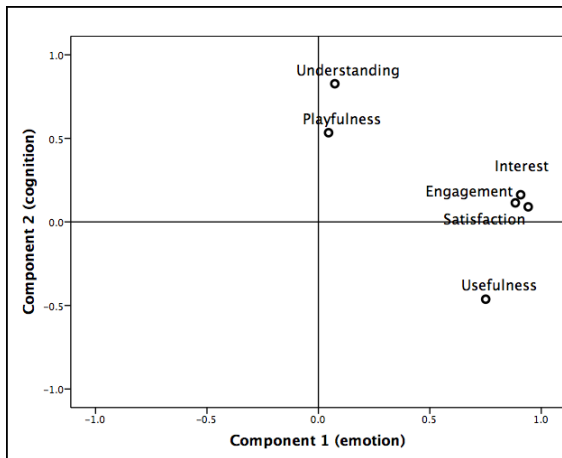


Fig. 7. Component plot in rotated space

Basic users may enjoy the application at a similar level (no significant difference between Group for the ratings of the emotion component, $t(33) = 1.67, p > .05$). However, difference between Group was significant for the ratings related to the cognition component, $t(33)=-2.04, p < .05, r = .101$. This result may be a paradox: one would actually expect an advanced user to be more acquainted with new technological devices and applications with respect to a more basic user. This result can otherwise be interpreted in another way: users with more experience in using multimedia functions of their mobiles are more demanding; they can therefore reveal more critics (with lower ratings) when considering the new functions offered by the mobile orchestra app.

Specific Items Related to the Active Listening Concept. Last items of the questionnaire specifically addressed issue related to the Active Listening Concept and its effect in terms of music discovery and learning: “Using this application allowed you to acquire a better knowledge of the instrument timbre?”, “The possibility given to explore physical the sound allowed you to memorize better the music piece?”. Results (means and confidence intervals) are shown for the Advanced vs Basic Users Groups in Figure 8. Ratings confirmed that overall evaluation was positive ($m = 6.4$). Independent samples t-test was conducted to compare the ratings of the music memorization for the Advanced and the Basic Users groups. Difference in ratings was significant, $t(20)=3.9, p < .01, r = .43$.

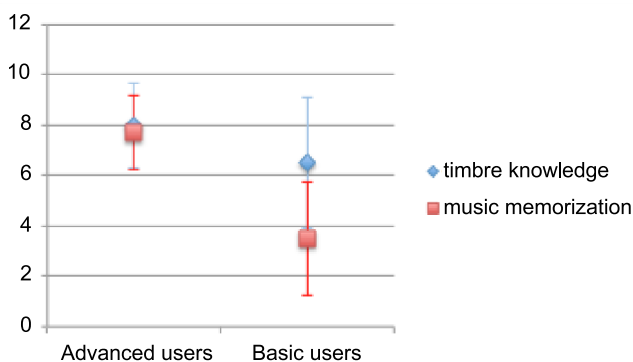


Fig. 8. Means and Confidence Intervals for timbre knowledge and music memorization items for Advanced and Basic Users Groups

These results suggest that using the Mobile Orchestra Explorer may help the participant recognize instrument timbre and memorize music sequences.

5 Conclusion

The Mobile Orchestra Explorer paradigm allows users to navigate and express themselves in a virtual Orchestra Space, populated by the sections of a pre-recorded music. We presented an evaluation study conducted during the Festival of Science 2010 in Genova, Italy. Several participants interacted with the Mobile Orchestra Explorer and filled questionnaires about their active music listening experience.

The results confirm that the Mobile Orchestra Explorer paradigm, considered as a proof-of-test of the active listening concept, has the potential to increase the user interest for music and to improve his capacity to distinguish between instrument timbres. Pre-Post tests should be specifically designed to investigate how the real-time temporal and spatial manipulation of a music material allowed by our application may facilitate the acquisition of these musical skills.

As a whole, this study may confirm the suitability of the active listening concept for entertainment and for didactic applications.

Acknowledgements. The work presented in this paper has been partially supported by the EU FP7 ICT Collaborative Project MIROR (Musical Interaction Relying On Reflexion) Grant n°258338, <http://www.mirrorproject.eu>. We thank Carlo Chiorri, Irene de Ferrari and Luca for their support as well as the anonymous reviewers for their useful suggestions.

References

1. Anttila, A.: SonicPulse: Exploring a Shared Music Space. In: 3rd International Workshop on Mobile Music Technology (2006)
2. Camurri, A., Canepa, C., Coletta, P., Mazzarino, B., Volpe, G.: Mapped Affetti Erranti: a Multimodal System for Social Active Listening and Expressive Performance. In: Proceedings of the 8th International Conference on New Interfaces for Musical Expression (2007)
3. Camurri, A., Canepa, C., Volpe, G.: Active listening to a virtual orchestra through an expressive gestural interface: The Orchestra Explorer. In: Proceedings of the 7th International Conference on New Interfaces for Musical Expression (2007)
4. Camurri, A., Volpe, G., Vinet, H., Bresin, R., Fabiani, M., Dubus, G., Maestre, E., Llop, J., Kleimola, J., Oksanen, S., Välimäki, V., Seppänen, J.: User-Centric Context-Aware Mobile Applications for Embodied Music Listening. In: Daras, P., Ibarra, O.M. (eds.) UCMedia 2009. LNICST, vol. 40, pp. 21–30. Springer, Heidelberg (2010)
5. Foss, R.D., Goodwin, A.H., McCartt, A.T., Hellinga, L.A.: Short-term effects of a teenage driver cell phone restriction. *Accident Analysis & Prevention* 41(3), 419–424 (2009)
6. Gaye, L., Mazé, R., Holmquist, L.E.: Sonic City: The Urban Environment as a Musical Interface. In: Proceedings of the 3rd International Conference on New Interfaces for Musical Expression (2003)
7. Goto, M.: Active Music Listening Interfaces Based on Signal Processing. In: Proceedings of the 2007 IEEE International Conference on Acoustics, Speech, and Signal Processing (2007)
8. MIROR, <http://www.mirrorproject.eu>
9. Östergren, M., Juhlin, O.: Sound Pryer: Truly Mobile Joint Listening. In: 1st International Workshop on Mobile Music Technology (2004)
10. SAME, <http://www.sameproject.eu>
11. Samkange-Zeeb, F., Berg, G., Blettner, M.: Validation of self-reported cellular phone use. *Journal of Exposure Science and Environmental Epidemiology* 14(3), 245–248 (2004)

As Wave Impels a Wave

Active Experience of Cultural Heritage and Artistic Content

Francesca Cavallero, Antonio Camurri, Corrado Canepa, Nicola Ferrari,
Barbara Mazzarino, and Gualtiero Volpe

Casa Paganini – InfoMus,
Piazza Santa Maria in Passione, 34, Genova
Università degli Studi di Genova
{cavallero,toni,corrado,bunny,volpe}@infomus.org

Abstract. This paper presents the interactive installation “Come un’Onda pre-muta da un’Onda” (“As Wave impels a Wave”, a citation from Ovidio’s “Metamorphoses” as a metaphor of time). The installation, presented in its early version at the Festival della Scienza 2009, introduces visitors to the rich history and artistic content of a monumental building: a virtual walk through the time. The core idea is to support an active experience based on novel paradigms of interaction and narration. The active experience is grounded on an informational environment characterized by an invisible “sound scent” map. The research is partially supported by the EU FP7 ICT I-SEARCH project.

Keywords: active experience of cultural and artistic content, multimodal audiovisual content search, Mixed Reality, museum ecology.

1 The Evolution of Museum Experience

The qualitative enhancement of a visit experience, according to the user behavior, is very important in a museographic context: for example, a dynamic control of lighting may influence the visitors flow, raise or lower the attention on specific exhibits, as well as the degrees and effectiveness of interactivity of an exhibit. The immersiveness can influence the understanding of a cultural message [1]. New technologies in the museums can be used to enhance the social interaction. Science centers and in general research on “infotainment” [2] contributed to a novel vision of “user experience”: e.g. the San Francisco Exploratorium APE-exhibits (for *Active Prolonged Engagement*) [3].

The aim is, in our case, to create novel adaptive cultural experiences to facilitate an active fruition of cultural heritage. We propose a mixed reality installation based on non-intrusive technology (without wearable devices) to enable an “explorative discovery” of a monumental site: the Casa Paganini building. The creation of a strong *sense of presence* [4, 5, 6, 7, 8, 9] in the mixed reality environment enhances a more conscious sense of the place in the user: the ancient building “talks” with a new voice.

In effect, the relationships between cultural heritage and its users should occur with all respects to (i) the artwork/historic building ontological state, and (ii) the new audience “infotainment needs”.

1.1 Active Experience of Cultural Content

With “active experience” we mean that users are enabled to interactively operate on audiovisual content in a cultural context, by modifying and molding it in real-time while experiencing. Two different perspectives contribute to achieve an effective active experience: *content-centric* and *user-centric*. The first concerns the need for a richer content modeling. The second concerns the various aspects of the users involved in the experience: for example, the understanding and exploitation of the behavior of the users while listening to audio content which is characterized by its stereo audio file and by a number of further data, including features obtained by means of advanced signal processing techniques (e.g., spatial rendering, to control the 3D audio localization of music sources). The active listening experience depends on and is shaped by the individual as well as the social behavior of the users. In this framework, the automated analysis of non-verbal user behavior (e.g. expressive gesture conveying emotions, social signals in small groups of people) supports the design of such multimodal systems. User Centric Media [10] will be able to analyze users' non verbal behavior, expressive gesture and intentions. In a museum context, for example, smartphones exploiting the growing number of sensors (e.g. videocamera, accelerometers, microphone...) can contribute to detailed real-time measures of visitors' behavior. State of the art approaches vary from wearable sensors to smart environments: the former are based on development of hand-held and context-sensitive prototypes. In this case, integrated sensors capture the current position of the user to make adaptive feedback (for example, with PDA devices). Ambient intelligence and user centric technologies were proposed to extend the possibility of interaction in museum spaces: for example, Chittaro and Ieronutti [11] focus on the tracing of users' behavior in virtual museum environments; Wakkary and Hatala [12] (see also [13]) explore “the design issues of 'situated play' within a museum through the study of a museum guide prototype that integrates tangible interfaces, audio displays and adaptive modeling” [12, p.171]; the ethnographic studies on museum visiting styles have been included in the research path of Zancanaro et al. [14], to personalize information through experimental mobile multimedia systems.

InfoMus–Casa Paganini developed since early nineties innovative interactive multimedia systems to enable active and social experiences of audiovisual content [15, 16]. Recent research provides novel engaging paradigms of interaction with pre-recorded music content, enabling a large number of non expert users to re-discover the musical heritage they may not be familiar with (EU-ICT Project SAME). Further, these active experience paradigms have been extended to the audiovisual content, in particular in a novel permanent interactive exhibition: *Viaggiatori di Sguardo* (Genova, Italy) enabling visitors to explore virtually the UNESCO Treasure of “Palazzi dei Rolli” in Genova (2010).

Liminality, Engagement and Place Identity. Bell [17] describes the museum in terms of cultural ecologies, identifying three components: *liminality*, *sociality* and *engagement*. The first feature defines museums as places where an experience apart from everyday life (rich in suggestion and occasions to pause and reflect) happens. The second defines museums as social places for groups such as pairs and families. The third proposes the museum experience as composed by learning and entertainment parts. Multimedia technologies increase these characteristics. In a Virtual Reality application, any action or interaction (and related feedback) happens in an inclusive space, in a 3D world where the “navigator” is able to freely move, following not pre-conditioned paths, but exploring in real time all available space [18]. The interactive installation we propose is based on Mixed Reality, “a particular subset of VR related technologies that involve the merging of real and virtual worlds somewhere along the ‘virtual continuum’ which connects completely real environments to completely virtual ones” [7, p.1321]. The full comprehension of the “place identity” [19, 20] of a monumental building is important to exploit the sense of liminality and engagement. From a phenomenological perspective it is possible to identify three dimensions [19]: *physical setting* (the concrete characteristics of the environment), *activities* (afforded by the place) and *meanings* (e.g. memories, associations, connotations and denotations linked to the place). A social aspect can be added. The dimension of *meanings* (through the history of the building) is particularly “dense” in the case of a monumental site, and the sense of place remains an *emergent property* [19] of interaction between individual and environment. We try to create an innovative experience of time through an experience of the space (having a strong “place identity”) (see also [21]).

Presence and Sense of Place. Presence is “a psychological state or a subjective perception in which (...) the subjects gets involved in the task, in objects, entities and event perception, as if technology was not present” [22, p.58]. Three among different characteristics of presence are very significant in our case: presence as *transportation*, *immersion* or *realism* [6]. The first concern the sensation of “you are there” (where the “there” is a real augmented place); the second involves the extent to which the senses are engaged by the mediated environment (but through disappearing technology); the third, finally, pertains to the degree to which a medium can seem perceptually realistic (the fresco fragments in the Spiral of the Time are in a display, but are coherent with the place and its characteristics, while the sounds evoke the past of the building). Presence is not only a *perceptual illusion of non-mediation* “produced by means of the disappearance of the medium from the conscious attention of the subject” [8, p.28]: presence can be also related to the concept of attention, especially in an installation where “only” fragments (audio and video) guide the interaction. The concentration and attentional factors are fundamental to determine the user’s sense of presence [23, 4]. The *embodied presence* theory [24] proposes a “mental representation of the environment in terms of pattern of possible actions, based on perception and memory” [9] (in “As Wave Impels a Wave”, for example, the user creates a mental map of sounds while perceives and discovers their position). The immersion is a result of the interaction between user and installation (man and environment): in turn, presence is a property that emerge from immersion, and the “being there” is enhanced by the possibility of “acting here” [25, 8].

Museum Ecology and Informational Environment. From an ecological perspective [26], the environment is, perceptually, correlated to the subject: if perceptions “support actions in the environment capturing opportunities, permissions or affordances” [8, p.30-31], we are able to use the spatialized sound to thinly guide the user along the discovery of the installation space, according movement and directionality in the 3D space (the approach or the moving away from the sound source) to loudness. In this “immersion perspective”, it is necessary to considerate the user, moving in a rich informational environment, as an “informavore” [27] to strongly engage her in the interaction. The installation designer has not to “extract” or “distract” the user from the “ecological” context (the monumental building): the challenge is to increase the sense of presence of the user in a context of MR, respecting (and communicating!) the strong “place identity” of the monumental site. The visitor is enabled to walk in a whispering, augmented space, following the “scent” of sounds, which guide her at the discovery of the multimedia content. In the interaction design discipline, “*information scent* is the (imperfect) perception of the value, cost, or access path of information sources obtained from proximal cues (...) representing the sources” [28, p.10]: “If the scent is strong, the information forager can make the correct choice; if there is no scent, the forager will have to perform a ‘random walk’ through the environment. The forager’s perception of which direction offers the optimal information source or patch is changed by sniffing for scent activities; so the forager is constantly adapting decision making and direction” [28, p.11].

2 Active Experience of a Monumental Building

We propose an active experience in sensitive spaces exploiting expressive body movement, in a perspective of valorization of cultural heritage and to enhance the level of engagement and communication of cultural content to visitors. The user is actively involved in her learning experience, watching and listening to cultural content embedded in a responsive environment.

The MR “virtual continuum” [7] is referred, in our case, to the “mixture” of classes of elements participating in the interaction experience: real environment (the Auditorium of Casa Paganini) is augmented by means of virtual objects (graphic fragments of a “Spiral of Time” in a large display and audio “spotlights” spatialized on the stage) evoking the history of the monumental building. “As Wave impels a Wave” (a citation from Ovidio’s “Metamorphoses” [XV, 181-184]), refers in its name to the chasing and coming waves as a metaphor for the time, which flees and follows as the past impels the present and the future. The goal of this installation is to increase the sense of presence of the user in a special, cultural place: it means, at the same time, increasing also the sense of place which “immerses” the visitor. In effect, a real place “is a particular space which is overlaid with meaning by individuals or group” [19, p.205], and it is possible to summarize this aspect in the equation “place=space+meaning” [29]: in a cultural heritage context, as the building of Casa Paganini, the challenge is to allow a dialogue between user and historical place. Therefore, on the one hand, the user is free from any device constraint (e.g. sensors, wearable

devices, handhelds, and PDA in general: increasing sense of presence by disappearing technology) and is enabled to interact only with the responsive environment; on the other hand, the building is respected as monumental and historical site (increasing sense of place “immersing” the visitor in the real, “ecological” environment).

The Monumental Site. Santa Maria delle Grazie La Nuova, named Casa Paganini, is a monumental site, rich in fresco paintings (from 15th to 18th century) and archaeological relics (6th-5th century BC). Along its history, the building was a roman villa, a convent (near to 15th century decorations and medieval persistences, the contributions of many artists actives in Genova between 16th and 18th century, e.g. Valerio Castello, Andrea Ansaldo, Giovanni Andrea Carlone, immediately stand out to the visitor crossing the ancient convent threshold. Religious decorations cycles centred on Marian main themes are common in the whole building), a typography, a ballroom, a school of music, a theatre, and finally, since 2005, after a long restoration (from 2000 to 2004), it hosts the Casa Paganini – InfoMus Intl. Research Centre (University of Genoa). The building includes an auditorium of 230 seats, a foyer, a gallery and museum rooms.

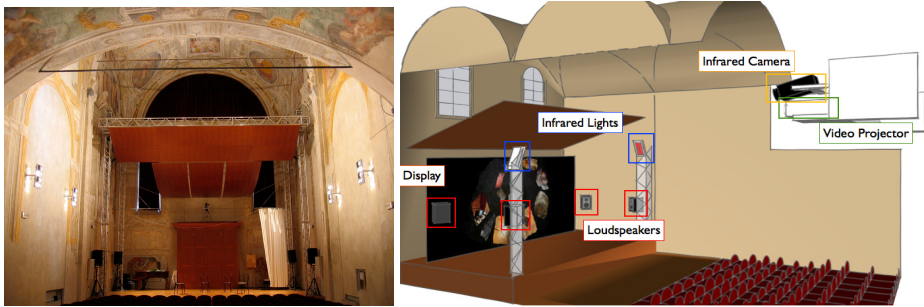


Fig. 1. The Casa Paganini Auditorium (the formerly church nave with fresco paintings) and the installation set up: an infrared camera detect the active area of the stage. The feedback of movement in this area is emitted by loudspeakers all around and on the big display of projector.

Set Up and Scenario. The dimensions of the sensitive space on the Auditorium stage are about 9.5x4.5 m. Infrared lighting up the surface where the users move, while 4 loudspeakers assure the sound spatialization. A large display (8,85x4m) is located on the wings as a “virtual wall” where the time-spiral appears during the interaction. The user path on the stage is detected by an infrared camera at 9,55m in vertical height. As is possible to mark from the Figure 1, each technological device is perfectly integrated in the monumental substrate.

The interactive experience is structured into a sequence of four layers, with growing levels of interactivity, each corresponding to a different paradigm of interaction. In this perspective, metaphors enhance the feeling of participation, the immersion and the sense of presence: for example, the “explorer”, the “archaeologist”, the “detective” are different roles that enable users to perform active experiences. The first layer of interaction is a simple walk in an “empty” space, to discover sonic objects hidden on the stage (*exploration paradigm*: the resonant environment); in the second layer,

visual fragments linked to the sounds emerge on the “virtual wall” (*discovery paradigm*: the spiral of time); at the third layer, finally, there is an in-depth experience of audiovisual narrative content related to the fragments discovered during the previous two phases of the experience (the “*archaeologist*” *metaphor* of interaction paradigm: to discover the hidden content). To achieve the experience of all the layers, it is necessary to discover all the fragments hidden in the sensitive space.

Layer 1–The Exploration. The visitor enters in the half-lighted Auditorium, then she approaches the stage. Under the fresco paintings of the formerly church single nave, an atmosphere of suspension wafts. The user now is immersed in the large, apparently empty space of the stage faintly whispering (a background, low-level soundscape is started as soon as the user enters the sensitive space). The user begins to move and, suddenly, mysterious short sounds emerge from the “whispering background”: a cannon shot, something like a brushstroke, a choral singing, etc.. The user walks and creates an invisible path on the space: she meets active places, and, step by step, she understands that every sound corresponds to specific regions on stage, like in a geographic map. Sounds are spatialized: a sound heard in a given space of the explored stage is perceived as its source were in that position, and the user motion towards a specific direction is “supported” by the “scent” of presence of audio content in others regions of the stage.

Layer 2–The Discovery. The visitor continues the exploration: whenever a sound previously discovered is met again, a corresponding visual fragment appears on a large display (a “virtual wall”) behind the stage. For example, the evoked image of a fragment of a fresco representing an angel is associated to a sound of brushstroke. The background projection on the display is similar to one of the ancient walls of Casa Paganini, which remains static until visual fragments emerge, like relics of semi-destroyed frescos. This evokes the experience of the visit to the real monumental building: in many cases only partial fragments on the walls and ceilings of the building are available. The visitor walks and discovers other fragments with their associated sonic materials. Each fragment is located like in a spiral: the Spiral of the Time.

Layer 3–The Archaeologist. Once the discovery phase is concluded, i.e. all the fragments of the building are brought to life, the third experience layer becomes available. By stopping a few seconds on a specific region of the stage, the user can now discover further hidden content, a short audiovisual clip explaining the meaning of the corresponding sonic and visual fragment discovered in the previous phases: when the visitor stops in correspondence to the “fresco angel” fragment, on the “virtual wall” a short documentary film about the painter starts. In the final version of the installation, the user can interact by her behavior with such audiovisual content, suspend, or interrupt to continue the exploration of the content linked to other fragments in the time-spiral. The interaction paradigm is the “archaeologist”: the user can go into more details, evoke other movie clips explaining and shading light to the “secrets” of the fragment she is exploring. The spiral-displacement on the display does not correspond to analogous positions on the active stage: the user has to explore the stage, remember the positions of the fragments to build the spiral of time, try to understand the content in order to discover hidden details (like an archaeologist) of the cultural content, buried under the virtual “dust”.

Layer 4—The Detective. The fragments of audiovisual content refer to objects (frescos, archaeological relics) which are concrete part of the building. Once the user terminates her “virtual” experience, she has collected elements to search and discover the real frescos and artworks in the building around her: the active experience on the stage has the role of stimulating the curiosity, in a “detective game”, aiming at discovering the building and its rooms in search of the tangible relics.

2.1 Crossing the “Limen”

In our system technology disappears, leaving the visitor the full control of experience, crossing the “limen” [17], in an “optimal flow” perspective [30]. According to the Gibson’s *affordance* [26], the installation exploits non-intrusivity, pervasivity and transparency of technical devices: the user body is the only interface to explore the space-time dimensions. The user confronts herself with a museographic site (the installation *with* and *into* the building), implements a creative editing between different temporal planes (the monumental site: to historically reconstruct and to analyze in its decorative and figurative context) and present (the laboratory of the Research Centre). We try to create an “invisible” scenery, where the information scent is represented, at the first layer, by the sound spatialization in 3D space: users should first “navigate” according to the sound (whispering in an increasing loudness up to the perceptual clairness), and when the location of sound source is founded, a bright fragment in the Spiral of Time appears, suggesting the possibility of others proximate significant regions (the second layer). We’re able to guide a “discovery walk” in the apparently empty space with a sort of “whispered sound texture” which acts as information scent. When the user approaches a meaningful area on the active space, a sort of sound-spot (a 2-3 seconds audio) progressively emerges from the background and attract her attention. The Auditorium is never really silent: a sound-map is virtually superimposed on the stage and the “information scent” [28] of every meaningful area contributes to create the background audio-texture (structured as a grid of interrelated gaussian regions). The user is

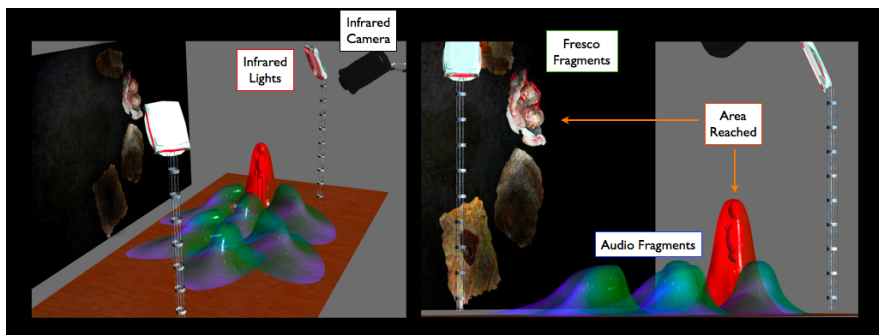


Fig. 2. In this installation scheme, the user discovers significant regions on the space, corresponding to specific sounds and, in the Spiral of Time, to different fresco fragments. Furthermore, in the whispering texture of the installation “soundscape”, the audio fragments surrounding the user current position increase their volume, attracting her in different areas.

free to follow any “sound-scent”, which includes the directional information on the related source, attracting her in one or another direction. The “scent” of informations is a “trigger” to explore, search, find audio and visual (virtual) fragments which the user is able to explore, search, find and discover also in the real environment around her. Any interaction layer allows a continuous information exchange [31] between installation (virtual space), monumental building (real place) and user.

3 System Implementation

EyesWeb XMI (for eXtended Multimodal Interaction) [32] is the open software platform supporting the design and development of the installation presented in this paper. The EyesWeb Trajectory Analysis Library contains a collection of modules for the extraction of features from trajectories in 2D (real or virtual) spaces, while the EyesWeb Space Analysis Library is based on a model considering a collection of discrete potential functions defined on a 2D space. In this case, the space is divided in a grid of active cells where the user is tracked (as a point moving in the space). Regions in the space can also be defined: it is possible that some regions exist on a stage in which the presence of movement is more meaningful than in other regions. The insisting of a user in a given place, or the trajectory to reach the place can influence the active mapping with audiovisual content. A certain number of “meaningful” regions (i.e., regions on which a particular focus is placed, a spot sound or a spiral fragment in our case) can be defined and cues can how much time a user occupied a given region to accede to hidden content.

4 Open Questions and Future Work

The first prototype of the installation provided useful feedback from the extensive experience with about 1700 of users during the Festival della Scienza (2009). Current work concerns a deeper use and exploitation of users' expressive gesture, which in the present prototype is used at a basic level. In effect, from observational analysis, user noticed a certain difficulty to follow exclusively the audio feedback in the empty space: a sort of “horror vacui” makes difficult the exploration, despite the “sound-scent” attracting the attention in different directions. To get over this problem, we avoided any silent area on the stage: when the user is far from any active region on the stage, she hears a diffused whispering resulting by the sonic scent of the nearest active regions. Such sonic scent is characterized by a 3D position and direction, and an emphasis of the loudness of its components corresponding to the user tendencies to move toward a given region. In this way, the directionality of the sonic scent is emphasized and anticipated.

Furthermore, we aim at extending the possibility to shape and mould the experience of audiovisual content (third layer), by taking into account the individual behavior and history (e.g. path trajectories) of the user: how a user behaves can influence the audiovisual content presented. In the basic version of the installation, multiple users cannot enter the space at the same time. In further versions of “As Wave Impels

a Wave”, we aim to introduce a social experience of the installation, by exploiting non verbal social signals: that is, to exploit the automated measures non-verbal social behavior of users during the experience [33], in order to adapt the cultural content presented to the group, depending, for example, on the activity, trajectory in the active space, and coherence of the behavior of the group.

Acknowledgments. We thank Prof. Lauro Magnani and our colleagues at Casa Paganini-InfoMus, in particular Paolo Coletta and Simone Ghisio. The research is partially supported by the EU FP7 ICT I-SEARCH project.

References

1. Dede, C.: Immersive Interfaces for Engagement and Learning. *Science Magazine* 323(5910), 66–69 (2009)
2. Delaney, J.: Ritual space in the Canadian museum of civilization. In: Shields, R. (ed.) *Life style Shopping*, pp. 136–148. Routledge, London (1992)
3. Tisdal, C., Perry, D.L.: *Going APE! At the Exploratorium*. Technical report, Selinda Research Associates, Inc. Chicago (2004)
4. Barfield, W., Weghorst, S.: The sense of presence within virtual environments: a conceptual framework. In: Salvendy, G., Smith, M. (eds.) *Human-Computer Interaction: application and case studies*, pp. 699–704. Elsevier, Amsterdam (1993)
5. Lessiter, J., Freeman, J., Keogh, E., Davidoff, J.D.: A Cross-Media Presence Questionnaire: The ITC Sense of Presence Inventory. *Presence: Teleoperators and Virtual Environments* 10(3), 282–297 (2001)
6. Lombard, M., Ditton, T.: At the heart of it all: The concept of Presence. *Journal of Computer Mediated Communication* 3(2) (1997)
7. Milgram, P., Kishino, F.: A Taxonomy of Mixed Reality Visual Displays. *IEICE Transactions on Information Systems* E77-D(12), 1321–1329 (1994)
8. Coelho, C., Tichon, J., Hine, T.J., Wallis, G., Riva, G.: Media Presence and Inner Presence: The Sense of Presence in Virtual Reality Technologies. In: Riva, G., Anguera, M.T., Wiederhold, B.K., Mantovani, F. (eds.) *From Communication to Presence: Cognition, Emotions and Culture towards the Ultimate Communicative Experience*, pp. 25–45. IOS Press, Amsterdam (2006)
9. Schuemie, M.J., Van Der Straaten, P., Krijn, M., Van Der Mast, C.: Research on Presence in VR: a Survey. *Cyberpsychology and Behavior* 4(2), 183–201 (2001)
10. Daras, P., Ibarra, O.M. (eds.): *UCMedia 2009*. LNCS, vol. 40. Springer, Heidelberg (2010)
11. Chittaro, L., Ieronutti, L.: A visual tool for tracing users’ behavior in virtual environments. In: *Conference on Advanced Visual Interfaces*, pp. 40–47. ACM Press, New York (2004)
12. Wakkary, R., Hatala, M.: Situated play in a tangible interface and adaptive audio museum guide. *Personal and Ubiquitous Computing* 11(3), 171–191 (2007)
13. Wakkary, R., Evernden, D.: Museum as ecology: A case study analysis of an ambient intelligent museum guide. In: Trant, J., Bearman, D. (eds.) *Proceedings of Museums and the Web Conference*. Archives & Museum Informatics, Toronto (2005)
14. Zancanaro, M., Kuflik, T., Boger, Z., Goren-Bar, D., Goldwasser, D.: Analyzing Museum Visitors’ Behavior Patterns. In: Conati, C., McCoy, K., Paliouras, G. (eds.) *UM 2007*. LNCS (LNAI), vol. 4511, pp. 238–246. Springer, Heidelberg (2007)

15. Camurri, A., Coglio, A.: An architecture for emotional agents. *IEEE-Multimedia* 5(4), 24–33 (1998)
16. Camurri, A., Ferrentino, P.: Interactive environments for Music and Multimedia. *Multimedia Systems, Special Issue* (1999)
17. Bell, G.: Making sense of museums. The museum as ‘cultural ecology’. *Intel* (2002)
18. Forte, M.: *Realtà Virtuale, beni culturali e cibernetica: un approccio ecosistemico*. *Archeologia e Calcolatori* (XV), 423–448 (2004)
19. Turner, P., Turner, S.: Place, Sense of Place and Presence. *Presence Teleoperators and Virtual Environments* 15(2), 204–217 (2006)
20. Relph, E.: *Place and Placelessness*. Pion Books, London (1976)
21. Rendò, L.A.: Presence and Mediated Space: a Review. *PsychNology Journal* 3(2), 181–199 (2005)
22. Lombard, M., Snyder-Duch, J.: Interactive advertising and Presence: a Framework. *Journal of Interactive Advertising* 1(2), 56–65 (2001)
23. Witmer, B.G., Singer, M.J.: Measuring presence in virtual environments: a presence questionnaire. *Presence: Teleoperators and Virtual Environments* 7, 225–240 (1998)
24. Schubert, T., Fiedmann, F., Regenbrecht, H.: Embodied presence in virtual environments. In: Paton, R., Neilson, I. (eds.) *Visual Representations and Interpretations*, pp. 268–278. Springer, London (1999)
25. Usoh, M., Alberto, C., Slater, M.: *Presence: experiments in the psychology of virtual environments* (1996), <http://www.cs.ucl.ac.uk/external/M.Usoh/vrpubs.html>
26. Gibson, J.J.: *The ecological approach to visual perception*. Houghton Mifflin, Boston (1979)
27. Dennett, D.C.: *Consciousness explained*. Little, Brown and Co, Boston (1991)
28. Pirolli, P., Card, S.K.: Information Foraging. *Psychological Review* 106(4), 643–675 (1999)
29. Harrison, S., Dourish, P.: Re-place-ing space: The roles of place and space in collaborative systems. In: *Conference on Computer-Supported Cooperative Work*, pp. 67–76. ACM, New York (1996)
30. Csikzentmihalyi, M.: *Flow: The Psychology of Optimal Experience*. Harper and Row, New York (1990)
31. Forte, M.: *Cibernetica e beni culturali: il problema della cornice*. In: *Workshop Interazione e Comunicazione Visuale nei Beni Culturali*, Perugia (2004)
32. Camurri, A., De Poli, G., Friberg, A., Leman, M., Volpe, G.: The mega project: Analysis and synthesis of multisensory expressive gesture in performing art applications. *Journal of New Music Research* 34(1), 5–21 (2005)
33. Varni, G., Camurri, A., Coletta, P., Volpe, G.: Toward a real-time automated measure of empathy and dominance. In: *International Conference on Computational Science and Engineering*, vol. 4, pp. 843–848. IEEE Press, Washington DC (2009)

An Intelligent Instructional Tool for Puppeteering in Virtual Shadow Puppet Play

Sirot Piman^{1,2} and Abdullah Zawawi Talib¹

¹ School of Computer Sciences, Universiti Sains Malaysia,
11800 USM Pulau Pinang, Malaysia

² Department of Business Computing, Faculty of Management Sciences,
Surat Thani Rajabhat University, Surat Thani Province 84100, Thailand
sirot.cod07@student.usm.my, azht@cs.usm.my

Abstract. Shadow puppet play has been a popular storytelling tradition for many centuries in many parts of Asia. In this paper, we present an initial idea and architecture of a software tool that allows people to experience shadow puppet play in the virtual world. Normally, a virtual puppet show is controlled automatically by the application. However, our tool allows the user to create storyline and control the puppets directly in real-time with a special device that can improve the skill of a puppeteer. This paper focuses in detail on the design and issues of a component of the software tool which is the intelligent instructional tool for puppeteering of virtual shadow puppet play. The result of the preliminary evaluation has shown that the tool is able to help users more beneficially and a higher degree of satisfaction among the respondents which include professional puppeteers and potential users.

Keywords: Shadow puppet play, virtual puppet, virtual storytelling.

1 Introduction

The history of the traditional shadow puppet play in Southeast Asia began several hundred years ago and has been in existence in various forms in both insular and mainland Southeast Asia. The show ranges from large-scale productions associated with classical court traditions to relatively small-size folk-art productions in small rural villages. This multifaceted performing art tradition that combines music, drama, literature and storytelling is presented with dramatic movements and visual effects. Most of the previously published materials on this subject deal with Indonesia (primarily Java and Bali) and to a lesser extent Thailand and Malaysia. Malaysian shadow puppet tradition can be considered as a bridge between Indonesian and Thai traditions since it shares certain characteristics with both [1]. Shadow puppet play is called Nang-Talung in Thailand, Wayang Kulit in Malaysia, and Wayang Kulit or Wangwayo in Indonesia. The word “wayang” is derived from a word meaning shadow or ghost. The puppeteer is called Nai-Nang in Thailand, and Tok Dalang in Malaysia and Indonesia.

The puppets are carved from cow's or buffalo's hide and painted differently for every puppet in order to portray a different character. These puppets are performed on a silhouette screen and the show is accompanied with dialogues and poetry narrated by the puppeteer. Each puppet is depicted differently. For examples, the main actress is graceful and serene; the main actor is playful and rather dandy; the king is typically powerful; and the queen looks noble and nice [2].

Currently, this tradition is becoming less popular and it is fairly difficult to learn and not easy for young generations to watch, appreciate and explore this traditional art. Furthermore, nowadays there are only a few professional puppeteers for this traditional shadow puppet play. The success of each show depends on the storyteller's ability to present the story to the audience effectively and make them happy throughout the show till the end of the performance. The key element in the performance is how the puppeteer presents the show in such a way that correct expression, intonation and gestures are used so that it looks real, and makes the values and morality behind the story understood and appreciated by the audience. However, nowadays, the traditional shadow puppet play is no longer performed frequently. Lack of expertise, difficulty in getting the traditional musical instruments, difficulty in producing the leather puppets, high cost of maintaining and producing the show, and proliferation of new media entertainment are some of the main factors that contribute to the lack of interest in this traditional show. Therefore, a tool is needed to transform this traditional storytelling into an interactive virtual storytelling environment. In this paper, we describe an initial idea on the architecture of the virtual shadow puppet play and concentrate on the intelligent instructional tool for puppeteering in the virtual shadow puppet play.

2 Related Work

The virtual storyteller, Papous [3] can tell a story and express some facial emotions, happiness, fear, and sadness according to the story just as real human storyteller. This virtual storyteller allows changes of the scene tags according to the environment. Virtual Puppet [4] is a virtual reality application that combines the ability of current graphics systems with creative and pleasant storytelling. Generally, it enables the user to control a puppet with simple movements in different locations. In Virtual Storyteller [5] and [6], the characters are implemented as semi-autonomous intelligent agents. Its advantage includes the incorporation of the narrative and the presentation levels as intelligent agents rather than as text-based story generation system. Furthermore, it uses an embodied speaking agent to present the generated stories using appropriate prosody and gestures.

An intelligent multimedia storytelling system, CONFUCIUS [7] creates human character animation from natural language. One of the advantages is that collision detection, autonomy, and multiple characters synchronization and coordination are

applied to this storytelling system. Besides, it uses Humanoid Animation (H-Anim) and MPEG 4 SHNC for realistic animation. An interactive storytelling using a mixed reality system by Cavazza et al. [8] provides more interaction for the user by allowing the user to take part as one of the roles in virtual storytelling.

From the perspective of shadow puppet play, Hsu and Li [9] utilized motion planning algorithms to generate Chinese shadow animation automatically according to user's high level input. The puppets cannot be controlled directly by the user, and thus the system does not allow interactive real-time play. Zhu et al. [10] used photon mapping to render out the shadow effect. However, the time taken to perform rendering does not allow real-time and interactive play.

Kim and Talib [11] described a practical framework for integrating the elements of the traditional shadow play environment in a virtual storyteller. This includes shadow rendering of puppets, challenges in mapping of the traditional shadow play to a virtual storyteller, and a methodology in undertaking such development. In virtual 'wayang' URL [12], the puppet can be moved anywhere but the arms and legs of the puppet do not swing while the puppet is moving. Lam and Talib [13] and [14], proposed a novel real-time method for modeling of puppet and its shadow image that allows interactive play of virtual shadow play using texture mapping and blending techniques. Special effects such as lighting and blurring effects for virtual shadow play environment were also developed using OpenGL platform.

In summary, we can conclude that most of the existing works on digital traditional shadow puppet play are not interactive, realistic and dynamic enough to allow user to play interactively. Most of the existing digital shadow play systems need the user to predefine the frames offline and in-between frames need to be generated using commercial software. A better way of presenting virtual shadow puppet play is by allowing interaction and real-time play of the shadow puppet play. In this paper, we describe an attempt to overcome these problems and limitations by describing the general architecture of a virtual shadow puppet play system and focusing on the intelligent instructional tool for puppeteering of virtual shadow puppet play.

3 Overview of the Virtual Shadow Puppet Play

The main aim of our work is to develop a software tool for puppeteering shadow puppet play that allows people to experience it interactively in real-time in an intelligent virtual world, creates realistic virtual shadow puppet play and environment, and provides a means of improving the skill of a puppeteer based on the knowledge and expertise of professional puppeteers. Fig. 1 provides a general architecture model of the software tool that enables users to play virtual shadow puppets interactively. It consists of three main components – an intelligent instructional tool for puppeteering, real-time animation and special effects of 3D virtual puppets and multi-puppets controlling and performing using a special device.

Generally, the software tools allows users to create a storyline with the help of the intelligent instructional tool which automatically generates iconic instructions called

“sequential iconic instructions” (SII). Users can control the puppets directly with a special device such as wii-remote device or other haptic devices through Multi-Puppets Controlling and Performing Module and Real-Time Animation and Special Effects of 3D Virtual Puppets Module (see also [14] and [15]) based on SII. The next section describes the main focus of this paper which is the architecture of the intelligent instructional tool for puppeteering in the virtual shadow puppet play.

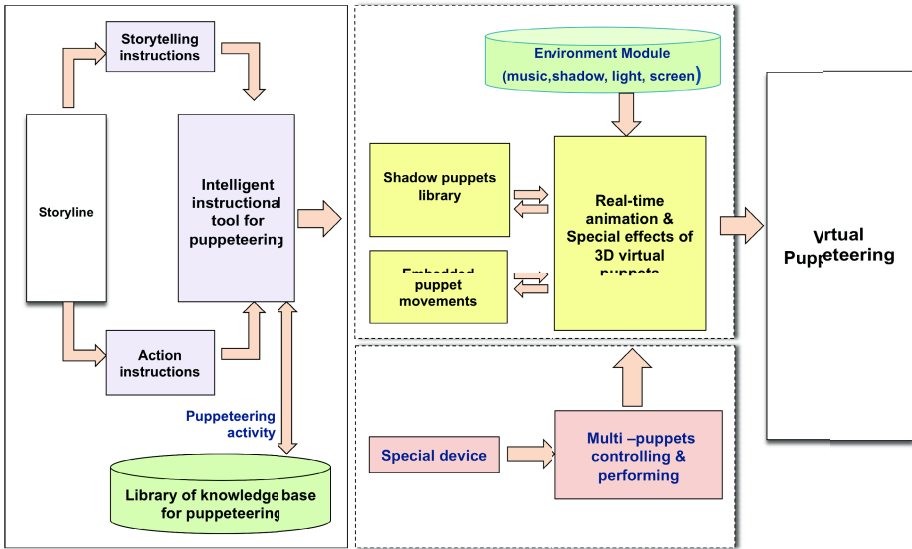


Fig. 1. General architecture model of virtual shadow puppet play

4 The Architecture of the Intelligent Instructional Tool for Puppeteer

This component enables the user to create storyline based on the knowledge bases of puppeteering. The knowledge bases were developed by conducting interviews on professional puppeteers and observing them during their performance. The knowledge base library stores three important database tables namely puppet table, action table and movement rule table. The puppet table stores all the information on each puppet such as its name, its picture, its description and level of its movement. The level of puppet’s movement refers to the different movement speed of each puppet. For example, the puppet “giant” moves very fast, the puppet “king” moves normally, the puppet “queen” moves softly and the puppets “tree” and “stone” do not move. We classify this into four levels: 0 – does not move, 1 - slow, 2 - normal and 3 – fast, depending on the character of each puppet. The action table stores the information on puppet movements such as ‘move to the left’, ‘move to the right’, ‘turn around’ and so on. The movement rule table stores all the action rules of the puppet’s movement

that we have gathered from professional puppeteers. This table is very important in Storyline Analysis and Action Checking Module of the intelligent instructional tool (Fig. 2). Then, the system automatically generates SII that guides the user to perform actions on the puppets according to the created storyline. The overall process of this component is shown in Fig. 2 and the next subsections describe in detail the modules of the intelligent instructional tool for puppeteering.

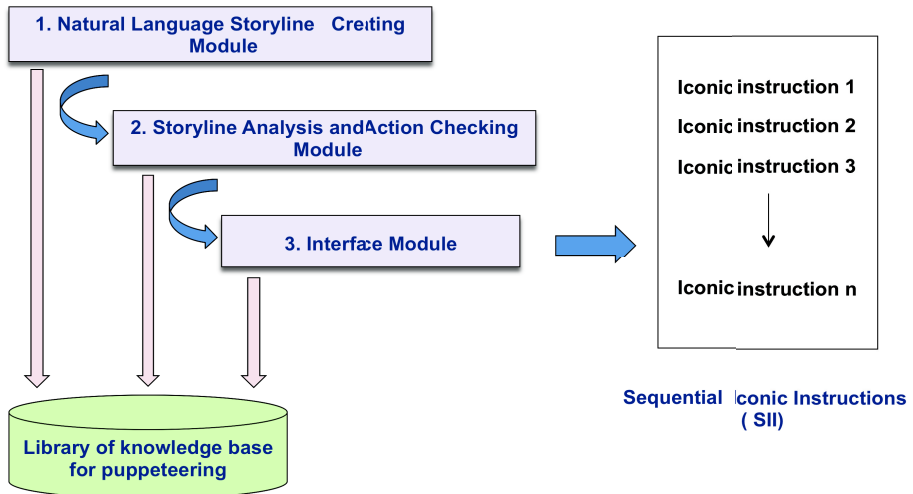


Fig. 2. The overall process of the intelligent instructional tool for puppeteering

4.1 Natural Language Storyline Creating Module

For experienced users, storylines can be created directly using natural language based on the semantic of the storyline sentences as shown in Fig. 3, and for novice and less experienced users, storylines can be created by using a software interface as shown Fig. 4. In the latter, users create the storyline by simply entering the data line by line on the left side of the interface. The storyline will show up on the right side of the interface. After pressing the “Save-F8” button, the system will create the text file of the storyline automatically which is similar to the one shown in Fig. 3.

The semantic of the storyline sentence is designed based on the knowledge base developed by conducting interviews on professional puppeteers and observing them during performance. Each line of the storyline consists of the puppet name and a sentence. The puppet name is separated from the sentence by the symbol “:”. There are two types of sentence. The first type is the action sentence which consists of two parts - the action part which is enclosed within the symbols “[“ and “]”, and the position part which is enclosed within the symbols “{“ and “}” as shown in Fig. 5. The second type is the storytelling sentence which consists of also two parts – the action part which is contained within the symbols “[“ and “]”, and the speech part which is enclosed within the symbols””” and “”” as shown in Fig.6.

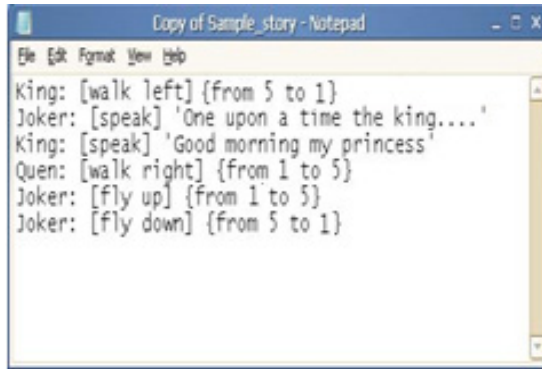


Fig. 3. Creating storyline using natural language (for experienced users)

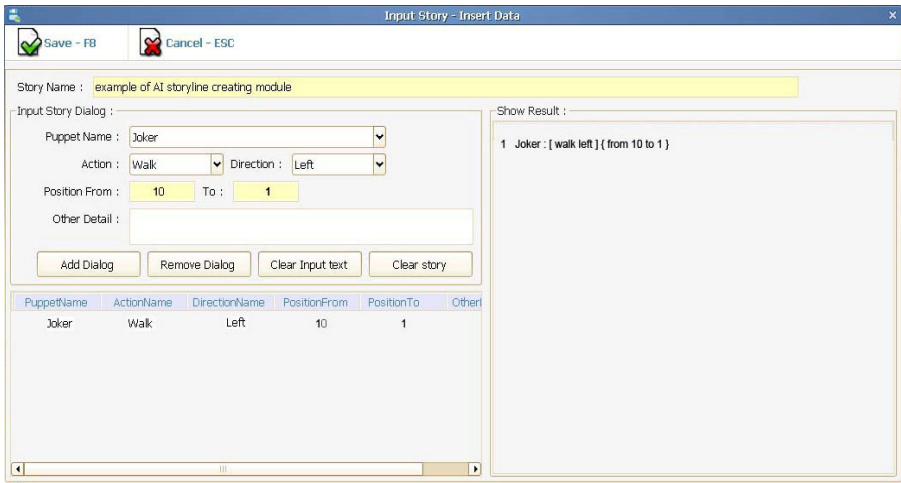


Fig. 4. A software interface for creating storyline (for novice and less experienced users)

Queen : [walk right]{from 1 to 5}

}
}
}

Puppet's name
Action
Position

Fig. 5. An action sentence

King : [speak] 'Good morning my princess'

}
}
}

Puppet's name
Action
Speech

Fig. 6. A storytelling sentence

4.2 Storyline Analysis and Action Checking Module

After the storyline is created, the Storyline Analysis and Checking Module will analyze line by line all the lines of the storyline, and separate them into two groups – storyline sentence and action sentence as described previously (See Fig. 7). For a storytelling sentence, it is analyzed immediately by the storyline analysis module by separating it into three separate information namely puppet’s name, action and speech. Then, they are stored in a table called “sequential instructions table”. The action sentence has to be processed further in order to ensure that only correct action is put in the table. Incorrect action sentence has to be corrected by the user and further checked by the action checking module. It is then analyzed by the storyline analysis module by separating it into three separate information namely puppet’s name, action and position.

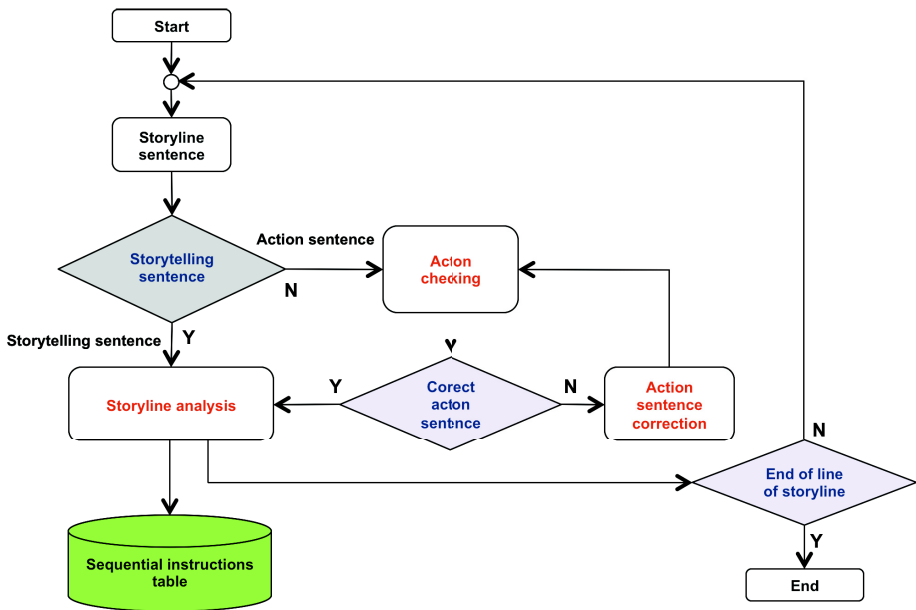


Fig. 7. Storyline Analysis and Action Checking Module

4.3 Interface Module

In Interface Module, all the processed storyline sentences in the sequential instructions table are displayed in Interface Module (Fig. 8(a)) and the table is translated into SII (Sequential Iconic Instructions) (Fig. 8(b)). SII will eventually be used to guide users to control the puppets according to the storyline. SII as shown in Fig. 9 is able to help the puppeteer to perform shadow puppet play and learn the skill needed in order to become a skilled puppeteer.

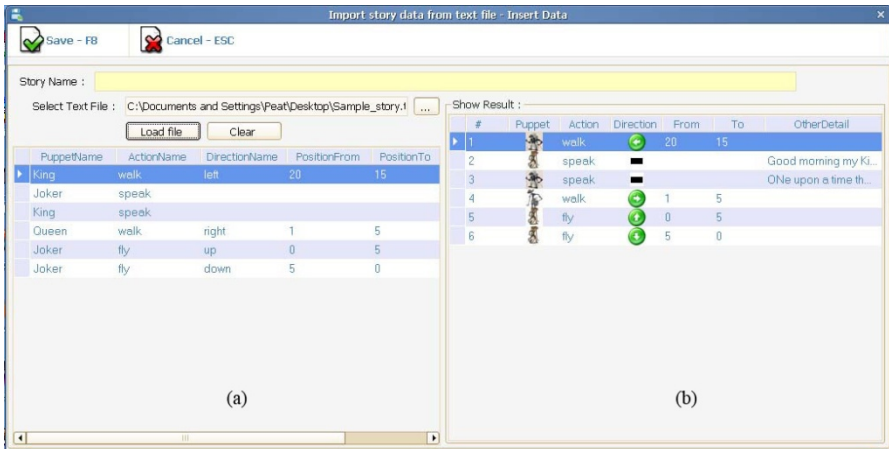


Fig. 8. Interface Module (a) Sequential instructions table, (b) Sequential Iconic Instructions (SII)



Fig. 9. Sequential Iconic Instructions (SII) as a guide in virtual shadow puppet play

5 Preliminary Evaluation and Discussion

In preliminary evaluation of the intelligent instructional tool, a questionnaire has been designed to obtain feedbacks and comments from two parties consisting of professional and experienced puppeteers, and general users. The respondents are required to rate their satisfaction based on a scale of 1 to 7 where 1 is the lowest and 7 is the highest in all of the questions. For the first part of the questionnaire which is on the use of the tool for creating storyline, 93% of the respondents have given a high

rating (6 or 7). It shows that our storyline creating interface helps users in creating storyline more beneficially. In the second part, we focused on the use of the storyline analysis and action checking technique. For this part, feedbacks and comments are only obtained from the puppeteers. Majority of the respondents that is 84% have given ratings of 6 or 7 which means that they are very satisfied with this tool. For some respondents, the outcome is not to their satisfaction. This may be due to the content of the knowledge bases for puppeteering. The last part of the questionnaire which focuses on the usefulness of Sequential Iconic Instructions (SII), 96% of general users have given a rating of 7 which shows that SII is very useful in guiding them to play the puppets according to the created storyline. 55% of the respondents from among the puppeteers have given a low rating on SII which means that for highly experienced puppeteers, SII is not very important and beneficial to them.

6 Conclusion and Future Work

In our work, we have provided a general architecture and an initial idea on the development of an intelligent tool that makes the users become a good virtual puppeteer. We have explored and investigated the possibility of developing virtual environment for puppeteering of traditional shadow puppet play that integrates the elements of the traditional shadow play with a virtual environment. The system includes an intelligent instructional tool that gives suggestions for puppeteer which is the main focus of this paper. The result of the preliminary evaluation has shown that the tool helps users more beneficially and a high degree of satisfaction among both the professional puppeteers and the general users.

For our future work, we are looking forward to improving SII such as by incorporating elements of animation in SII. We also aim to gather more knowledge for the library of knowledge bases for puppeteering. It is also hope that the action checking procedure in Storyline Analysis and Action Checking Module will be able to correct any incorrect action sentences automatically. Other future work includes the development of a complete system for the virtual shadow puppet play. The system is expected to provide a new form of presentation tool and storytelling tool to the younger generations for better appreciation of the traditional art, and promote and preserve the art of traditional shadow play since it provides wider access to the people. Besides, localization of the software tool is also required in order to reach various communities in many different countries.

References

1. Matsuki, P.: Malaysian Shadow Play and Music - Continuity of an Oral Tradition. South-east Asian Social Science Monographs. Oxford University Press, Kuala Lumpur (1977)
2. Chat Chai, S.: The folk plays of Southern Part. Cultural Center of Rajabhat Nakhon Si Thammarat University, p. 14 (2000)
3. Silva, A., Vala, M., Paiva, A.: Papous: The Virtual Storyteller. In: de Antonio, A., Aylett, R.S., Ballin, D. (eds.) IVA 2001. LNCS (LNAI), vol. 2190, pp. 171–180. Springer, Heidelberg (2001)

4. Biedermann, M., Geimer, M., Langs, A., Ritschel, T., Trappe, R., Muller, S.: Virtual Puppet: A physically based, real-time shading experience, Computer Graphics Group Institute for Computational Visualistics, University of Koblenz-Landau (2006)
5. Theune, M., Faas, S., Nijholt, A., Heylen, D.: The Virtual Storyteller. ACM SIGGROUP Bulletin 23(2), 20–21 (2002)
6. Theune, M., Faas, S., Nijholt, A., Heylen, D.: The Virtual Storyteller: Story Creation by Intelligent Agents. In: Proc. the Technologies for Interactive Digital Storytelling and Entertainment Conference, pp. 204–215 (2003)
7. Ma, H., Mc Kevitt, P.: Building Character Animation for Intelligent Storytelling with the H-Anim Standard. In: Proc. of Eurographics, Ireland, pp. 9–15 (2003)
8. Cavazza, M., Martin, O., Charles, F., Mead, S.J., Marichal, X.: Users Acting in Mixed Reality Interactive Storytelling. In: Balet, O., Subsol, G., Torguet, P. (eds.) ICVS 2003. LNCS, vol. 2897, pp. 189–197. Springer, Heidelberg (2003)
9. Hsu, S.-W., Li, T.-Y.: Planning Character Motions for Shadow Play Animations. In: Proc. International Conference on Computer Animation and Social Agents (CASA 2005), pp. 184–190 (2005)
10. Zhu, Y.B., Lee, C.J., Shen, I.F., Ma, K.L., Stoppel, A.: A New Form of Traditional Art: Visual Simulation of Chinese Shadow Play. In: International Conference on Computer Graphics and Interactive Techniques Sketches and Applications, p. 1 (2003)
11. Kim, J.C.M., Talib, A.Z.: A Framework for Virtual Storytelling Using Traditional Shadow Play. In: Proc. International Conference on Computing and Informatics (ICOCI 2006), pp. 1–6 (2006)
12. Rahman, K.A.: Wayang “Virtual” Integration of Computer Media in Traditional Wayang Kulit (Shadow Play)
<http://www.itaucultural.org.br/invencao/papers/Rahman.html>
13. Lam, T.K., Talib, A.Z., Osman, M.A.: Real-Time Visual Simulation and Interactive Animation of Shadow Play Puppets Using OpenGL. Int. J. Comp. & Inf. Eng. 4(1), 52–58 (2010)
14. Lam, T.K., Talib, A.Z.: Shadow Image and Special Effects Implementation Techniques for Virtual Shadow Puppet Play. In: Mastorakis, N.E., Mladenov, V. (eds.) Proc. 3rd WSEAS International Conference on Visualization, Imaging and Simulation, Advances in Visualization, Imaging and Visualization, pp. 80–81. WSEAS Press (2010)

A Tabletop Board Game Interface for Multi-user Interaction with a Storytelling System

Thijs Alofs¹, Mariët Theune¹, and Ivo Swartjes^{2,*}

¹ University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands
t.alofs@gmail.com, m.theune@utwente.nl

² Ranj Serious Games, Lloydstraat 21m, 3024 EA Rotterdam, The Netherlands
ivo@ranj.nl

Abstract. *The Interactive Storyteller* is an interactive storytelling system with a multi-user tabletop interface. Our goal was to design a generic framework combining emergent narrative, where stories emerge from the actions of autonomous intelligent agents, with the social aspects of traditional board games. As a visual representation of the story world, a map is displayed on a multi-touch table. Users can interact with the story by touching an interface on the table surface with their fingers and by moving tangible objects that represent the characters. This type of interface, where multiple users are gathered around a table with equal access to the characters and the story world, offers a more social setting for interaction than most existing interfaces for AI-based interactive storytelling.

1 Introduction

In this paper we present a generic tabletop board game interface for interactive storytelling. The setting of a tabletop board game stimulates face-to-face contact and social behaviour, which we think is important in multi-user interactive storytelling. The idea of using tabletop interfaces for digital storytelling is not entirely new. Several tabletop interfaces exist for storytelling systems [1,2,3], but they only focus on facilitating storytelling, trying to stimulate collaboration and creativity. Unlike our system, they do not use AI to contribute to the story.

The interfaces of current AI-based storytelling systems are mostly like those of computer games, where a single user interacts from a first person perspective with 2D or 3D virtual characters on a computer screen [4,5,6]. A few systems focus on collaborative storytelling by multiple users, but these also have interfaces similar to computer games [7,8]. To our knowledge, our *Interactive Storyteller* is the first AI-based storytelling system with a table-top interface.

The idea behind *The Interactive Storyteller* is to let users control the actions of one or more of the characters in a simulated story world. Each user can play one character, but collaborative control is also possible, with users co-operating to make decisions for the characters. From the ongoing interaction

* The work was carried out while the third author was working as a postdoctoral researcher at the Human Media Interaction group of the University of Twente.

between the characters (which can be either player or computer controlled) and through their choices of actions a story emerges; this approach to interactive storytelling is called ‘emergent narrative’ [4]. The technical framework underlying *The Interactive Storyteller* is a multi-agent system for story generation, in which intelligent agents act out the role of characters in the story. These agents can plan and execute sequences of actions to satisfy their character’s goals, taking into account the current state of the story world, and the mental state of the character. For more information on the storytelling framework, see [9].

By using a multi-touch table we aim to achieve interactive storytelling that resembles the social setting of a tabletop board game. Like existing digital tabletop board games [10,11,12], we try to combine the dynamics and intelligence of computer games with the social advantages of traditional board games. To reinforce the resemblance with board games, we investigate the use of tangible playing pieces that represent characters for physical interaction.

Next, we discuss the interface design and design choices based on related work and preliminary user tests. We end with a discussion on future work.

2 Interface Design

The interface design of *The Interactive Storyteller* is meant to be generic and easily adaptable to different story domains. We use a multi-touch table based on infrared reflection that is capable of identifying tangible objects through fiducial markers. The MT4j framework¹ is used for multi-touch support.

Story World and View. The visual representation of the story world is presented to users and possible spectators on a shared visual surface. Just as with traditional board games, it is important that people on all sides of the table have a similar view on the story world, therefore we chose a top-down map view.

Two story domains are currently available in our storytelling framework: a domain about pirates, and a domain based on the “Little Red Riding Hood” (LRRH) story [13]. We used the latter domain in the prototype, because we consider it to be more coherent, easier to understand for new users, and easier to visualise (see the screenshot in Fig. 1). Given that LRRH is a children’s fairytale, we decided to focus our research predominantly on children aged 6 to 11. However, a goal kept in mind was that the interface should also be appropriate for adults (with more adult domains and corresponding images).

The story world of LRRH contains three characters (Red, grandma, and the wolf) and five locations (Red’s house, grandma’s house, the clearing in the forest, the lake, and the beach). These locations are marked by blue circles on the map. Typical actions for a character currently available in the LRRH domain are amongst others: walking, greeting someone, stealing things, crying, eating something, baking a cake, and poisoning food. Characters can plan a series of such actions to try to achieve a goal they have. For instance, for the goal of

¹ Multi-touch for Java, <http://www.mt4j.org/>

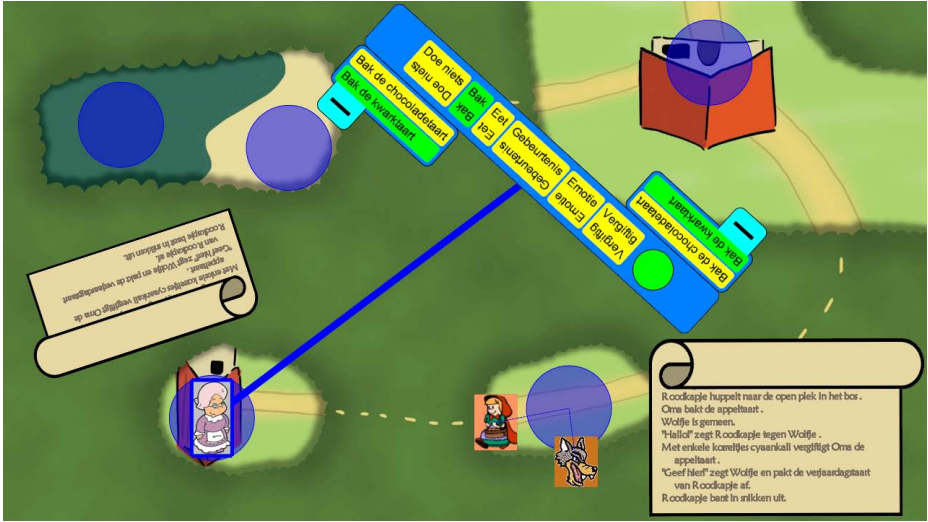


Fig. 1. Screenshot of interface design. On the actual tabletop the character images are covered by the tangible objects representing the characters.

‘Red’ wanting to poison the wolf, a possible series of events might be: Grandma bakes a cake, Red poisons it with cyanide because she expects the wolf to steal it, which he does, the wolf eats the cake and dies.

For aesthetic reasons, we decided to draw the characters and houses on the LRRH map not strictly from their top-side view, but from a more recognisable angle. However, to prevent one side of the table from being optimal for perceiving the story world, characters are displayed in one direction and houses are projected the other way (see Fig. II). Another solution for this orientation issue is autorotation, as was used in KnightMage [10]. Autorotation is only possible when there is always just one user that controls each character and the position of this user at the table is known, both of which do not apply to our system.

To enable an existing story world to be used in *The Interactive Storyteller*, the only required additions are a map and pictures that represent the characters in the world. The coordinates of the locations on the map, which link the story world to its visual representation, have to be provided in a properties file.

Physical Interaction. Like in the systems KnightMage [10] and False Prophets [11], users can change the locations of characters by moving physical toys that represent the story characters across the surface of the multi-touch table. These tangibles provide tactile interaction that is expected to be intuitive because it very much resembles the interaction offered by many familiar board games.

In our storytelling system locations are always discrete: characters are at one location, or the next, but never half-way in between. When a user moves a tangible to an adjacent location, the blue circle of the destination location turns green to indicate that this is an allowed action. When the user moves the tangible

to a location that is not in direct reach of the character's current location, the circle of the destination turns red to indicate that this is not an allowed action.

Users might put tangibles outside the circles that mark locations on the map. The system is unable to physically move tangibles away from such non-locations. Interventions, like the system asking to move a tangible to a particular location, are not used because they distract the user from the story. This means the system has to be able to deal with tangibles being anywhere on the map. When a tangible is not at its character's actual location in the story world, a thin dotted blue line is shown connecting the tangible to its character's location.

Interface Elements. For the selection of non-move actions by users, there is an Action Selection Interface (ASI). An important requirement we had for this interface was that it should be quickly usable for multiple storytelling domains. A very specific and intuitive ASI can be developed by focussing on one particular domain to fit the users' needs in that particular virtual world. This is usually done in computer games, but we consider it more important that the interface stays generic. Therefore, we decided to refrain from icon-based or other graphics-based ways for action selection, and use a flexible text-based approach.

Every time the turn goes to a new character, the knowledge base retrieves the set of all possible actions for that character, at that location, at that time, that are available in the story world. The ASI can be seen in the centre of Fig. 1. Because of the young target user group, all text is displayed in our native language Dutch. The user first selects a category in the centre bar of the ASI and then an action within that category. After that, the round confirmation button in the centre bar is enabled and can be used to confirm the selected action and pass the turn to the next character.

Users and spectators can read the results of actions that characters perform in the story areas, which are the two scrolls that can be seen on the screenshot in Fig. 1. If users want more or fewer lines of text to be visible in a story area, this can be achieved by touching the end of the scroll and rolling it up or down.

The ASI and the story areas occlude the view of the part of the map behind them. A balance has to be found between good visibility of the map and its contents, and the readability of the ASI and story areas. To achieve this subjective balance, we decided to keep the user in control of the size and placement of the ASI and the story areas. The user can move these elements around by dragging them with a finger. By dragging with two fingers at the same time, it is possible to rotate and resize them. The user can choose to find a static arrangement that generally works well in a particular story world, or keep changing sizes and arrangements depending on the current state of the story and places of interest.

Because we consider it to be important that users or spectators from all sides of the table have an equal view, all text in the ASI is presented in two directions instead of one. When users or spectators are standing on all four sides of the multi-touch table, the optimal layout is to position the ASI under an angle of 45 degrees with the sides of the table. We expect this angle to be acceptable for most readers. Having a shared ASI saves much space compared to having separate control areas on all four sides of the tabletop for different users, as in

the SIDES system [12]. Moreover, if everybody is standing on one side of the tabletop, the text at the opposite side of the ASI can be hidden by touching the minus symbol on that side, conserving even more valuable screen estate.

Preliminary User Tests. We performed some informal user tests with five test subjects (three boys aged 8, 8 and 10; and two girls aged 10 and 11) interacting with an early prototype of the system. We found that despite the limited graphics, the children were very engaged by the system and enjoyed playing with it. With only a very limited explanation they understood how to interact with the system. Several children discussed possible actions together and some even planned a sequence of actions to pursue a particular storyline.

Based on observations, several improvements to the system were made. For example, we discovered that users often lost track of turn-taking. To make clear to which character the actions in the ASI belong, the ASI is now connected by a solid blue line to the character that has the turn.

Another observation was that fingertips were often badly recognised because they were very small. At the same time the rest of the hand did get recognised while hovering above the surface. By fine-tuning some recognition parameters we managed to reduce these issues, but the used hardware is a limiting factor.

Although told to do so, not every child looked at one of the story areas to read the results of actions performed by other characters. This often resulted in these users ending up confused and less immersed in the story. To address this issue, we decided to offer the same information in another modality by vocalising it with Loquendo text-to-speech. The story areas are kept as a time-independent source of the same information about the story. After introducing the new modality, we decided to also allow the addition of action specific sounds. Although domain specific, associating actions with sounds is a quick and easy way to present audible feedback of an action to enrich the user experience.

3 Discussion and Future Work

Because our interface combines tabletop interaction with the advantages of computer- and board games, we expect the system to be a suitable interface for interactive storytelling. The system facilitates group play as opposed to solitary game play. Our next step will be a formal user test to evaluate whether *The Interactive Storyteller* provides a suitable interface for interactive storytelling for children. Because questionnaires are not very suitable for young children, we intend to use an observation scheme like the Play Observation Scale [14].

We also intend to investigate whether the use of tangibles in our system setup has advantages over a touch-only approach. To answer this question, we made a touch-only version of our prototype which only uses images to represent and move characters. Both versions will be compared in the user tests. These tests should also show whether users prefer to play one character each, like actors playing a role, or to choose the actions for all characters together in deliberation.

The multi-touch table used in this research makes use of an ordinary video projector, non-diffuse IR-beams, and an average webcam. The lack of precision of

the used setup irritated users in the preliminary user test. In the future we would like to test the system on a high-end multi-touch table or to port *The Interactive Storyteller* to the new generation tablet PC's. These devices live up to the high expectations and increasing demands of present-day users and provide more accuracy for new directions in multi-touch research. One of the research challenges to be addressed is allowing the possibility to add engaging visual elements (e.g., object inventories, animations) while keeping the framework generic.

References

1. Alves, A., Lopes, R., Matos, P., Velho, L., Silva, D.: Reactoon: Storytelling in a Tangible Environment. In: 3rd IEEE International Conference on Digital Game and Intelligent Toy Enhanced Learning, pp. 161–165 (2010)
2. Cappelletti, A., Gelmini, G., Pianesi, F., Rossi, F., Zancanaro, M.: Enforcing Cooperative Storytelling: First Studies. In: 4th IEEE International Conference on Advanced Learning Technologies, pp. 281–285 (2004)
3. Helmes, J., Cao, X., Lindley, S., Sellen, A.: Developing the Story: Designing an Interactive Storytelling Application. In: ACM International Conference on Interactive Tabletops and Surfaces, pp. 49–52 (2009)
4. Aylett, R.S., Louchart, S., Dias, J., Paiva, A., Vala, M.: FearNot! - An Experiment in Emergent Narrative. In: Panayiotopoulos, T., Gratch, J., Aylett, R.S., Ballin, D., Olivier, P., Rist, T. (eds.) IVA 2005. LNCS (LNAI), vol. 3661, pp. 305–316. Springer, Heidelberg (2005)
5. Mateas, M., Stern, A.: Façade: An Experiment in Building a Fully-Realized Interactive Drama. In: Game Developers Conference: Game Design Track (2003)
6. Pizzi, D., Cavazza, M.: Affective Storytelling based on Characters' Feelings. In: Intelligent Narrative Technologies: Papers from the AAAI Fall Symposium, pp. 111–118 (2007)
7. Kriegel, M., Lim, M., Aylett, R., Leichtenstern, K., Hall, L., Rizzo, P.: A Case Study In Multimodal Interaction Design For Autonomous Game Characters. In: 3rd Workshop on Multimodal Output Generation, pp. 15–25 (2010)
8. Prada, R., Paiva, A., Machado, I., Gouveia, C.: You Cannot Use My Broom! I'm the Witch, You're The Prince": Collaboration in a Virtual Dramatic Game. In: Cerri, S.A., Gouardères, G., Paraguaçu, F. (eds.) ITS 2002. LNCS, vol. 2363, pp. 913–922. Springer, Heidelberg (2002)
9. Swartjes, I., Theune, M.: The Virtual Storyteller: Story Generation by Simulation. In: 20th Belgian-Netherlands Conference on Artificial Intelligence, pp. 257–265 (2008)
10. Magerkurth, C., Memisoglu, M., Engelke, T., Streitz, N.: Towards the Next Generation of Tabletop Gaming Experiences. In: Graphics Interface 2004, pp. 73–80 (2004)
11. Mandryk, R., Maranan, D.: False Prophets: Exploring Hybrid Board/Video Games. In: Conference on Human Factors in Computing Systems, pp. 640–641 (2002)
12. Piper, A., O'Brien, E., Morris, M., Winograd, T.: SIDES: A Cooperative Tabletop Computer Game for Social Skills Development. In: 20th Anniversary Conference on Computer Supported Cooperative Work, pp. 1–10 (2006)
13. Swartjes, I., Theune, M.: Iterative Authoring Using Story Generation Feedback: Debugging or Co-creation? In: Iurgel, I.A., Zagalo, N., Petta, P. (eds.) ICIDS 2009. LNCS, vol. 5915, pp. 62–73. Springer, Heidelberg (2009)
14. Rubin, K.: Play Observation Scale. *Child Development* 53(3), 651–657 (1982)

Design of an Interactive Playground Based on Traditional Children’s Play

Daniel Tetteroo, Dennis Reidsma, Betsy van Dijk, and Anton Nijholt*

University of Twente, Department of Human Media Interaction
d.tetteroo@utwente.nl, {d.reidsma,e.m.a.g.vandijk,a.nijholt}@utwente.nl

Abstract. This paper presents a novel method for interactive playground design, based on traditional children’s play. This method combines the rich interaction possibilities of computer games with the physical and open-ended aspects of traditional children’s games. The method is explored by the development of a prototype interactive playground, which has been implemented and evaluated over two iterations.

1 Introduction

Many governments pursue health care programs that promote a healthier youth. Still, children are unlikely to give up computer time in favor of outdoor play [8]. A solution may be the development of interactive playgrounds, consisting of “one or more interactive objects that use advanced technology to react to the interaction with children and actively encourage them to play” [17]. They possess the rich interaction possibilities from computer games as well as the physical and social aspects of traditional outdoor play. Being an *environment for play*, rather than a *game*, they do not force strict rules upon the players, but provide possibilities for the children to define their own games. This form of “open-ended” play benefits children in many aspects of their development, such as social skills, problem solving, creativity, and a better understanding of the physical world [10,3].

This paper presents a design method for developing interactive playgrounds, based on elements of traditional children’s play, that should actively stimulate the development of children’s physical, social and creative skills, without forcing game specific rules upon them, by combining the open-endedness of traditional outdoor play and the interactivity of modern computer games.

2 Related Work

Flash-Poles [17,2] are interactive poles, placed on a fixed position on a field. User tests indicated that they were successful in stimulating both cooperative and competitive physical play amongst children. The authors touch on a paradoxical issue in the design of installations for open-ended play. The interactive behaviour

* This research has been supported by the GATE project, funded by the Dutch Organization for Scientific Research (NWO) and the Dutch ICT Regie.



Fig. 1. The interactive slide (left) and the interactive pathway (right)

of the objects needs to be understandable for the children, but too many rules limit the possibilities for open-ended play. The *Ledball* [17] was developed to investigate the creation of new games by children using *mobile* play objects, but has not yet been evaluated.

The *Playware* playground tiles contain sensors, LED's and/or loudspeakers, plus processing capabilities to allow one to configure multiple tiles into different games [9]. Its evaluation focused on clear-defined games with fixed rules.

In the *Interactive Slide* (see Figure 1), a display is projected on a large, inflatable slide [15]. The projection area is monitored by an IR camera, which allows for interactivity between the slide's users and the projection. Using an interaction driven design strategy, two games have been developed which aim at encouraging physical activity amongst children.

Space Explorers are a new category of animated playground props that allow children to explore the space around them in a playful manner [13]. The prototype consists of a ball which moves around autonomously in a space, while interacting with the children present in the space. Children keep interacting with the ball as it moves around the space, such that they gradually discover the space around them. This shows there is potential for such playground props as mediating objects between a child and its play setting.

The *Interactive Pathway* (see Figure 1) is a rail-way like construction consisting of two wooden beams with a series of narrow pressure-sensitive mats connecting them [14]. On the wooden beams, "spinning tops" were placed, hand crafted by the children themselves. When a child steps on a mat, a motor causes one of the spinning tops to rotate. Evaluation showed that these quite simple interactions led to a wide range of open-ended play behaviour from the children.

3 Traditional Children's Play

Our goal was to design an intelligent, interactive playground, based on elements of *traditional children's (outdoor) play*. In this section we will discuss which dimensions of traditional play can be relevant in the design process. The project focuses on children in the age group between 8 and 12 years old. These children are able to perform advanced physical activities, define games rules and socially interact with each other [5, p.226] [12, p.251-252]. Furthermore, they are capable of actively taking part in evaluations involving group discussions and interviews.

3.1 Play, Games, and Playgrounds

Huizinga, in the book *Homo Ludens*, argues that it is almost impossible to capture the properties of *play* in a single definition. Because the act of play does not necessarily have a goal and is by definition not bound to rules, almost any act could be considered an act of play [7]. Games are a more formalized and strict form of play. “*The game has a beginning, a middle, and a quantifiable outcome at the end. The game takes place in a precisely defined physical and temporal space of play. Either the children are playing Tic-Tac-Toe or they are not*” [11, p. 25]. Open-ended play can evolve into a game over time, as players try to enforce certain play behaviour by defining rules and limitations.



Fig. 2. Examples of traditional children’s games: running the hoop, playing with marbles and the skip rope. [4, pgs. 33/3/36]

Children’s games often require few attributes, consisting of little more than some simple toys, the players themselves and their creativity [16]. They often require physical activity and can be played with basic materials which can be carried along (hoop, ball, stick). Finally, the rules of such games are often few and simple, adaptable by the players themselves. Whether a stick is to be used for hitting, drawing or waving is up to the children inventing their game. The simple rules make the games very accessible and easy to understand for anyone eager to join a game, and make it easy to adapt a game by replacing, removing or adding a rule. It also makes it easy to convey these games from generation to generation. Sometimes, simple games – made-up by children themselves – can even have a higher appeal to them than more complex and predefined games [3].

Although play is potentially always and everywhere possible [20], most societies know the concept of *playground* as an environment specifically designated for play. The space may contain one or more playground artefacts. Traditional playground artefacts do not offer feedback and do not actively interact with the user. For example, a slide is often a rigid wooden structure which just ‘sits’ there in a playground. However, its presence within the playground allows for a number of ways to interact with it (climbing up the slide, sliding downwards, hiding beneath the slide), and children, when incorporating the artefact in their play, may assign any meaning to it when it fits their current play.

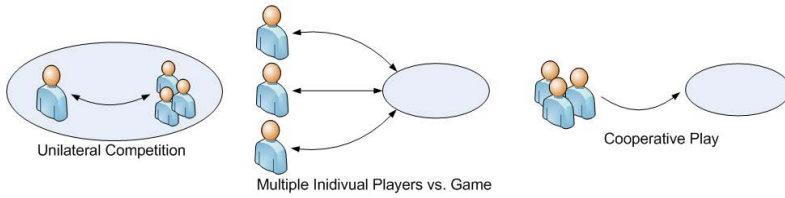


Fig. 3. A subset of the player interaction patterns from Fullerton et al. [6, p.46].

3.2 A Taxonomy for Playground Play

In order to design our interactive playground using elements of traditional children’s play, we need a taxonomy to describe these elements in a structured way. For this, we draw from related work in interaction design, we introduce the idea of *Gamespace* in playgrounds, and finally elaborate our taxonomy based on an analysis of many types of traditional playground play.

Elements from Related Work in Interaction Design. Sturm et al. address five key issues for the successful design of an intelligent, interactive playground of open-ended play: *social interaction, simplicity, challenge, goals* and *feedback* [17]. Soute et al. [16] discuss that one should address *social interaction, fun, physical activity, and flexible and adaptable rules* in order to combine the appeal of indoor digital games with the benefits of traditional outdoor play in Head Up Games. Fullerton et al. [6, p.46], finally, defined seven different player interaction patterns (see Figure 3). Social interaction between the players could be enhanced by facilitating competitive or cooperative play.

Gamespace. We introduce *Gamespace* as a one-dimensional measure defining to which degree an act of play, or a game, is related to the play-environment or playground in which it occurs. We define gamespace on a sliding scale on which three global levels can be defined (see Figure 4):

- *Fully external*: the playground is irrelevant to the game, except as the location where it takes place. For example, children throw a ball back and forth without looking at, or making use of what is in the playground.
- *Partially contained*: the game is not mainly dependent on the playground, but incorporates elements of the playground. For example, the children play cops and robbers, and use the climbing frame as the robbers’ home base.
- *Fully contained*: the game takes place entirely within the playground. For example, the children are swinging or using the seesaw.

We aim to merge traditional children’s play and modern computer gaming into an interactive playground. Ideally, a passer-by would see children playing together, and only with a closer look would discover that the playground is technologically enhanced. In a similar fashion, the children playing on the playground would incorporate elements of the interactive playground in their play, but would not let the digital enhancements overshadow their play. Ideally, the games played

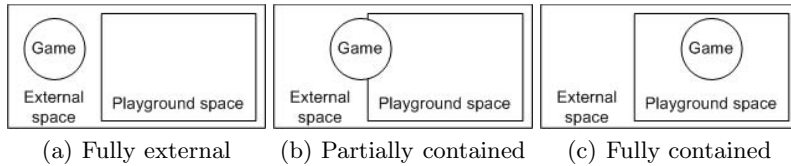


Fig. 4. Gamespace levels

within the playground would be *partially contained*. It should be stimulated and facilitated by the playground, but not be fully dependent on it.

A Design Taxonomy for Playgrounds. Based on the key issues described above, we define our taxonomy in three layers (see Figure 5). The **top level** is made up of three highly abstract ‘classes’, that also define global goals for the playground: (1) Gamespace, (2) social interaction, and (3) physical activity. We want our interactive playground to encourage physical activity and social interaction, and we want the resulting children’s play to be partially contained in the playground’s gamespace. We argue that the other key issues can be handled by carefully planning the social and physical dimensions of a playground. For example, by providing possibilities for competition (which is a form of social interaction), both a challenge and a possible goal (being ‘better’ than other players) are created for the players. The **bottom level** of our taxonomy is made up of a set of single interaction patterns between children and the (enhanced) playground. An example of a single interaction is the playground responding to shouting children by changing the colour of the playground’s surface. The distance between the interactions and the classes is very large, so it is very difficult to derive the one from the other without resorting to an **intermediate level**. To bridge the gap, we have analysed a large number of traditional outdoor children’s games. By analysing these games along the aforementioned classes and by comparing and weighing their individual properties, we constructed a set of 20 *dimensions*. Some example dimensions are *competition*, *collaboration* and *item possession*. The list of dimensions can be found in Appendix A; more extensive details on the dimensions can be found in [19].

4 A Novel Approach to Interactive Playground Design

Although a lot is known about key issues for interactive playgrounds through evaluation studies of various prototypes (cf. the sections above), a fully structured method leading from a playground concept to interactions for a working prototype is still lacking. This section aims to fill this gap by introducing a novel approach for interactive playground design that results in a tight integration of traditional children’s play and modern computer gaming. This section introduces the four-stage design process. In the next section we will describe a case study that we performed using this method.

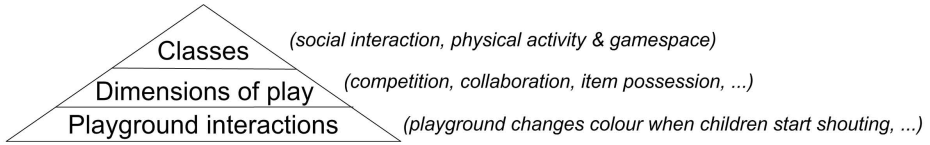


Fig. 5. Playground interaction design levels

Concept Generation. The method starts with *Concept Generation*. This phase aims to arrive at an overall ‘story concept’ that will drive the design, using the *classes* defined in the previous section as guidelines. First, a number of candidate ‘story concepts’ are described. Out of these, one concept is selected for further development, choosing on the basis of suitability for open-ended play: the concept must be concrete enough to be able to derive possible interactions from it, but it should not be so concrete as to block the children from evolving their own play in the playground. For example, a story concept might center on “make a playground that is like a giant complex machine with moving parts”, or “make a playground inhabited by many creatures”.

Interactions Generation. The story concept sketches the rough contours along which we can design the playground’s interactions. The second phase is to *develop single Interactions* that children can have with the playground. For each of the 20 *dimensions* determined earlier, a single interaction possibility is designed. An interaction may be related to more than one dimension, but at least we make sure that every dimension is related to at least one interaction. For example, an interaction, related to dimension 16 (see appendix), might be “If you step on a spinning gear in the machine, it will start making a noise”.

Systematic Variation on Interactions. The fourth step in the design process is one of *Systematic Variation*. In this phase, every interaction developed during the previous phase is analysed along all 20 dimensions. Wherever possible, a new interaction is derived by adding the dimension if it was not yet present in the interaction, or by inverting the role of the dimension, if it was. This phase adds yet more structure to the design and yields a vastly more extensive set of interactions, with a better coverage of the various dimensions, and therefore of the three abstract ‘classes’.

Selection. The final step in the design process is the *Selection* of interactions which will be implemented. Because the previous phase will tend to create numerous contradicting and opposing interactions, selection is not a trivial task. Criteria to guide the selection process, besides practical reasons of feasibility, are: (1) Which dimensions are covered by the selected interaction methods? (2) Do the chosen interactions form a balanced system? If there are, for example, mechanics for introducing new (virtual) objects into the playground, there should be mechanics for reducing their amount as well, to avoid clutter.

Initial situation:	Two or more players with a ball within the playground
Action performed:	Two players let their balls touch for a moment
System reaction IM2:	A new shape is created near both players
Rationale:	Introduces element of collaborative play
System reaction IM22:	Tails of both players switch owners
Rationale:	Competitive instead of collaborative play

Fig. 6. IM22 is a variation on IM2; collaboration becomes competition

5 Case Study

Using the design approach summarized above, we developed an interactive playground and evaluated it in a user study with 19 children.

Concept Generation. The initial story concept was as follows^[4]:

“The playground space is initially empty. When a player enters the playground, (s)he gets surrounded by a simple unique shape in an arbitrary colour. The shape follows him/her wherever (s)he goes. When two players get in touching distance, a smaller ‘offspring’ shape is created, based on the players’ shapes. The offspring get their own ‘life’ travelling around within the playground. Players are allowed to interact with the shapes through collecting, stealing, killing or moving offspring shapes. The shape ecosystem contains control mechanisms, such as a predator shape, to provide additional dynamics to the playground.”

Interaction Generation and Systematic Variation. Systematic variation of the initial set of 20 interaction methods (one for each dimension) resulted in 32 additional interactions. Full details can be found in [19]; for reasons of space we only mention one example of variation (Fig. 6). In the original interaction, physical contact between two players triggers the ‘birth’ of a new shape. The variation shows a different system response: the players’ tails switch owners, causing collaborative play to become competitive play.

Selection. Of the 52 interactions resulting from systematic variation, 13 were implemented in our playground. The playground interactions are centered around shapes that normally lead a life floating around the playground freely, but which can be captured by players who chase them. Captured shapes follow a player in a tail, but can also be stolen by other players that chase the tail for a while. Players can create new shapes by standing together, and destroy each others’ shapes by running through another player’s tail. Each player can, by shaking a ball (s)he carries, create a pool of poisonous venom that destroys other players’ shapes. All of these actions influence a player’s status, which is expressed by the size of a circle projected around them.

¹ cf. <http://hmi.ewi.utwente.nl/showcase/anemone>: emergent entertainment/

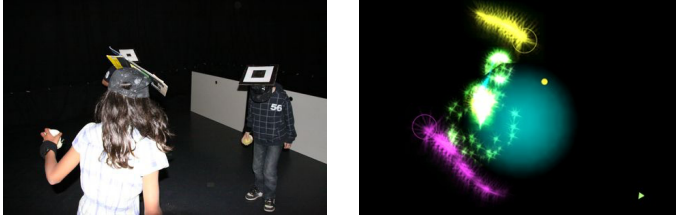


Fig. 7. The interactive playground

5.1 Implementation

The interactive playground was implemented in the SmartXP laboratory of the University of Twente, using top projection and an infrared camera over an area of 6x6 meters. The players' positions were tracked by the infrared camera, filming reflective markers mounted on the children's heads. Small foam balls were equipped with Sun SPOT sensors [18]. The sensors allowed us to detect when a ball is shaken and by whom (because the players were wearing another Sun SPOT sensor around their wrist).

5.2 Evaluation

The playground was evaluated to test the relation between the children's play and the implemented interactions. Children were invited to play in the playground during 30 minute sessions in 2-4 person groups (see Figure 7). A modified version of the OPOS observation scheme [1] was used to record observations along the three classes defined in section 3.2. We observed numerous games, such as throwing a ball and games of tag. Most games were strongly related to the children's presence in the playground. Examples are *catch-the-shapes*, *scare-the-monster* and *switch tails*. Most observed games were fully contained within the playground's *gamespace*; only a few were partially contained. This effect might be due to novelty of the playground, which incited children to focus on exploration of the playground's possibilities. Concerning (*social interaction & physical play*), there are strong hints that the implemented interactions influenced children's play behaviour in the intended way. Concluding interviews with the children indicated that they were very fond of the interactive playground and were willing to give up computer gaming time to play in the playground.

6 Conclusions and Recommendations

This paper presented the design, implementation and evaluation of an interactive, intelligent playground. A novel design approach has been used in which the design of the interactive playground is based on elements of traditional children's games. Ideally, a passer-by would see children playing together on a playground and only with a closer look, (s)he would discover that the playground is

technologically enhanced. In a similar fashion, the children playing on the playground should feel the same excitement they feel when playing computer games and only on second thought notice that they are participating in socio-physical play. A playground resulting from the approach should combine the merits of both traditional outdoor play and modern computer gaming.

An example playground was designed and implemented using the novel method, and was evaluated in a user study with 19 children. The children who participated in the evaluation showed great enthusiasm. Almost without exception, they were prepared to give up computer gaming time in exchange for spending time with the interactive playground. Considering the motivation of this project, this promises a hopeful future for interactive playgrounds. To further verify the validity of the proposed design method, both the current playground concept and alternative concepts should be tested in experiments focussing on long-term usage of the designed playground in a realistic setting.

References

1. Bakker, S., Markopoulos, P., de Kort, Y.: OPOS: an observation scheme for evaluating head-up play. In: NordiCHI 2008, pp. 33–42 (2008)
2. Bekker, M., van den Hoven, E., Peters, P., Hemmink, B.k.: Stimulating children's physical play through interactive games: two exploratory case studies. In: Proc. of IDC 2007, pp. 163–164. ACM, New York (2007)
3. Bekker, T., Sturm, J., Wesselink, R., Groenendaal, B., Eggen, B.: Interactive play objects and the effects of open-ended play on social interaction and fun. In: Proc. of ACE 2008, pp. 389–392. ACM, New York (2008)
4. Cornelisz, J.H.: Kinderspelen, in leerzame gedichtjes. Ten Brink en De Vries, Amsterdam (1837)
5. Del Alamo, M.R.: Design for Fun: Playgrounds. Links International (2004)
6. Fullerton, T., Swain, C., Hoffman, S.: Game design workshop: designing, prototyping, and playtesting games. Focal Press (2004)
7. Huizinga, J.: Homo ludens. Proeve eener bepaling van het spel-element der cultuur. H.D. Tjeenk Willink & Zoon (1950)
8. Jansen, W., Mackenbach, J.P., van Zwanenburg, E.J., Brug, J.: Weight status, energy-balance behaviours and intentions in 9-12-year-old inner-city children. *Journal of Human Nutrition and Dietetics* 23(1), 85–96 (2010)
9. Lund, H.H., Klitbo, T., Jessen, C.: Playware technology for physically activating play. *Artificial Life and Robotics* 9(4), 165–174 (2005)
10. Rogers, Y., Price, S.: Extending and augmenting scientific enquiry through pervasive learning environments. *Children, Youth and Env.* 14(2), 67–83 (2004)
11. Salen, K., Zimmerman, E.: Rules of play: game design fundamentals (2003)
12. Schenk-Danzinger, L.: Entwicklungspsychologie. BV (1977)
13. Seitinger, S.: Animated props for responsive playspaces. Masters's thesis in Media Arts and Sciences, Massachusetts Institute of Technology (2006)
14. Seitinger, S., Sylvan, E., Zuckerman, O., Popovic, M., Zuckerman, O.: A new playground experience: going digital? In: CHI 2006, pp. 303–308 (2006)
15. Soler-Adillon, J., Ferrer, J., Parés, N.: A novel approach to interactive playgrounds: the interactive slide project. In: Proc. of IDC 2009, pp. 131–139. ACM, New York (2009)

16. Soute, I., Markopoulos, P., Magielse, R.: Head up games: combining the best of both worlds by merging traditional and digital play. *Pers. and Ubiquit. Comput.* 14, 435–444 (2009)
17. Sturm, J., Bekker, T., Groenendaal, B., Wesselink, R., Eggen, B.: Key issues for the successful design of an intelligent, interactive playground. In: *Proc. of IDC 2008*, pp. 258–265. ACM, New York (2008)
18. Sun Microsystems Inc. Sun spot system: Turning vision into reality. Technical White Paper on Sun SPOTs (June 2005)
19. Tetteroo, D.: Design of an interactive playground based on traditional childrens play. Masters’s thesis in Human Media Interaction, University of Twente (2010)
20. Wigley, M.: *Constant’s New Babylon: The Hyper-architecture of Desire* (1998)

A Game dimensions

1. The dominant **player interaction pattern** [6] associated with the game.
2. **Physical skills**: the amount of physical activity involved in the game.
3. **Social skills**: the amount of social activity involved in the game.
4. **Creative skills**: the amount of creativity involved in the game.
5. **Tactical skills**: the extent to which tactics can be applied to the game.
6. **Is finite?**: whether the game is limited by some intrinsic rule or condition.
7. **Has goal?**: whether the game has a concrete and defined goal.
8. **Is competitive?**: whether competition plays a role in the game.
9. **Single / multiplayer**
10. **Amount of space required**
11. Whether **chasing** other players is a factor in the game.
12. **Player’s visibility is essential part of the game?**
13. Whether the game contains a **promotion / degradation mechanism**.
14. **Allows ‘game over’**: whether a player can lose or not win the game.
15. **Time limit**: whether the game is strictly limited in time.
16. **Sound**: whether sound plays (or can play) a determining role in the game.
17. **Physical contact between players required?**
18. **Requires extra (physical) items or resources**
19. **Shared / individual items**: whether ownership of the item(s) is shared.
20. **Item possession is a (sub)goal in itself?**

Designing a Museum Multi-touch Table for Children

Betsy van Dijk, Frans van der Sluis, and Anton Nijholt

Human Media Interaction, University of Twente
PO Box 217, 7500 AE Enschede, The Netherlands
{e.m.a.g.vandijk,f.vandersluis,a.nijholt}@utwente.nl

Abstract. Tangible user interfaces allow children to take advantage of their experience in the real world with multimodal human interactions when interacting with digital information. In this paper we describe a model for tangible user interfaces that focuses mainly on the user experience during interaction. This model is related to other models and used to design a multi-touch tabletop application for a museum. We report about our first experiences with this museum application.

Keywords: tangible user interfaces, multi-touch table, tabletop, information access, children.

1 Introduction

A few decades ago, Human-Computer Interaction was largely restricted to traditional graphical user interfaces on computers with rectangular screens and mouse and keyboard as input devices. In that time [1] proposed to make computing truly ubiquitous and invisible and they introduced tangible user interfaces as a way to make digital information tangible. In these interfaces physical controls of digital information play a key role. This allows people to take advantage of their experience in the real world with multimodal human interactions.

The theory of embodied cognition shows the salience of tangible interaction for children. It is the merging of cognition and action which allows to easily manipulate the world and offload cognition while doing so [2]. Hence, for children, who in general have less abstract reasoning skills, tangible interfaces are in particular useful. The smaller the gap between real-world manipulation and digital manipulation, the easier the access to the digital information becomes, especially for young children.

Tabletop environments have been shown to allow natural interaction using tangible interaction elements [3]. All kinds of physical objects in the environment can be equipped with unobtrusive sensors in such a way that the children interact with their tangible surroundings. For instance in the Navigational Blocks project [4], the use of physical objects to represent data queries allowed people to explore relationships between topics and retrieve information. The tangible objects help users, especially children, to learn through touch and direct manipulation of objects [4].

Within the European project PuppyIR access to information for children is central. Unfortunately the current tools for information access offered on the Internet are not adequate for children. Interfaces are typically created for adults, information retrieval methods are based on the perception of relevance by adults, and services are typically constructed based on the idea of the information needs of adults. Part of easing the access is through the use of intuitive interfaces. A tangible tabletop is, as explained, suitable to this aim.

The context for this paper is an educational museum that aims at a broad audience. Its main theme is the human and his relation with nature, culture, society, science and technology. This is the theme of the permanent exhibition of the museum as well. The museum functions as a test environment to alter the access to the information contained in the permanent exhibition.

This paper explores the possibility to enhance the access to information using tangible interfaces. Models for tangible interfaces are discussed in Section 2. A tabletop interface will be used to direct the visitors through the museum, adapted to the interests of the user. The development of this multi-touch tabletop interface is described in Section 3. Finally, Section 4 discusses preliminary experiences with the interface within the museum context.

2 A Children Specific Design Model for Tangible Interfaces

In the model we propose for PuppyIR, the design concepts and heuristics are grouped in four themes: (1) Physical and digital representations; (2) Actions and effects; (3) Exploration and collaboration; and (4) Engagement and fun. These themes are built upon other, related, models for tangible interaction. See Table 1 for an overview of the model and its related models. The model is tailored for the design of tangible interfaces for children.

2.1 Physical and Digital Representations

The theme *Physical and digital representations* refers to the appearance of the physical objects and the relation with the digital representation of the objects. Are representations naturally coupled? Are they meaningful and built on the user's experience? Do they invite them into interactions?

Part of this theme is related to what [5] refer to as expressive representation: the interrelation of physical and digital representations and to how users perceive them. Often hybrid representations combine physical and digital elements. [6] calls this semantic mappings: the mapping between the information carried in the physical objects and the digital aspects of the system. Young children (under seven) have difficulty relating physical manipulatives to other forms of representation. The ability to understand that one object can have multiple representations develops slowly.

To make an interface more suitable for young children, perceptual mappings can be exploited [6]. Various kinds of mappings between physical and digital space can be afforded by tangibles. The mapping between the perceptual properties of the physical

and digital aspects of the system can rely on perceptual affordances or designed affordances. Designs that rely on perceptual affordances allow even very young children to explore these mappings. Designed affordances are opportunities for actions that are created through mindful design of artificial objects and environments. These affordances may be meaningful for adults, but for children age appropriate perceptual, cognitive and motor abilities and limitations need to be considered.

Table 1. Models of Tangible Interfaces

Theme	Related concepts of Hornbecker & Buur [5]	Related concepts of Sharlin et al. [7]	Related themes of Antle [6]
Physical and digital Representations: The appearances	Perceived coupling Representational signific. Tailored representation (Inhabited space – partly)		Perceptual mappings Semantic mappings
Actions and effects: How tightly are they related?	Isomorph effects Externalization Haptic direct manipulation (Inhabited space – partly) (Configurable materials)	Successful spatial mappings Unify input and output space	Space for action Behavioural mappings Semantic mappings
Exploration and Collaboration: How are they facilitated?	Lightweight interaction Embodied constraint Multiple access points Non-fragmented visibility	Enable trial-and-error activity	Space for friends Semantic mappings
Engagement and Fun: Are presence, motivation and user experience positively affected?			

2.2 Actions and Effects

The theme *Actions and effects* refers to the relation between the manual actions of users and their effects. This can be characterized by the following concepts [5]:

- Haptic direct manipulation: can users grab, feel and move the interaction objects?
- Externalization: can users use the objects as props to act with or think and talk with or through? Are tangible interactions salient to the overall use process?

- Isomorph effects: how easy is it to understand the relation between the manual actions of users and their effects? For instance because they are close in time, visibly nearby or of the same shape.

For children the relation between manual actions and their effects becomes more complicated. Children's developing repertoire of physical actions and spatial abilities for direct system input and control can only be applied successfully if the design is based on an understanding of how and why children's actions in space relate to changes in cognitive and motoric development.

An example of the benefits of a close coupling between action and effect comes from the repeatedly connecting and disconnecting of Lego blocks to better understand how different configurations relate to stability of the construction. Children use epistemic actions to facilitate the understanding of how things work. Hence, direct physical interaction with the world, by means of bodily engagement with physical objects, facilitates active learning and is a key component of cognitive development in childhood. External scaffolding (aids that include interactions with other children, adults, or aspects of the environment) is often used when executing epistemic actions [6].

The relation between action and effect can also be looked at from a behavioral perspective: the mapping between the input behaviors and output effect of the physical and digital aspects of the system. This is discussed by [7] with regards to their spatial mappings. [7] showed two conclusions: Physical/digital mappings must be successful spatial mappings, and physical/digital mappings must unify input and output space. A spatial mapping is successful if the spatial relationship between a physical object and its digital use is natural and intuitive and exploits spatial abilities known innately or learned early in life. And, when we play with a physical object the action space (our hands moving the object) and perception space (view and weight of the object) are perceived in the same time and place: tangible user interfaces designed to maximize input and output unification have a tight action-perception coupling leading to increased user identification between physical interface components and digital application objects.

2.3 Exploration and Collaboration

The theme *Exploration and collaboration* refers to the suitability of tangible user interfaces to facilitate exploration and collaboration. [7] clearly indicated the importance of exploration with the following design guideline: Physical/digital mappings must enable trial-and-error activity. Good physical tools enable people to perform goal-oriented activities as well as trial-and-error activities meant to explore the task space. They make sure that the cost of trial-and-error explorations is low. [5] further specify that tangible interfaces facilitate exploration and collaboration by:

- Lightweight interaction: a conversational style of interaction with rapid feedback, allowing users to proceed in small experimental steps.
- Embodied constraint: a physical set-up (such as size, form or location of objects) that leads users to cooperate by easing some activities and limiting others.

The importance of collaboration is clear from what [6] calls *space for friends*. This refers to tangible user interfaces which have both the space and affordances for multiple users. More explicitly, [5] define this as multiple access points: to distribute control such that all users can get their hands on objects of interest. This gives the opportunity to facilitate collaboration and imitation. Since collaboration and imitation are important ways for children to develop schemata level knowledge acquisition, it is important for designers of tangible user interfaces to understand the importance and mechanisms of imitation in experiential learning and to understand how to facilitate children's collaboration. Tangible systems have space and handles for co-located collaboration without the need to share input devices. Another topic belonging to this theme is imitation. Learning through imitation is very important for young children. When young children observe another person using unfamiliar objects they try to discern what the other person is using the artifact for. Tangible user interfaces are very suitable to foster imitative learning processes because of the physicality of tangibles combined with space for others and digital feedback.

2.4 Engagement and Fun

The *theme Engagement and fun* refers to the suitability of tangible user interfaces to increase presence, to be intrinsically and extrinsically motivating, and to create a positive user experience. Each of these aspects will be described.

Presence is the feeling of being in a mediated world; i.e., being fully engaged in a mediated activity. Accordingly, the degree of presence can be seen as the degree to which normal (psychological) processes are applied to a mediated world [8]. Tangible interfaces have the ability to increase presence because they allow to use normal (physical) processes to interact with a mediated world. Presence is closely related to the first two themes, where a decrease between action and effect and between physical and digital representations increases the use of normal psychological processes.

Motivation is an important user state, influencing the effort exerted and the persistence shown in solving a problem. It has been found to be a strong predictor of problem-solving success [9]. Motivation is often divided in intrinsic motivation; i.e., a genuine interest, and extrinsic motivation; i.e., some external incentive. An interface can be made extrinsically motivating by making it more game-like; i.e., with fantasy, challenge, rules and goals, sensory stimuli, mystery and active user control [10]. Intrinsic motivation can be explained by two determinants: a task performance that leads to a sense of mastery and competence, and a novelty that leads to a sense of curiosity, attention, and interest [11]. Moreover, it should be noted that feedback on and an overview of the progress with the activity are a key element of intrinsic motivation which should be supported by an interface [12]. An example of motivation based design is given by [5] as tailored representation: are the representations built on the user's experience and skills and do they invite them into interaction? A correct balance between the user's skills and the system's challenge (to use) leads to a motivating state.

The final perspective, user experience, takes a holistic view on engagement and fun. User experience is a rather fuzzy concept, often defined as technology use beyond its instrumental value. In other words, stating that it is the whole experience

of an interactive system, contrary to only the instrumental, which creates value. Hence, the focus on user experience is broader than merely heightening the enjoyment of a system; the experience has many facets joining together. Several aspects of user experience have been identified; e.g., usability, beauty, affect, temporality, and situatedness. Together, these aspects explain part of the eventual user experience [13]. Hence, an interface should be usable (i.e., intuitive), support positive emotions, and be aesthetically appealing.

The presented themes will be useful to guide the design of a PuppyIR prototype and they will also be used to develop measures for user-centred evaluation.

3 Development of a Museum Application for Children

In this section we describe ongoing work on the design of a prototype touch-table application for a museum context. We explain design decisions by relating to the themes presented in the model in Section 2. The museum (Museum in The Hague, The Netherlands) is an educational museum with a permanent exhibition ‘Your World, My World’ about humans and their relation with nature, culture, society, science and technology. In the museum a multi-touch table is used at the beginning and at the end of the museum visit, by groups consisting of two to four children. Often one or two parents are also part of the group. The application at the multi-touch table is used to determine a route through this exhibition, based on the interests of the visitors. The goal is to give the children guidance and to optimize their museum-going experience.

The main screen that people see when they arrive at the table has the solar system as a background. When people touch this screen, particles appear with colored backgrounds. See Fig. 1 for a screenshot of this screen.

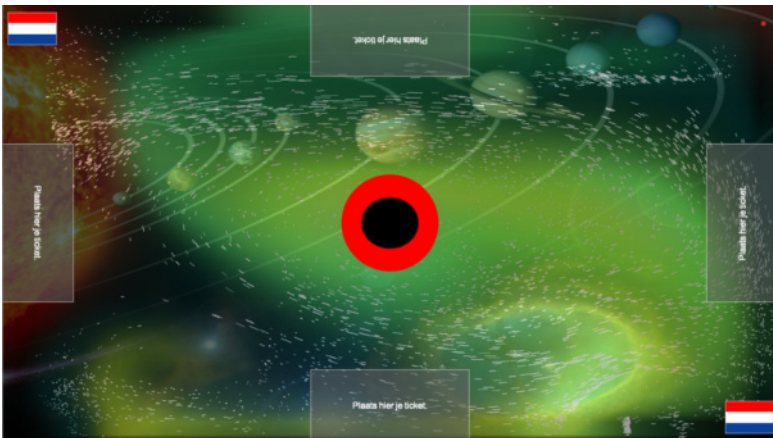


Fig. 1. The main screen of the touch table before interaction started

Interaction with this screen is not really part of the registration procedure but it is meant to be *engaging* and to *encourage exploration* by enabling trial and error activity and lightweight interaction. At the middle of each side of the table there are

virtual boxes. These boxes can be used to register. *Collaboration* is facilitated by *multiple access points*, *non-fragmented visibility* and *space for friends*. Children who take part in the experiment get a ticket that fits in these boxes and that has a marker on one side that is recognized by the table and a barcode on the other side that will be used in the quest (see below). The initial game starts when people put the tickets in the boxes. Fig. 2 shows the situation that two people already registered successfully (the red circle becomes partly green then) and two people are still busy registering.



Fig. 2. Registration with the personal tickets

When all group members have registered, the participants get a screen (see Fig. 3) where they drag the characters of their name to a bar on the table, in front of them.



Fig. 3. Putting in the name of the children in the game

In the initial game people choose categories of subjects (i.e. parts of the exhibition) they are interested in. There are twelve categories represented by round images. Everybody chooses six of these categories. Here the theme *physical and digital representations* of our model is relevant. Are physical representations in the exhibition naturally coupled to the digital representations in the images? To answer this question more research is needed. Fig. 4 shows the screen where the people can drag the images they choose to the circles near their registration ticket. The theme *actions and effects* is also relevant here. By *direct manipulation* users choose images and move them to their own area. The relation between this manual action and its effect is easy to understand (*isomorph effects*) and the action-perception coupling is tight (*unification of input and output*).



Fig. 4. Choice of categories of subjects children are interested in

The chosen categories are used to determine a route through the exhibition room of the museum. In the exhibition room many (around 120) touch screens with barcode reader are available, close to the exhibits. The registrations tickets are used here to identify the children and to transfer information from and to the table applications. Based on the results of the initial game the participants receive a personalized quest of twelve questions to be answered at twelve different exhibits. After each good answer, the children choose a virtual object they like. People can help each other whenever they want. They are near to each other and interacting with the other group members. After all members of the group have finished the quest, they go to the multi-touch table again to do the end game.

Coming back to the multi-touch table the children use their tickets to login again. From the virtual objects collected during the quest the group chooses twelve different objects. In the end game these objects are in the middle of the table. Twelve boxes with words are positioned at the edge of the table (see Fig. 5). *Collaboration* is facilitated by visibility of the words from two sides. The task is to connect the words

to the matching virtual objects by drawing lines. The children have limited time for this. After two minutes the connections are checked showing an animation: one by one the virtual objects are highlighted and the connecting lines become green when a connection is correct and red when it is not correct. The animation is meant to be *engaging* and *motivating*. After this animation the end score of the group is shown.



Fig. 5. Boxes with words that have to be connected with virtual objects in the end game

4 Observations and Conclusions

Interaction with the solar system on the main screen attracted many visitors and appeared to be *engaging*: children kept producing colors and stars for up to five minutes. While doing this they talked about fireworks, stars and imitating Harry Potter. Without much hesitation most children interacted with the table with both hands and together with other children. Only some very young children (younger than six years of age) started to interact very cautiously, with one finger. They cooperated while they tried to find out how the table worked. An often heard hypothesis was that the table reacted on heat. Some of the children discovered that the table already reacted when they hovered over it, which they found intriguing. We conclude that the solar system on the main screen *enabled trial and error activity* and *facilitated exploration and collaboration*.

The choice of characters and images appeared to be *intuitive*. Hence *actions and effects* were tightly related. In the end game children connect the words to the matching virtual objects by drawing lines. This appeared to be *less intuitive* than the other interactions on the table, especially for children under eight years old. Here the *actions and effects* relationship can be improved. However, with extra explanations and feedback, most children found out how it worked quickly.

During the animation that checked the connections drawn, nobody touched the table and the children were very attentive to see their results. This might indicate the animations were *engaging* and the children were *motivated* to have a high score.

In conclusion, these preliminary results indicate that for three out of four themes of the PuppyIR model derived in Section 2 the interactions at the multi-touch table seem to be well-designed: The interactions at the multi-touch table are *engaging and fun, facilitate exploration and collaboration* and most of the interactions are intuitive, hence *actions and effects are tightly related*. Only in the end game this relationship should be improved. Currently the interactions designed for the multi-touch table in the museum use no tangible, physical, objects (except for the registration tickets). Hence the first theme, *the appearance of the physical objects in relation to the digital representation*, seems to be irrelevant here. Tightly related, however, is the representation of parts of the exhibition of the museum in the round images children use to choose subjects they are interested in. If this coupling was clear is one of the questions we hope to be able to answer after we studied all the results of the experiments we did in the museum.

Acknowledgments. This work is part of the Puppy-IR project, which is supported by a grant of the 7th Framework ICT Programme (FP7-ICT-2007-3) of the European Union.

References

1. Ishii, H., Ullmer, B.: Tangible bits: towards seamless interfaces between people, bits and atoms. In: Proc. CHI 1997, pp. 234–241. ACM (1997)
2. Antle, A.N.: LIFELONG INTERACTIONS Embodied child computer interaction: why embodiment matters. *Interactions* 16, 27–30 (2009)
3. Sluis, R., Weevers, I., van Schijndel, C.: Read-It: five-to-seven-year-old children learn to read in a tabletop environment. In: Proc. IDC 2004, pp. 73–80 (2004)
4. Camarata, K., Do, E.Y.-I., Johnson, B.R., Gross, M.D.: Navigational blocks: navigating information space with tangible media. In: Proc. IUI 2002, pp. 31–38. ACM (2002)
5. Hornecker, E., Buur, J.: Getting a grip on tangible interaction: a framework on physical space and social interaction. In: Proc. CHI 2006, pp. 437–446. ACM (2006)
6. Antle, A.N.: The CTI framework: informing the design of tangible systems for children. In: Proc. of the 1st International Conference on Tangible and Embedded Interaction, TEI 2007, pp. 195–202. ACM (2007)
7. Sharlin, E., Watson, B., Kitamura, Y., Kishino, F., Itoh, Y.: On tangible user interfaces, humans and spatiality. *Personal and Ubiquitous Computing* 8, 338–346 (2004)
8. Nunez, D.: A connectionist explanation of presence in virtual environments, PhD thesis University of Cape Town (2003)
9. Jonassen, D.H.: Toward a Design Theory of Problem Solving. *Educational Technology Research and Development* 48, 63–85 (2000)
10. Garris, R., Ahlers, R., Driskell, J.E.: Games, Motivation, and Learning: A Research and Practice Model. *Simulation & Gaming* 33, 441–467 (2002)
11. Reeve, J.: The interest-enjoyment distinction in intrinsic motivation. *Motivation and Emotion* 13, 83–103 (1989)
12. Csikszentmihalyi, M.: *Flow: The Psychology of Optimal Experience*. Harper Collins (1991)
13. Hassenzahl, M., Tractinsky, N.: User experience - a research agenda. *The American Journal of Psychology* 25, 91–97 (2006)

Automatic Recognition of Affective Body Movement in a Video Game Scenario

Nikolaos Savva and Nadia Bianchi-Berthouze

UCLIC, University College London, MPEB Gower Street, London, WC1E6BT, UK
{nikolaos.savva.09,n.berthouze}@ucl.ac.uk

Abstract. This study aims at recognizing the affective states of players from non-acted, non-repeated body movements in the context of a video game scenario. A motion capture system was used to collect the movements of the participants while playing a Nintendo Wii tennis game. Then, a combination of body movement features along with a machine learning technique was used in order to automatically recognize emotional states from body movements. Our system was then tested for its ability to generalize to new participants and to new body motion data using a sub-sampling validation technique. To train and evaluate our system, online evaluation surveys were created using the body movements collected from the motion capture system and human observers were recruited to classify them into affective categories. The results showed that observer agreement levels are above chance level and the automatic recognition system achieved recognition rates comparable to the observers' benchmark.

Keywords: Body movement, automatic emotion recognition, exertion game.

1 Introduction

The gaming business is changing with one of the latest highlights being the inclusion of body movement in their games (e.g., Nintendo Wii, Microsoft Kinect). As more and more companies move towards this new type of technology, researchers are exploring new ways to improve and measure the player's experience by considering the role of body movement in the game [1, 21]. An important aspect of the user experience is the affective one. Until recently, the main modality used to measure the affective state of people was their facial expressions [4]. Recent psychology studies, however, have revealed that body expressions are also a very good indicator of affect [e.g., 2, 3, 10]. These studies encouraged us into researching the possibility to create an automatic recognition system that would use the players' body movement to detect their affective state.

Previous work on this subject has been carried by various researchers even if on a smaller scale than automatic recognition of affect from facial expression. An interesting work is presented in [5] and aims at detecting emotions from non-stylised acted body motions. The movement analysed in this study are cyclic knocking arm

movements expressing either basic emotions (i.e., angry, happy, sad) or a neutral state. Using SVMs classifiers, the correct recognition rate of affective states reached 50%. However, by taking into account individual idiosyncrasies in the description of the movement, the performances increased to 81%. The recognition performances were comparable to human observers' performances (varying between 59% and 71%) for the same set of stimuli, as discussed in [2]. Another interesting study aimed at recognizing affective states is the one by Gunes and Piccardi [20]. It exploits both facial expressions and upper-body gestures. The expressions considered are anger, anxiety, disgust, happiness and uncertainty. Using BayesNet, the recognition performances reached 90% by using body expressions only.

Using acted affective postures, Kleinsmith et al. [24] explored cultural differences in expressing and recognizing affect from body expressions. The analysis, based on the set of low-level descriptive features proposed in [24], highlighted some differences between the cultures but showed also the possibility to build automatic recognition models that reflect the recognition of human observers from different cultures. Similar results were obtained for the Japanese culture on affective dimensions as discussed in [25].

In all these studies, like many others [6, 9, 10, 11, 7], the affective states are acted and hence very stereotypical and even exaggerated making the generalization of these studies to real application scenarios more difficult.

Recently, there have been some attempts to model non-acted body expressions. A study that aims at detecting emotional states from non-acted body expression is presented in [19]. This is very relevant to our work as the scenario considered is whole-body computer games. However, the body expressions used in this study are static postures rather than movement. The recognition rates for the automatic systems were 60% on average for four affective states (concentrating, defeated, frustrated and triumphant). Their results were comparable with the human observers' level of agreement (i.e., 67% recognition rate) reached for the same set of stimuli. In the same work, the authors explore the possibility to recognize the level of arousal and valence from the postures of the players. Again, the results are comparable to human observers' agreement levels and well above chance level.

All these studies obtained quite interesting results highlighting the importance and the feasibility of using body expressions for automatic affect recognition. However, each of these studies explores a very particular type of body movement or body expression making the generalization of the results limited. Also, most of these studies focus on acted expressions.

Our focus on this study is to create a system to automatically recognize non-acted, affective expressive movements in the context of computer games. A benchmark is created from an analysis of the agreement between human observers in order to evaluate the system. The benchmark and the system are built using a dataset collected from players playing Nintendo Wii tennis games. The body movement of the players is collected during matches and represents the affective expressions that occur between the start and the end of a match point (winning or losing the point). The next session presents the method used to create the data. Section 3 describes the surveys used to build the benchmark on human observers. Finally, section 4 presents the automatic recognition system and its performance. We conclude with a result discussion and a comparison with human observers' agreement level.

2 Methodology and Data Analysis

The first step of our methodology was to obtain the body movement data. A motion capture system, Animazoo IGS-190, was used to record the movements of the participants during game play. The motion capture system has 17 sensors placed on the head, neck, spine, shoulders, arms, forearms, wrists, upper-legs, knees and feet. Nine players, ranging in age from 20 to 30 years old were recruited for the experiments. Since psychology studies suggest that players feel more emotionally engaged when they are familiar with their opponent [12], we asked the participants to bring a friend to compete with. The participants were asked to play the Wii Grand Slam Tennis game for 15 minutes while being recorded with the motion capture system and by a camera.

After collecting the motion captured data, we segmented them into ‘playing’ and ‘non-playing’ frame windows. We were able to collect 423 significant playing windows containing either winning or losing points. Each window time length varied between 10 and 40 seconds (i.e., between 600 and 2400 frames per window). By examining all the data, it was found that 248 out of 423 windows were very noisy (due to gimbal lock problem [13]) and we decided to exclude them as sufficient data would be available. As a result, our final data set consisted of 175 windows.

In order to identify the set of affective states to focus on, we first asked the participants to freely list the emotions they had felt during the game. Furthermore, we also observed the set of collected videos. At the end, eight emotion labels were selected: *Frustration*, *Anger*, *Happiness*, *Concentration*, *Surprise*, *Sadness*, *Boredom* and *Relief*.

In order to collect the affective ground truth for the collected movement, i.e. assign an affective label to each movement, and build the automatic recognition system, an online evaluation survey was conducted using computer animated avatars (See Fig. 1). These animations were built using the motion captured data corresponding to the selected 175 windows. Computer animated avatars were used instead of the video of the actual human participants to create a faceless non-gender, non-culturally specific ‘humanoids’ in an attempt to eliminate bias. The reason to use external observers rather than the players to build the ground truth is due to the unreliability of post-task reported feelings and to the fact that it is not possible to stop players during the gaming session to ask them their current affective state. Furthermore, because the complete affective state is expressed through a combination of modalities in a non-acted scenario, it is difficult for the players to be aware through which modality affect was expressed [25].

A forced choice survey was created and nine observers were recruited for the labeling task. The survey required the observers to assign one of the eight labels to each animated avatar according to the affective expression its body conveyed. We then used the most frequent label associated by the observers to an avatar as the representative affective state for that avatar. We call this label ground truth following the approach used in [19]. Fig. 2 shows the distribution of the 175 windows (animated avatars) grouped according to the associated ground truth. The results of the survey were also used to set a benchmark for evaluating the performances of the automatic recognition system. This will be discussed in section 5.

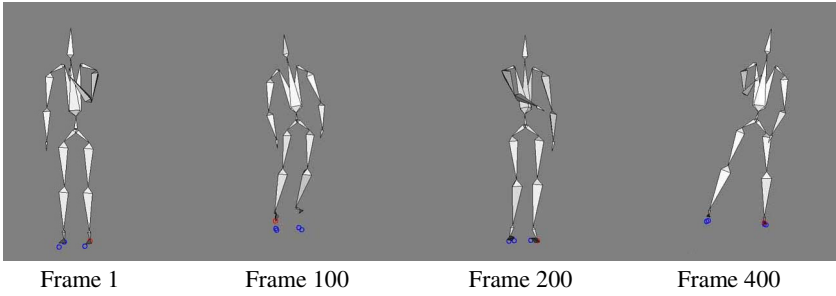


Fig. 1. The figure shows four frames of one of the avatar animations

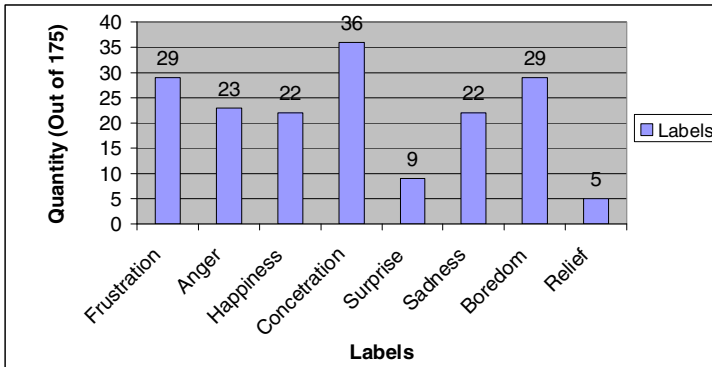


Fig. 2. Distribution of the most frequent labels associated to the 175 avatar animations

3 Low-Level Motion Description

In order to build our recognition system, the following dynamic features were selected on the basis of previous studies [e.g., 14]: *Angular Velocity*, *Angular Acceleration*, *Angular Frequency*, *Orientation*, *Amount of Movement*, *Body Directionality* and *Angular rotations*. As the motion capture data provided the 3D rotational information for each segment of the body (17 sensors were used as discussed on Section 2), a visual analysis of these set of features (see Fig. 3 for examples) was conducted for each body segment along the 3 rotational axes (x, y, z). An extensive graph analysis (by calculating all the features for each of the 17 sensors and for all the emotional states) was conducted in order to find the most discriminative features. From this analysis, we noticed that there was excessive variability between the participants for the data gathered from their leg sensors, so these data were discarded as they were contradictory and inconsistent. The final set of the most discriminative features (listed in Table 1) were selected to build the automatic recognition system.

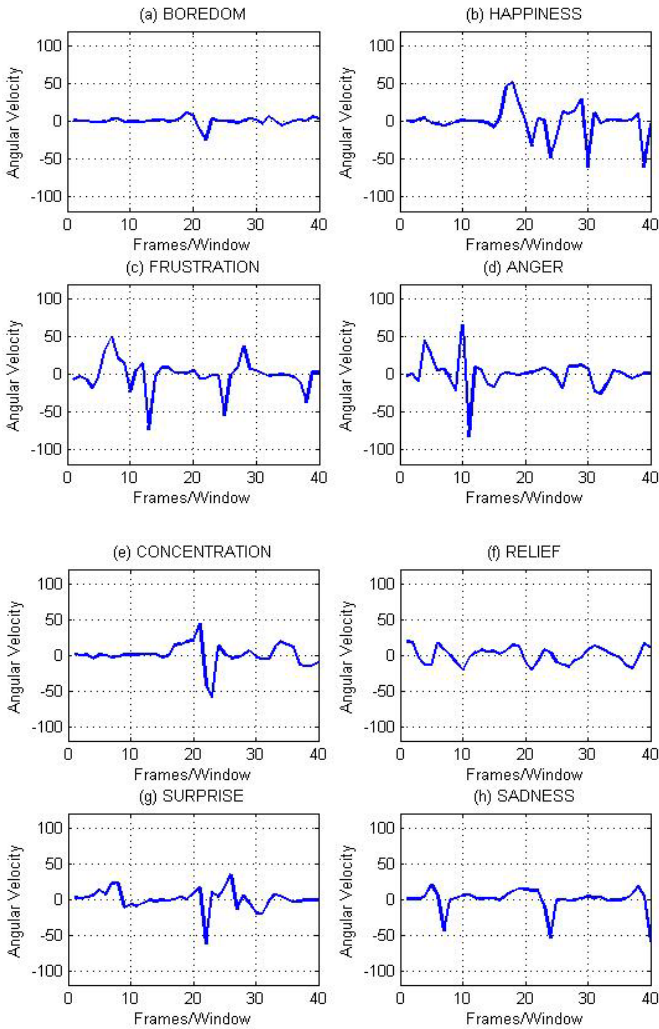


Fig. 3. Angular velocity for the X-rotation of the right forearm (Window=10)

Table 1. Identified set of discriminative features

Motion Features	Frame Interval Features
Angular Velocity _{XYZ} : Right Forearm, Arm, Hand	Amount of movement with respect to each sensor
Angular Accellaration _{XYZ} : Right Forearm, Arm, Hand	
Angular Frequency _{XYZ} : Right Forearm, Arm, Hand	
Body Directionality _X :Spine, Head	
BodySegment Rotation _{XYZ} : Right Forearm, Arm, Hand	
Angular Speed _{XYZ} - Right Forearm, Arm, Hand	

4 Automatic Recognition System and Evaluation

Since we are dealing with time related features, a dynamic learning algorithm was better suited for building our system. A Recurrent Neural Network algorithm (RNN) [15, 16, 17] was selected. The parameters of the RNN can be seen in Table 2. The inputs to the network correspond to the set of features listed in Table 1. The number of output nodes corresponds to the number of selected emotion labels.

The testing of the learning algorithm was conducted for both the ability to generalise to new observers and to new data. The 5-fold cross validation method was employed to ensure that. The training was conducted using four subsets of the training set and then the remaining subset was used to test the algorithm's ability to generalise to new data as well as to new observers. Our first experiment showed recognition rate lower than 35%. The analysis of the results showed that most of the errors were due to misclassifications of very similar expressions: frustration with anger, sadness with boredom. Furthermore, the low number of data for surprise and relief (see Fig. 2) was also one of the main causes of misclassifications. It was hence decided to refine the set of labels to be recognized.

Table 2. Initial Network Parameters

Parameter	Value	Parameter	Value
Input nodes:	47	Momentum:	0.3
Hidden layer nodes:	90	Recurrency parameter:	0.5
Output nodes:	8	Network window size:	10
Learning rate:	0.7		

According to the literature, affective states can be divided into larger categories such as negative, positive, and neutral affective states. According to Storm et al. [18], frustration and anger are both negative and high intensity emotions and their main difference is in the intensity levels of the expression. Anger normally has higher intensity than frustration. As a result, we decided to group these two emotions into one category called 'high intensity negative emotion'. Instead, sadness and boredom are negative emotions characterized by low energy/intensity. Thus, we grouped these two emotions into one category called 'low intensity negative emotion'. Furthermore, given the low number of samples for 'surprise' and 'relief', these two labels were removed from the data set. Therefore, we are left with four classes: 'high intensity negative emotion', 'happiness', 'concentration' and 'low intensity negative emotion' and 161 windows as data set. The distribution of affective states with respect to the data set is illustrated in Fig. 4. These 4 classes of emotions cover the four quadrants of valence-arousal space generally used to describe emotional states, with Concentrated being a neutral state and Happiness, in this case, representing the high intensity positive emotions.

Various experiments to identify the best set of features were executed. The best results were obtained by using only *angular velocity*, *angular speed* and *amount of*

movement as input features. Angular velocity is a vector quantity which specifies the angular speed (a scalar) of an object along with the axis which the object is rotating around. The 175 windows in our data set were further segmented into smaller frame intervals. Since each window varied between 600 and 2600 frames (10 to 40 seconds), we segmented them into smaller, equal frame intervals in order to import them into the RNN. The best performance was achieved using a network window size equal to ten. For example, a window consisting of 600 frames was segmented into 60 frame intervals containing ten consequent frames each. As a result, for one data point we extracted sixty sub-windows.

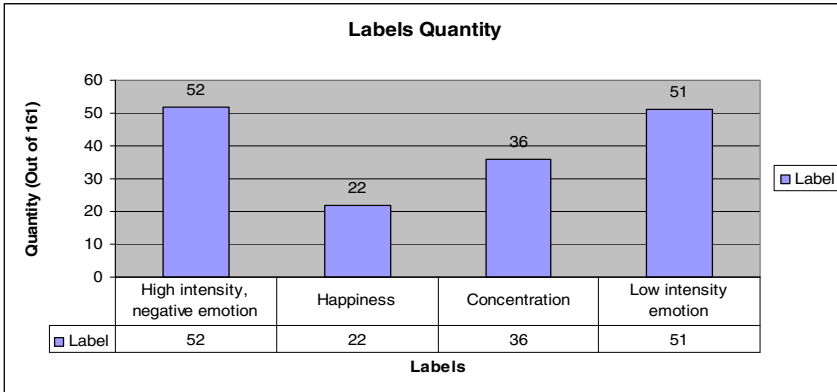


Fig. 4. The graph shows the number of animated avatars for each emotion category

Table 3. Confusion matrix for the testing set

		<i>Predicted</i>				
		High intensity, negative emotion	Happy	Concentr.	Low intensity emotion	Multiple classes
<i>Actual</i>	High intensity, negative emotion	223 (64%)	42	46	6	31
	Happy	13	124 (58%)	24	11	43
	Concentration	51	37	102 (36%)	32	62
	Low intensity emotion	9	33	49	263 (67%)	38

The resulting data set was split in two parts (training and testing set) reserving the 1/3rd, 1239 samples, to be used as a testing set and the remaining 2/3rd, 3720 samples, for the training set, from overall 4959 instances. Table 3 shows the recognition performance over the testing set. Overall the network was able to categorize correctly 712 samples corresponding to 57% of the testing set. In particular, 64% of the ‘high intensity negative emotion’ samples were correctly classified, 58% accuracy was obtained for ‘happiness’ and 67% for low intensity negative emotion. Only 36% accuracy was, instead, obtained for ‘concentration’. The low accuracy obtained for ‘concentration’ could be due to the fact that the human observers may have used this label when the avatar’s expression did not express any of the other affective states as discussed in [19]. Finally, the column named as ‘Multiple classes’ in table 3 contains the number of the test samples that our algorithm was not able to categorize into only one class. An analysis of the results highlighted, also, the large variability between expressions belonging to the same category. This was due to the large diversity of the players’ playing styles. Thus, for every class we had a variety of different input patterns. An example is provided in Fig. 5.

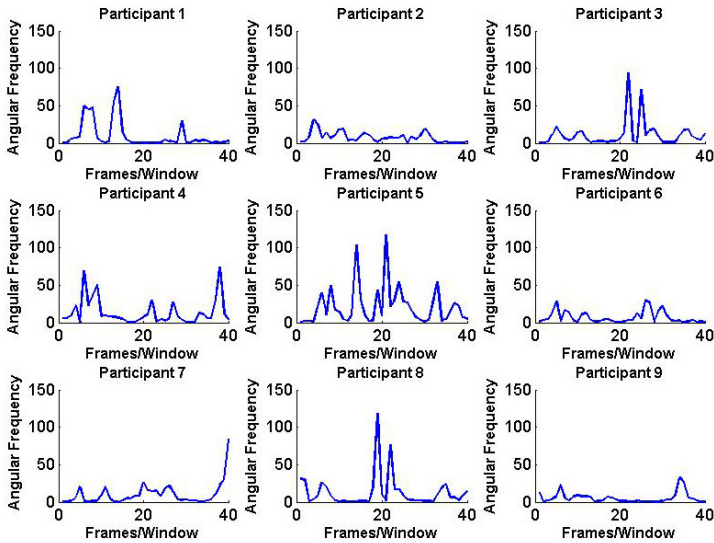


Fig. 5. The differences in angular frequency between the participants that portray anger

From Fig. 5, we can notice that players (participants) can be categorized into two groups, the ones that do not move a lot during the game (P2, P6, P7 and) and the others that move a lot (P1, P3, P4, P5 and P8). This difference between them exists because some of the participants tend to play the game using only their hand/wrist in comparison with the other group that uses their arm and shoulder as well. As discussed in [22], players adopt different body movement strategies according to their level of expertise but also according to their motivations for engaging in the game play.

5 Discussion

To evaluate the performance of the automatic recognition system, we followed a simplified version of the method proposed in [19]. The evaluation method proposed in [19] requires three groups of observers in order to fully separate the computation of the benchmark from the testing of the system. This was not possible in this case as the number of observers available was quite small. Hence, we divided the observers into two groups; the first group of observers was used to create the training set and the second group of observers was used for the testing set. The agreement level between the two groups of observers resulted in 61.49% since the two groups agreed only on 99 out of the 161 instances. Finally, we can observe that our system's accuracy (57.46%) is comparable to, even if slightly lower than, the observers' agreement. The results are hence very encouraging given the complexity of our data set. The results are also comparable to the results obtained for complex expressions in the acted and non-acted studies discussed in the introduction.

Bernhardt et al. [5] is one of the studies we can compare with ours since they used motion data instead of single postures. The researchers used arm movement features to recognize emotions from 'knocking' movements reaching similar performance with our system (59% accuracy). However, when individual idiosyncrasies were considered, their results increased to 81%. As we pointed out in Fig. 5 and various studies [19] show that, players not only have their own idiosyncrasy but they employ different strategies when playing. By taking into account such differences in the modelling process, it could be expected that the performance of our system would improve. We still have however to remember that in [5], the expressions are acted and hence simpler to discriminate, whereas in our study the expressions are non-acted and often very subtle making even the human observer recognition task much harder. Also, differently from our study, their movements were repeated and hence easily to segment into movement phases before describing them. Hence, by adding a segmentation of playing movement in our study, it is possible that our method could reach much better results.

Besides discussing the evaluation of our system, we should consider the limitations of our approach. By analysing the features visually and individually, we have possibly discarded some important ones. It is possible that combination of features that individually appear to have low discrimination may instead result in being very informative. Therefore, it would be important to perform a more thorough statistical analysis of the features and their combinations (e.g. by using PCA). Finally, by adding to the recognition system information about the type of shots being played (back-hand, fore-hand, etc) together with its features may bring better performances in the recognition of each emotion. In fact, biomechanical aspects of the type of shot may have an effect on the kinematic features considered independently of the emotion expressed. These observations will be our guide for our next step.

References

1. Kim, J.H., Gunn, D.V., Schuh, E., Phillips, B., Pagulayan, R.J., Wixon, D.: Tracking real-time user experience (TRUE): a comprehensive instrumentation solution for complex systems. In: Proceedings of the 26th Annual SIGCHI Conference On Human Factors In Computing Systems, pp. 443–452. ACM, New York (2008)

2. Pollick, F., Paterson, H., Bruderlin, A., Sanford, A.: Perceiving affect from arm movement. *Cognition* 82, 51–61 (2001)
3. Mehrabian, A., Friar, J.: Encoding of attitude by a seated communicator via posture and position cues. *Journal of Consulting and Clinical Psychology* 33, 330–336 (1969)
4. Mandler, G.: *History of Psychology. Emotion*, vol. 1, ch. 8. Wiley (2002)
5. Bernhardt, D., Robinson, P.: Detecting Affect from Non-stylised Body Motions. In: Paiva, A.C.R., Prada, R., Picard, R.W. (eds.) *ACII 2007. LNCS*, vol. 4738, pp. 59–70. Springer, Heidelberg (2007)
6. Castellano, G., Villalba, S., Camurri, A.: Recognising Human Emotions from Body Movement and Gesture Dynamics. In: Paiva, A., Prada, R., Picard, R.W. (eds.) *ACII 2007. LNCS*, vol. 4738, pp. 71–82. Springer, Heidelberg (2007)
7. Kleinsmith, A., Fushimi, T., Bianchi-Berthouze, N.: An incremental and interactive affective posture recognition system. In: Carberry, S., De Rosi, F. (eds.) *International Workshop on Adapting the Interaction Style to Affective Factors*, in conjunction with the *International Conference on User Modeling* (2005)
8. Kleinsmith, A., Bianchi-Berthouze, N.: Recognizing Affective Dimensions from Body Posture. In: Paiva, A.C.R., Prada, R., Picard, R.W. (eds.) *ACII 2007. LNCS*, vol. 4738, pp. 48–58. Springer, Heidelberg (2007)
9. Coulson, M.: Attributing emotion to static body postures: Recognition accuracy, confusions, and viewpoint dependence. *Journal of Nonverbal Behavior* 28, 117–139 (2004)
10. Kleinsmith, A., De Silva, R., Bianchi-Berthouze, N.: Cross-cultural differences in recognizing affect from body posture. *Interacting with Computers* 18(6), 1371–1389 (2006)
11. Camurri, A., Mazarino, B., Ricchetti, M., Timmers, R., Volpe, G.: Multimodal Analysis of Expressive Gesture in Music and Dance Performances. In: Camurri, A., Volpe, G. (eds.) *GW 2003. LNCS (LNAI)*, vol. 2915, pp. 20–39. Springer, Heidelberg (2004)
12. Mandler, G.: *History of Psychology. Emotion*, vol. 1, ch. 8. Wiley (2002)
13. Kitagawa, M., Windsor, B.: *MoCap for Artists: Workflow and Techniques for Motion Capture*, pp. 190–194. Focal Press (2008)
14. Roether, C., Omlor, L., Christensen, A., Giese, M.A.: Critical features for the perception of emotion from gait. *Journal of Vision* 8(6), 15, 1–32 (2009)
15. Elman, J.L.: Finding Structure in Time. *Cognitive Science* 14, 179–211 (1990)
16. Haykin, S.: *Neural Networks: A Comprehensive Foundation*, 2nd edn., pp. 754–777. Prentice-Hall (1999)
17. Bodén, M.: A guide to recurrent neural networks and backpropagation, in *The DALLAS project. Report from the NUTEK-supported project AIS-8: Application of Data Analysis with Learning Systems, 1999-2001*. Holst, A. (ed.), SICS Technical Report T2002:03, SICS, Kista, Sweden (2001)
18. Storm, C., Storm, T.: A taxonomic study of the vocabulary of emotions. *Journal of Personality and Social Psychology* 53(4), 805–816 (1987)
19. Kleinsmith, A., Bianchi-Berthouze, N., Steed, A.: Automatic Recognition of Non-Acted Affective Postures. *IEEE Transactions on Systems, Man and Cybernetics, Part B* (2011)
20. Gunes, H., Piccardi, M.: Bi-modal emotion recognition from expressive face and body gestures. *Journal of Network and Computer Applications* 30, 1334–1345 (2007)
21. Muller, F., Bianchi-Berthouze, N.: Evaluating Exertion Games Experiences from Investigating Movement Based. *Human-Computer Interaction Series, Part 4*, pp. 187–207. Springer, Heidelberg (2010)

22. Pasch, M., Bianchi-Berthouze, N., van Dijk, B., Nijholt, A.: Movement-based Sports Video Games: Investigating Motivation and Gaming Experience. *Entertainment Computing* 9(2), 169–180 (2009)
23. De Silva, R., Bianchi-Berthouze, N.: Modeling human affective postures: An information theoretic characterization of posture features. *Journal of Computational Animation and Virtual Worlds* 15(3-4), 269–276 (2004)
24. Kleinsmith, A., de Silva, R., Bianchi-Berthouze, N.: Cross-cultural differences in recognizing affect from body posture. *Interacting with Computers* 18, 1371–1389 (2006)
25. Kleinsmith, A., Bianchi-Berthouze, N.: Recognizing Affective Dimensions from Body Posture. In: Paiva, A.C.R., Prada, R., Picard, R.W. (eds.) *ACII 2007*. LNCS, vol. 4738, pp. 48–58. Springer, Heidelberg (2007)
26. Russell, J.A., Feldman-Barrett, L.: Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *J. Pers. Social Psychol.* 76, 805–819 (1999)

Towards Mimicry Recognition during Human Interactions: Automatic Feature Selection and Representation

Xiaofan Sun¹, Anton Nijholt¹, and Maja Pantic^{1,2}

¹Human Media Interaction, University of Twente
PO Box 217, 7500 AE Enschede, The Netherlands

²Department of Computing, Imperial College

180 Queen's Gate, London SW7 2AZ, UK

{x.f.sun,a.nijholt}@ewi.utwente.nl. m.pantic@imperial.ac.uk

Abstract. During face-to-face interpersonal interaction people have a tendency to mimic each other, that is, they change their own behaviors to adjust to the behavior expressed by a partner. In this paper we describe how behavioral information expressed between two interlocutors can be used to detect and identify mimicry and improve recognition of interrelationship and affect between them in a conversation. To automatically analyze how to extract and integrate this behavioral information into a mimicry detection framework for improving affective computing, this paper addresses the main challenge: mimicry representation in terms of optimal behavioral feature extraction and automatic integration.

Keywords: mimicry representation, human-human interaction, human behavior analysis, motion energy.

1 Introduction

Mimicry plays an important role in human-human interaction. Mimicry refers to the coordination of movements in both timing and form during interpersonal communication. Behavior matching, synchronized changes in behavior and facial expressions, matching in posture and mannerisms are examples of mimicry. But there can also be vocalic mimicry and matching of verbal style. Mimicry is ubiquitous in daily interpersonal interaction. For example, when two interactants are facing each other and one of them takes on a certain posture such as moving sideways or leaning forward, then the partner may take on a congruent posture [1], [2], [12], and when one takes on certain mannerism such as rubbing the face, shaking the legs, or foot tapping, the partner may take on a congruent mannerism [2]. Another example, if one is crossing his legs with the left leg on top of the right, the other may also cross his legs with the right leg on top of the left leg (called “mirroring”) or with the left leg on top of the right leg (called “postural sharing”).

Mimicry enhances social interaction by establishing rapport and affiliation [2] and by observing mimicry behavior conclusions can be drawn about the quality of the interaction and about interpersonal relationships between conversational partners. For that reason mimicry has become object of study of social psychology. What behavioral cues show mimicry, how to rate mimicry, and what different kinds and functions of mimicry can be distinguished are among the main questions that are studied. Mimicry, as it can be perceived from facial expressions, vocal behavior, and body movements, affects human-human interaction.

It is interesting to look at a possible role of mimicry in human-computer interaction. It is well known that humans can consider computers as social actors and in particular in agent-oriented interfaces designers anticipate such behavior. Moreover, we see more applications where the role of the computer is not so much to be efficient or only efficient, but also being social or entertaining, for example in health and well-being situations where the computer plays a coaching function, in domestic situations where a social robot needs to be trusted in order to accept his help and advice, and, of course in gaming and entertainment applications where we play and communicate with virtual humans (avatars, embodied conversational agents, ...). More human-like behavior of a virtual human allows for more natural interaction and modeling mimicry makes it possible to understand and generate mimicry behavior in human- virtual human or human-social robot interaction.

Many researchers from psychology have investigated mimicry. Until now, research in affective computing has been concerned with the affective role of facial expressions, body postures, gaze directions, prosody, and (neuro-)physiological information. But, the role of mimicry in human-human interaction and how this role can be exploited in human-machine interaction (where, machine can be a computer, a robot, a virtual human, an environment, et cetera) to improve the interaction and the experience, has not been explored. It requires automatic (machine) detection of mimicry, automatic understanding of mimicry, automatic prediction of mimicry, and also automatic generation of mimicry. And, obviously, then the role of mimicry in human-human interaction should be completely understood.

In current and future game and entertainment environments we will meet people. Their characteristics and their behavior will not always be fully mediated. There will probably be a lack of subtle social signals that play important roles in human-human interaction and that are hard to mediate. Our research aims at understanding these subtle social signals, in particular mimicry, in order to mediate them in human-nonhuman interactions. This will help improving natural interaction (in natural situations) and establishing interpersonal relationships that people would like to have and maintain, whether it is with a human or with a social and intelligent human-like device. Mimicry is an informative and communicative act that helps to convey and recognize intentions and affect that are important for interaction and establishing relationships.

In our experiments on the role of mimicry in social interaction we have conversational partners that are being observed in a laboratory setting. Data such as location, body orientation, head pose, gestures, and vocal activities is obtained from camera and audio input. Behavioral patterns are analyzed to detect people's relationships, individuals' affect and assessing the quality of the interaction.

In this paper, reporting about work in progress, we show that we can find and represent behavioral mimicry in conversations by analyzing human actions in prediction models. In section 2 we have some observations on factors affecting mimicry. A short description of the corpus that we collected for mimicry analysis is presented in section 3. A more comprehensive description will appear elsewhere. The corpus is used for extracting and detecting of features for mimicry recognition. We shortly discuss our annotation steps and the automatic extraction of mimicry episodes. In section 4 we present some preliminary conclusions, including the conclusion that automatic mimicry identification is possible.

2 Mimicry to Be Expected in Social Interaction

The first and most important aspect in this study is to collect data which includes various behavioural mimicry or interactional synchrony in social interactions. The social interaction scenarios that aim at elicitation of behavioural mimicry or interactional synchrony need to be natural in terms of the different factors that may affect the likelihood or increase the chance of mimicry occurring. However, the factors that affect mimicry are not unique and they cannot account for everything. We illustrate this with a few examples. For example, in daily life, when we talk with our boss, we mimic his or her behavior or repeat what he or she said. Not necessarily because you really agree with him or her, but there may be a desire to affiliate for personal benefits and even without awareness. Moreover, when we share similar opinions in a meeting, we also have a strong tendency to mimic other members' behaviors in an attempt to gain acceptance. In some cases, there is a strong mimicry tendency because of directly active goals, sometimes we mimic to improve a harmonious interrelationship, but usually we mimic without consistent awareness. Mimicry occurs in our daily life all the time, and most of the time this mimicry behavior signals important social attitudes and affects.

Mimicry is sensitive to social context, so automatic mimicry behavior changes according to one's active goals in a realistic social situation. Mimicry responses are modulated by the social signal value of the behavior. That is, many social signals may be implied or signaled by various mimicry behaviors. For example, expressive behavior is more present in conversations about positive experiences than in conversations about negative experiences. And usually participants are more active and more willing to show facial expressions and body language when they seem to be familiar with the topic. That is, they show their opinions, both verbally and non-verbally, more actively when they are familiar with the topic. Hence, to choose a topic which is familiar with both interactants is important for collecting more mimicry episodes. Behavioral mimicry plays an important role in identifying interactants' attitude, affect and even roles played in conversations. Previous studies showed a higher mimicry tendency when people perceived themselves as similar, would like to be similar, or want to display themselves as similar [10]. In addition, they may have aligned goals [8] and lean forward, they may share attitudes [11] and lean forward and nod, they may like their conversational partner and show it by synchronous head

nodding and shaking [8], they may want the other to have a positive perception of them and display matching smiles [7], or empathize with the other and show this in matching behavior [6]. Moreover, mimicry also helps in identifying the roles people play in a conversation, for example, people always expand themselves unconsciously when they are perceived as dominant, however, constrict themselves when they are perceived as submissive [15], [16].

Thus, in our experiments a first scenario designed for collecting behavioral mimicry and interactional synchrony is about discussing a familiar topic that makes it possible to share attitudes with each other. In the second scenario, given that most participants in our experiments are students, we give a hypothetical conversational topic which is familiar with their actual daily life. They are given a non-task-oriented communication assignment which requires self-disclosure and emotional discovery.

3 Experiment Setup

For extracting and detecting of features for mimicry recognition in our prediction model, we used a corpus of 53 human-to-human interactions. This corpus is described in Section 3.1. Section 3.2 is devoted to the description of the features that are annotated in our experiments to be used for mimicry representation. Section 3.3 presents the algorithm used for tracking mimicry in terms of the features annotated. Finally Section 3.4 discusses our methodology for automatic mimicry extraction.

3.1 Data Collection

Our data is drawn from a study of face-to-face discussions and conversations. 43 subjects from Imperial College, London participated in this experiment. They were recruited using the Imperial College social network and were compensated 10 pounds for one hour of their participation.

The experiment included two sessions. In the first session, participants were asked to choose a topic from a list, which had several statements concerning that topic. Participants were then asked to write down whether they agree or disagree with each statement of their chosen topic. Participants were then asked to present their own stance on the topic, and then to discuss the topic with their partners, who may have different views on the topic. Participants could talk about anything they wanted, that is, the statements we listed were just a reference. In the second session, the intent is to simulate a situation where participants wanted to get to know their partner a bit better and they needed to disclose personal and possibly sensitive information about themselves. Participants were given a non-task-oriented communication assignment that required self-disclosure and emotional discovery. Participant A played a role as a student in university who was looking for a room to rent urgently. Participant B played a role as a person who owns an apartment and wants to let one of the rooms to the other person.

We collected synchronized multimodal data for each session. In each session we recorded data from the participants separately and from the two participants together,

including voice and body behaviors. In the visual-based channel we recorded data using 7 cameras for each person and 1 camera for both persons at the same time. The camera for both persons was used for recording an overview of the interaction, while the other 7 cameras were used for recording the two participants separately, including far-face view, near-face view, upper-body view, and whole body view with and without color. See Fig. 1 for some camera views. Both participants wore a lightweight and distance-fixed headset with microphone. For detecting head movements both participants wore rigs on their heads during recording. The rig is a lightweight, flexible metal wire frame and fitted with 9 infrared LEDs. Given the face location and orientation, the nine LEDs allow us to get detailed information about the characteristics of the head movements.

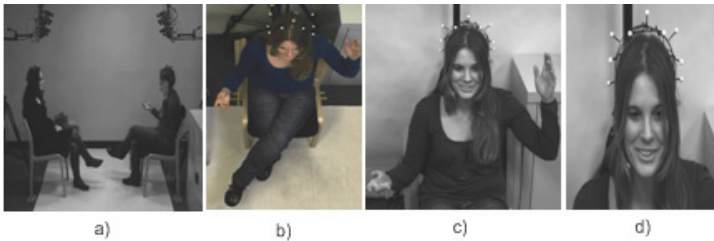


Fig. 1. a) Setup for corpus collection. b) Higher-view for whole body recording for each participant separately. c) Recording upper-body movement. d) Recording head movements and facial expressions.

3.2 Annotation

As discussed in the previous section, the corpus is a collection of face-to-face interactions designed with the aim to study mimicry behavior and interactional synchrony. Hence, the main focus of the annotation scheme is the labeling of the behavior expressions and in particular behavioral mimicry.

The annotators' job is to look at videos of these interactions and annotate them with information about the "human behavioural expressions" and "social signals" of the participants. This means that they continuously try to answer the questions "How the actions of those participants display: is he/she nodding, head shaking, etc.?" and "Do they mimic each other?"

For each annotation assignment, the main annotation steps are based on widely accepted concepts of mimicry. Firstly, mimicry is dynamic, hence, signals of mimicry behavior occur successively. Secondly, mimicry is about one conversational partner imitating the other [1]. That is, mimicry is when people express or share similar behavior during interaction, at the same time or one after another, in response to the other.

The main annotation steps are briefly introduced below:

1. Annotation of speakers and listeners (usually listeners and speakers take turns) based on the utterances.

2. Segmentation into episodes, where each episode consists of a sequence speaker1, listener2, speaker2, listener1, hence, each of the two participants appears in the sequence both as speaker and as listener.
3. Annotation of visual-based behavioral expressions for the two partners such as smile, nod, head shake, hand gesture, and body leaning.
4. Annotation of mimicry cues: we have predefined notions of behavioural cues; after manually annotating episodes and behavioural cues, we use an algorithm (see below) to automatically compare whether the selected notions match or not; if they match label mimicry (YES), if not, label mimicry (NO).

Hence, after the first step of annotation, the utterance token of a participant is labeled as listener or speaker. In the second step, we select the conversation segments in such a way that each participant is seen as a speaker and a listener, because their (amount of) mimic behavior can be dependent on their role in the conversation (speaker or listener). Then, in the third step, behaviors expressed by participants are labeled, using visual cues, for analyzing behavioral mimicry. Finally, in terms of mimicry perception we annotate those behaviors expressed by paired participants as mimicry or not. After annotating conversation segments and visual cues for detecting mimicry, based on these annotation results we extract mimicry episodes. In each mimicry episode visual cues are extracted to identify behavioural mimicry. This will be discussed in more detail in section 3.3.

Algorithm to automatically extract mimicry episodes

```

Given: i: current episode index;
SB[i]: the array of mimicry cues displayed by the
speaker during the current (ith) episode;
SB_N[i]: the total number of mimicry cues
displayed by the speaker during the current (ith)
episode;
LB[i]: the array of mimicry cues displayed by the
listener during the current (ith) episode;
LB_N[i]: the total number of mimicry cues
displayed by the listener during the current (ith)
episode.

```

Detect speaker's mimicry:

```

For each frame t
Do int SB[1]=0 if SB1<SB_N[i] and apply ++SB[1]);

```

Mimicking the previous episode's speaker:

```

For (int SB[2]=0; SB[2]<SB_N[i-1]; ++SB[2])
If (SB[i][SB1] == SB[i-1][SB2]), and label
mimicry;

```

```

Mimicking the current episode's listener:
  For (int lb1=0; lb1<SB_N[i-1]; ++lb1)
    If (SB[i][sb1] == LB[i][lb1]), and label
mimicry;

Detect listener's mimicry (only consider current
round)
  For(int lb1=0; lb1<LB_N[i]; ++lb1)
  For(int sb1=0; sb1<SB_N[i]; ++sb1)
    If(LB[i][lb1] == SB[i][sb1]), and label
mimicry.

```

3.3 Methodology

In this section, we first describe the human action recognition technique we use to extract motion features and represent the motion cycle [19] for identifying behavioral mimicry. Then, by analyzing our results, we show that in our annotated mimicry episodes, mimicry indeed occurs more frequently. Moreover, we investigate that similarity is indeed an important factor that increases mimicry. In this study we only annotated the episodes on one aspect of similarity, That is, the role participants play in a conversation. In fact, similarity was manipulated in various ways in previous studies: status, appearance, attitudes, sport interests, leisure interests, et cetera.

We calculated the motion cycle in each manually annotated episode in our attempt to detect behavioural mimicry. The motion cycle is extracted in terms of the accumulated or averaged motion energy (AME) which only is computed in areas that include changes [16], [19]. Hence we propose to represent the motion cycle by computing a group of accumulated motion images (AMIs). In detail, AMI represents the time-normalized accumulative and average action energy and contains pixels with intensity values for representing motions [21]. In the AMI, the regions containing pixels with higher intensity values denote that motions are more complex and occur more frequently. Although AMI is related to MEI and MHI [19], a fundamental difference is that AMI describes the motions by using the pixel intensity directly. That is, instead of giving all equal weights for all changing areas in MEI or assigning higher weights for new frames but lower weights for older frames in MHI.

$$AMI(x, y) = \frac{1}{T} \sum_{t=1}^T |D(x, y, t)| \quad (1)$$

where $D(x, y, t) = I(x, y, t) - I(x, y, t-1)$ in which T denotes the length of the query action video (i.e., total number of frames) and I stands for the intensity of the current frame. Fig 2 illustrates visual behavioural mimicry, extracted from consecutive sets of frames of a recording.



Fig. 2. A group of behavioural mimicry extracted from consecutive sets of frames (frame 92, 96, 98, 102, 103, 105, 108, 113, 120, and 123) of a recording in our database

The figure illustrates hand gesture mimicry behavior. This behavior is visualized by presenting the results of motion intensity calculation for hands movement. Motion cycle images are calculated by AMI in several successive frames for each annotated mimicry behavior in our data.

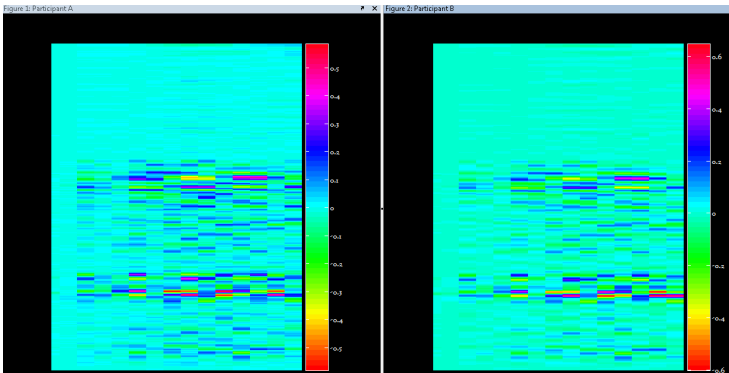


Fig. 3. The cross-correlations of body movements between two persons who interact with each other. The vertical axis shows the motion energy, the horizontal axis shows the frame numbers.

Fig 3 demonstrates the cross-correlations of movements between two persons, generated from a fragment of 580 windows (20 sec) in a conversation on looking for a suitable roommate. The vertical axis shows the motion energy, the horizontal axis shows the frame numbers. The left part of the figure shows the motion energy calculated in each frame for participant A; the right part shows the motion energy calculated in each frame for participant B.

Summarizing, in Fig. 2 we demonstrate that visual-based mimicry can be visually extracted in a short time period of around 5 seconds in our data. In Fig. 3, we accumulate all movements during a longer period (20 sec) to see the general motion tendency expressed by two people who interact in a conversation. We can see rather similar cross-correlations of body movements between conversational partners. Hence, we can safely assume that behavioral mimicry probably occurs with a high chance in this period.

4 Conclusions and Future Work

Our results show that behavioral information from conversational partners can be extracted and integrated in order to demonstrate mimicry. Moreover, it became clear that mimicry is indeed ubiquitous in human-human conversation. Methods to analyze motion energy can be applied and improved to deal with mimicry in a machine understanding approach [18]. From our mimicry episode annotation we have learned about the role of similarity, that is, the similarity of roles played in interactions. Moreover, mimicry analysis does contribute to recognizing the role of affect and empathy in social interaction.

For future work, we plan to extract relevant features from audio and visual channels for detecting more mimicry cues in our database. The aim is to automatically identify mimicry when people mimic facial expressions, vocal productions, and body movements with their conversational partners or others around them in daily interaction. In affective computing research the detection of nonverbal cues has been considerably improved in previous years. The role of verbal and nonverbal expressions has been investigated, including their necessity for understanding behavioural patterns, mental states, attitudes and personality traits. It has also been demonstrated that people tend to mimic and synchronize vocal utterances during a conversation. Usually, people with different personalities probably prefer different interaction tempos. In a conversation, if the communication goes well or is improving, the speech cycles of conversational partners become mutually entrained. The study of vocalic mimicry is not so much to find out the attribution of each feature of speech, such as spectral features or non-spectral features to specific human affect, but the focus is rather on the changing of and the similarity of speech utterances. Including vocalic mimicry is a next step in our research on modeling mimicry.

Acknowledgments. We gratefully acknowledge the help of Michel Valstar, Jeroen Lichtenauer, and many others from Imperial College who helped to realize the experiments for the data collection. This work has been funded in part by the European Community's 7th Framework Programme [FP7/2007-2013] under the grant agreement no 231287 (SSPNet). The work of Maja Pantic is also funded in part by the European Research Council under the ERC Starting Grant agreement no. ERC-2007-StG-203143 (MAHNOB).

References

1. Bernieri, F.J.: Coordinated movement and rapport in teacher student interactions. *Journal of Nonverbal Behavior* 12(2), 120–138 (1998)
2. Bernieri, F.J., Reznick, J.S., Rosenthal, R.: Synchrony, pseudosynchrony, and dissynchrony: Measuring the entrainment process in mother-infant interactions. *Journal of Personality and Social Psychology* 54(2), 243–253 (1988)
3. Chartrand, T.L., van Baaren, R.: Chapter 5 Human Mimicry, pp. 219–274. Academic Press (2009)

4. Chartrand, T.L., Bargh, J.A.: The chameleon effect: the perception-behavior link and social interaction. *Journal of Personality and Social Psychology* 76(6), 893–910 (1999)
5. Chartrand, T.L., Jefferis, V.E.: Consequences of automatic goal pursuit and the case of nonconscious mimicry, pp. 290–305. Psychology Press, Philadelphia (2003)
6. Chartrand, T.L., Maddux, W., Lakin, J.L.: Beyond the perception behavior link: The ubiquitous utility and motivational moderators of nonconscious mimicry, pp. 334–361. Oxford University Press, New York (2005)
7. Giles, H., Powesland, P.F.: *Speech style and social evaluation*. Academic Press, London (1975)
8. Gueguen, N., Jacob, C., Martin, A.: Mimicry in social interaction: Its effect on human judgment and behavior. *European Journal of Sciences* 8(2), 253–259 (2009)
9. Hess, U., Blairy, S.: Facial mimicry and emotional contagion to dynamic emotional facial expressions and their influence on decoding accuracy. *Int. J. Psychophysiology* 40(2), 129–141 (2001)
10. Jefferis, V.E., van Baaren, R., Chartrand, T.L.: The functional purpose of mimicry for creating interpersonal closeness. Manuscript, The Ohio State University (2003)
11. Kopp, S.: Social resonance and embodied coordination in face-to-face conversation with artificial interlocutors. *Speech Communication* 52(6), 587–597 (2010)
12. LaFrance, M.: Nonverbal synchrony and rapport: Analysis by the cross-lag panel technique. *Social Psychology Quarterly* 42(1), 66–70 (1979)
13. Lakin, J.L., Chartrand, T.L., Arkin, R.M.: Exclusion and nonconscious behavioral mimicry: Mimicking others to resolve threatened belongingness needs (2004) (manuscript)
14. Lakin, J.L., Jefferis, V.E., Cheng, C.M., Chartrand, T.L.: The chameleon effect as social glue: Evidence for the evolutionary significance of nonconscious mimicry. *Journal of Nonverbal Behavior* 27(3), 145–162 (2003)
15. Miles, L.K., Nind, L.K., Macrae, C.N.: The rhythm of rapport: Interpersonal synchrony and social perception. *Journal of Experimental Social Psychology* 45(3), 585–589 (2009)
16. Briassouli, A., Kompatsiaris, I.: Robust temporal activity templates using higher order statistics. *IEEE Transactions on Image Processing* 18(12), 2756–2768 (2009)
17. Chandrashekhar, V.H., Venkatesh, K.S.: Action energy images for reliable human action recognition. In: *Proceedings of the Asian Symposium on Information Display (ASID 2006)*, pp. 484–487 (October 2006)
18. Nagaoka, C., Komori, M., Nakamura, T., Draguna, M.R.: Effects of receptive listening on the congruence of speakers' response latencies in dialogues. *Psychological Reports* 97, 265–274 (2005)
19. Bobick, A.F., Davis, J.W.: The recognition of human movement using temporal templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23(3), 257–267 (2001)
20. Shechtman, E., Irani, M.: Matching local self-similarities across images and videos. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2007)*, Minneapolis, Minn, USA, pp. 1–8 (June 2007)
21. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005)*, San Diego, Calif, USA, vol. 1, pp. 886–893 (June 2005)

A Playable Evolutionary Interface for Performance and Social Engagement

Insook Choi¹ and Robin Bargar²

¹ Emerging Media Technologies Program, Entertainment Technology Department,
New York City College of Technology of the City University of New York
300 Jay St V205, Brooklyn, NY, 11201 USA
insook@insookchoi.com

² School of Media Arts, Columbia College Chicago, Chicago, IL, USA

Abstract. An advanced interface for playable media is presented for enabling both musical performance and multiple agents' play. A large format capacitive sensing panel provides a surface to project visualizations of swarm simulations as well as the sensing mechanism for introducing human players' actions to the simulation. An evolutionary software interface is adapted to this project by integrating swarm algorithms to playable interface functionality with continuous auditory feedback. A methodology for using swarm agents' information to model sound synthesis is presented. Relevant feature extraction techniques are discussed along with design criteria for choosing them. The novel configuration of the installation facilitates a unique interaction paradigm that sustains social engagement seamlessly alternating between cooperative and competitive play modes.

Keywords: evolutionary interface, agents, swarms simulation, sound model, interactive, playable media, social engagement.

1 Introduction

Advances in novel interfaces present a vast range of playable configurations for human players. Contemporary users and players adapt to the advances in many areas of engagement from the use of mobile phones to the play of games. Performance and artistic venues are one of the forefronts of advancing interface technologies. Experimental design of interface configurations can facilitate interactive *playability* in a media experience, an alternative to standard gameplay. Beyond the artistic achievement for look and feel, design must pursue meaningful interaction that sustains interest beyond initial attraction to the novelty of an interface and its artful constellation. An ongoing challenge for designing interactive media installations is their affordance of sustained engagement with systems and other players beyond their initial attraction. The present project inquiry is to identify and integrate the elements to sustain play and social engagement without extensive rules, scenarios, and explicit valorization mechanisms that big budget games offer. We introduce playable media with two novel interfaces, one physical and the other a software-based play interface. For this paper we focus on the latter. The play scenario involves no winning or losing

and there is no overarching goal dictating state transitions during play. The installation adapted an evolutionary algorithm into an interface function. Preliminary exploration of playability uses this interface configuration for group collaborative play. We refer to the project as *Wayfaring Swarms*.

1.1 Play Scenario

The *Wayfaring Swarms* play table is a rectangle of 36 inches by 48 inches, providing access for up to four players at one time. Using capacitive sensing the table surface can measure multiple players' hand positions. Players are invited to interact with swarm agents, which are animated graphics of small colored particles projected onto the table surface from above. Figure 1 illustrates the experimental infrastructure. Synthesized sounds are generated using algorithms that model musical patterns and transitions, and interactions with swarms cause changes in sound patterns. The swarm agents' dynamic properties exhibit emergent group behavior. Players may cooperatively gather agents or steer agents away from other players. The play surface displays a simple graphic map for a play sequence, a serial presentation of predefined play sections. Each section defines a swarm and accompanying sounds. Players simply play with each section and at some point they must decide to move on. Social interactions develop around this process.

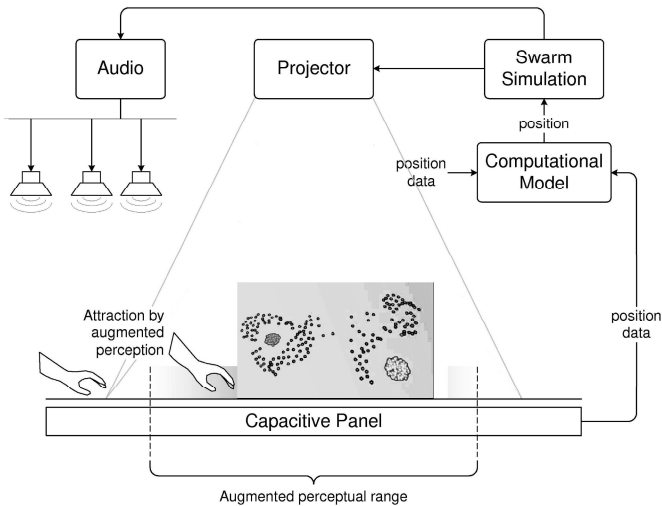


Fig. 1. *Wayfaring Swarms* playable system configuration

This simple play scenario is yet non-trivial for novice players as the play involves some learning curve to become attending listeners. Progression through the sequence of play sections is controlled by players using a simple signal of two hands pressed in a specific area of the capacitive panel. Players' determining when to change sections becomes either an awkward moment of unilateral action or an engaged consensual moment among players. Players can take new positions around the table, or leave and return without disrupting the interface functions.

The remainder of section 1 presents background and prior work. Section 2 introduces the original simulation and our modifications to enhance playability. Section 3 presents methodology for swarm feature extraction for application to sound models. Section 4 presents methodology for developing sound models to enrich play experience. Section 5 concludes with a narrative of informal observations projecting a future research scenario.

1.2 Background

Wayfaring Swarms advances an evolutionary interface as a viable application for sustained engagement for players. This interface implements a “breeding” model to generate a new swarm combining features of two existing swarms. The model is applied in relation to biological research for understanding social behaviors of the kind known as flocking behavior [1]; one of the simplest behaviors seen in nature revealing social and collective dynamics. It is described as self-organizing because the collective behavior is governed by a set of simple rules applied to each agent and there is no centralized control agent. It is also observed as exhibiting emergent behavior as its evolving patterns are unknown to each agent and there is no high level prescription dictating the resulting complexity. Mathematical modeling of swarms has been implemented in numerous versions and applied to many areas of study. The model usually combines Reynolds’ “boids” algorithm [2] and a self-propelled particle model [3]. These two methods combined make an excellent application to interactive and evolutionary play scenarios. While the boids algorithm provides microstructure of global patterns, the self-propelled particle model provides a lively oscillatory quality, and more importantly, an opportunity to introduce a high-level agent into the system, such as a human-controlled agent, as an influencing force to the global dynamics of swarms. This type of high-level influence evokes swarm behavior such as encountering predators [4], and provides a means to introduce human energy into the interactive pathway of swarm dynamics (see section 2.2).

1.3 Prior Work

Swarms and evolutionary algorithms are widely explored in areas of visual art and media, and developed to the extent of applications in commercial production including computer graphics, film special effects, and computer games, with use cases ranging from movements of crowds and armies to growth of vegetation-like scenic elements to genesis of game denizens (see Will Wright’s *Spore*). Interactive art examples include works by Sims [5], Sommerer and Mignonneau [6], and McCormack [7]. The Wayfaring Swarms system is not applied in this tradition. Swarm visualization serves not for advanced visual arts content but for basic and extensible capacity for listening and enactive movements not unlike those of musical performers (see section 5). The perceived kinesthetic energy of performers is an important aspect of musical reception [8]. Rudolf Laban has similarly articulated this through the theory of effort applied to dance performance [9][10]. Musical instruments are refined and fit to the human scale: the detailed interactive gestures between performers and their instruments are very intimate, not visible to audiences. However those gestures are perceived through the corresponding tones as musical expressions [11]. Our previous works on enactive interfaces explore various novel

configurations to “enlarge” performers’ tone producing gestures to make them accessible to observers [12]. Consistent with the previous works, the current project prioritizes the design criterion: configure the system so that players’ movements are clearly visible.

Wayfaring Swarms incorporates a touch free capacitive sensing panel. Leon Theremin originated capacitive sensing for music performance by direct touch-free tone control. The touch surface repertoire of interface paradigms opens other retrospective references to analog electronic interfaces with tone generators, such as systems constructed by Buchla [13] and Martirano [13], each of whom produced unique capacitive control surfaces and logic control gates for tactile transformation of sound synthesis. Following this line the Wayfaring Swarms interface engages upper body movements guided by human hands to interact with swarm agents projected on the playable surface. Unique from prior work, we introduce a layer of indirection from hand movements to tone control, through the swarm model. The interface also differs from the recently discontinued Lemur™ controller with animated physics-based graphic object trajectories. We are concerned not with graphic controllers but with 1) swarms’ emergent patterns, and 2) many to many mapping design between swarms and audible features.

2 Wayfaring Swarms Overview

Sayama [14] developed the swarm simulation used in Wayfaring Swarms. Sayama’s swarms are populated by agents exhibiting simple, semiautonomous movement rules in a continuous two-dimensional space. Each agent is assigned movement rules and a perceptual range for detecting the positions and velocities of other agents. Agents’ awareness within their perceptual ranges determines individual position updates with only decentralized control. The simple kinetic interactions among agents result in spontaneous large-scale pattern formation.

2.1 Agent Properties and Evolutionary Design of Swarms

Swarm agents consult all other agents in their perceptual range at each discrete time step. Agents can instantaneously change velocity according to the following rules, adopted from Reynolds’ “boids” system [1]:

Outside of an agent’s perceptual range: Straying:

- Agents move randomly if there are no other agents within perception range.
- Within an agent’s perceptual range:
- Cohesion: an agent moves toward the average position of local agents
 - Alignment: an agent moves towards the average velocity of local agents
 - Separation: an agent avoids collision with local agents
 - Whim: an agent moves randomly with a given probability
 - Pace keeping: each agent approximates its speed to its own normal speed.

Table 1 from Sayama enumerates kinetic parameters used to simulate agent behavior. Each agent i is assigned a set of unique values to define its dynamic properties. The

pixel is the atomic unit of spatial coordinates for agent position and movement. Tendency is an agent's rate of approximation of its current speed to its own normal speed. Maximum values were determined by Sayama heuristically and are arbitrary for implementation purposes. The total number of agents in a swarm is limited to 300.

Table 1. Kinetic parameters used to simulate agent behavior (explained in detail in Sayama [14])

Name	Min	Max	Meaning	Unit
R^i	0	300	Radius of local perception range	pixel
V_n^i	0	20	Normal speed	pixel step ⁻¹
V_m^i	0	40	Maximum speed	pixel step ⁻¹
c_1^i	0	1	Strength of cohesive force	step ⁻²
c_2^i	0	1	Strength of aligning force	step ⁻¹
c_3^i	0	100	Strength of separating force	pixel ² step ⁻²
c_4^i	0	0.5	Probability of random steering	—
c_5^i	0	1	Tendency of pace keeping	—

A set of these parameter values is referred to as a *recipe*. Multiple agents that share a common recipe are referred to as a *species*, and assigned a common color. Heterogeneous swarms are composed of multiple species. The emergent patterns of heterogeneous swarms are encoded in the agents' multiple sets of kinetic properties and the proportions of each recipe in the swarm. Sayama avoids automated fitness evaluation methods that would necessarily limit the diversity and novelty of potential outcomes. Evolutionary operators enable mutation of a single parent swarm by random re-sampling of up to 80% of the population size. Evolutionary design starting from two parent swarms will generate a new swarm by randomly determined ratios between all agents of both parents. The heuristic design process resembles musical improvisation, and is incorporated into the playability design of the Wayfaring Swarms system.

2.2 Creating a Playable Media Configuration

We embedded Sayama's simulation in a software and hardware environment designed for playability. To extend playability as social interaction, a large format capacitive panel is used for touch-free, multi-player, multi-point control. The panels were developed by Philippe Jean of Les Ateliers Numériques [15] for use in live performances by Cirque du Soleil. The Swarm graphics are projected on this surface so that players observe the social formations of swarms and interact with them by hand movements. The capacitive panel senses multiple hands as conductive objects in 1:1 ratio to the surface area; the maximum number of objects is determined by the surface area and objects' sizes. To align players' hands as play agents with swarm agents in simulation space, we bounded the simulation pixel region to the dimensions of the capacitive panel, and scale the projected swarm image to fit the capacitive play surface area. The projected image frame is calibrated with the capacitive surface by projecting markers in each corner of the frame, then touching the capacitive surface at

each marker point, and registering the touch points. Swarms reaching the edge of the capacitive surface are reflected from an invisible barrier and maintained within the playable area.

Hand position data is determined at the center of each area where a hand is detected, and transmitted to the corresponding position in the swarm simulation. Each hand position is represented in the simulation as a “super agent”. A super agent is directly controlled by a player’s hand, and is not influenced by the other kinetic rules. Swarm agents do not recognize super agents differently; they respond to super agents as they do to all other agents, by proximity-based kinetic rules. “Player control” is in this way an emergent property of a simulation where a control agent moves independently of the kinetic rules of swarm agents. Acting as super agents, players’ hands manipulate swarm shapes such as deformation and extrusion, separation and combination of multiple groups of agents. Performers engage a swarm’s emergent behavior but cannot directly manipulate agents’ positions or swarm formations independent of agents’ social relations. Figure 2 shows swarms gathered to the hands of four players.



Fig. 2. Three players around the playable swarm surface. A small robot with LEDs also plays.

Sayama’s original code was modified to isolate the simulation code from the frame loop of the swarm animation, and from the control flow of the graphical user interface. Then the isolated simulation code was embedded in an architecture for organizing multi-agent play sequences. Embedding the simulation included the addition of boundary conditions and calibration points mentioned above. The simulation was assigned an independent frame rate, required as the graphics rate cannot run fast enough to support parallel audio synthesis. The new software architecture simplified the routing of hand position data to the simulation, and enabled the specification of the play sequences described in Section 1.1.

3 Methodology: Swarm Feature Extraction Applied to Sound Models

The integration of the evolutionary interface requires a methodology to extract and use the swarm state information to enrich play experience for human players. Players in Wayfaring Swarms are intimately aware of swarm dynamics through graphics display and also through continuous auditory feedback. To provide an auditory feedback sound synthesis is applied to sonify the swarm state information. One of the criteria for choosing feature extraction techniques for sound representation was to complement the visual representation of swarm dynamics rather than duplicating it. Control strategies for sound models are adapted to use data of features extracted from emergent behaviors in swarm simulations. To test these adaptations, mappings are made between a set of emergent swarm states and a set of synthesis parameter states. The design decision associates the selected states of a sound model to selected states of a swarm. Thereafter other patterns that emerge in the swarm will generate corresponding sound patterns. Data selection is based on salient feature analysis. The process of establishing initial correspondences may be thought of as “tuning” the interface. Tuning in this sense is calibrating the relationship between swarm dynamics and sound dynamics. As an example, positional data of swarm clusters are used to localize positions of sound sources. The following discusses the procedure for swarm feature recognition and extraction in preparation of control structure to apply to sound models.

The use of swarm patterns as sound control data presents a significant challenge because in the code there are no numerical representations of the patterns that can be readily applied to sound models. The swarm simulation does not internally represent emergent patterns in classes of control parameters or in feature data variables. Emergent behavior from complex systems has been referred to as the result of *unspecific control parameters* [16]. As system parameter values change the resulting patterns vary, but variation and pattern emergence are not classifiable using systematic, linear representations. Patterns recognized by human observation are not represented in the simulation itself. Instead we extract data from visible features of swarms, and apply the data to sound models to enrich the playability.

3.1 Recognizing Clusters

Swarm denotes the total agents in a simulation. *Cluster* refers to a visibly coherent aggregate of agents. The relevance of cluster formation is that agents in a cluster are responding to mutual proximity, whereas agents in separate clusters are mutually unaware unless the clusters are in close proximity. Clusters are a primary feature to recognize and measure: they are emergent and temporary. Their spontaneous subdivisions and formations provide a highly configurable and playable dynamic. Players tend to focus attention on clusters and how to merge them or separate them, as well as moving them across regions in the interface.

Clusters are independent of species and recipes: in a swarm composed of multiple species, a cluster may be heterogeneous or homogenous. Membership of agents in clusters changes over time, so clusters are identified and tracked only by persistence. We examine the position of each agent at each time step and compare it to the positions of all other agents. A proximity threshold determines when an agent is a member of a cluster or a non-member roaming between clusters. An agent may only be a member of one cluster at each time step. For all agents in a common cluster we determine the average center position and provide this data for use in sound synthesis control. Shape is not a consideration in identifying a cluster. When two clusters' agents have sufficiently close proximity they are considered merged, regardless of shape. Clusters are identified by integer; a cluster keeps its number while it persists; when two clusters merge the lower number prevails, and the higher number is returned to the pool for re-use. The maximum number of simultaneous clusters recognized is set by a run time parameter. A cluster must have at least six agents; this limit is set by a run time parameter. We transmit cluster membership size for sound control; we track but do not transmit data of agents' individual cluster memberships.

3.2 Temporal Variation of Emergent Features

The variation of clusters over time creates challenges in applying cluster data as media control signals. Clusters are created by subdivision of larger clusters, and terminated when a cluster breaks into many pieces or when its agents become members of a larger cluster. The transitions at the creation or termination of a cluster often require multiple time steps before the stability of the new state can be ascertained. A time window confirmation parameter is used to track the number of consecutive time steps that a new cluster state is maintained. Initial emergence or disappearance of a cluster is flagged as the first frame of the time window. If the state is continuously present over the time window duration, a confirmed cluster state is reported. A duration threshold acts as a smoothing function to prevent rapid-fire series of messages of alternating cluster states. Alternation can occur when two clusters skirt one another and their perimeters temporarily overlap, or when a larger cluster is pulled apart to form new clusters.

Time windowing introduces unwanted latency in the transmission of cluster data to sound control. Latency undermines synchronization between swarm visualization and corresponding sounds. Latency is imposed by the cluster confirmation time window parameter set at run time. At minimum two frames are required to confirm cluster formation; a simulation frame rate of 50 frames per second provides 25 Hz latency multiplied by the number of frames in the confirmation time window. In practice when two clusters merge or when one cluster divides, two data streams must be managed, one appearing or disappearing, and the other persistent but changing in number of members. The varying size and behavior of the persistent cluster, as well as data of a new cluster, will impact corresponding sound. Cluster history is determined by tracking agent membership across consecutive time steps; history preserves coherence of sounds when a cluster divides or when two clusters merge. Cluster history tracking is applied to a cluster data stream in order to

terminate when it is absorbed into a larger cluster. Cluster history also provides a reference for smooth transition when a data stream bifurcates upon a cluster's subdivision.

3.3 Measuring Emergent Shape

A cluster's expressive features are shape and internal distribution of agents. Rings, elongations, "dumbbell" or "twin star" shapes, and internal rotation patterns are often prominent features. These emerging patterns are not represented in the simulation and must be detected as features by measuring the positions and velocities of swarm agents. In distinction to shape recognition, we determined the essential approach responds to displacement or perturbation. We arrived at this approach by observing that clusters do not achieve a wide range of shapes in terms of geometric primitives, and clusters cannot be forced into shapes other than their stable and emergent properties. We determined to tune sounds in a range corresponding from stable or symmetrical clusters to unstable or distorted clusters. This approach was selected rather than tuning sounds for target shapes unrelated to a cluster's inherent properties. We adopted this initial approach from Sayama's decision to avoid the use of fitness evaluation methods (see page 4 and [14]).

As proof of concept our prototype applies simple statistical measures to explore the application of perturbation recognition. Separate statistics are provided as sources of sound control data: 1) measured across the entire swarm, 2) measured by species regardless of cluster, and 3) measured by cluster. We measure average velocity; average distance from the statistical center; average distribution angle in radians; and average planar coordinate positions on the play surface. We also provide a histogram of each measure to show distribution of agents across the full ranges of these dimensions. For example, while most unperturbed clusters are circular in shape, their symmetry is distorted by interactions with players and with other clusters. Distortions of shape are easily visible features and are detected in uneven histogram distributions of agents' positional angle and velocity.

Figure 3 provides a sequence showing a separation of one cluster into two. The cluster is heterogeneous having two species forming a ring pattern. Figure 3a shows a player's hands bringing about the separation, which requires 8 to 12 seconds to complete. Figure 3b-3d shows the sequence of separation and the corresponding statistics and histogram data, displayed in a companion diagnostic tool, with multiple histograms overlaid by color. The red bars indicate average velocity; green indicates angle of distribution around the center; blue indicates distance from the center; and yellow and pink are average horizontal and vertical position respectively. From Figure 3b to 3d the cluster separation is reflected in the histogram of average velocity (red) and distribution angle (green). Not easily visible in Figure 3d, the distance-from-center histogram (blue) indicates the separation of species in the ring structure. In Figures 3b and 3c the visualization includes a small circle at the center of the ring formation. This circle is a diagnostic of the cluster center as determined by the feature detection system. In Figure 3d two circles denote two clusters are detected.

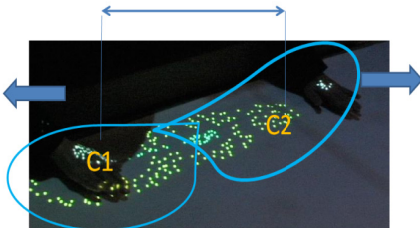


Fig. 3a. Separating one cluster into two

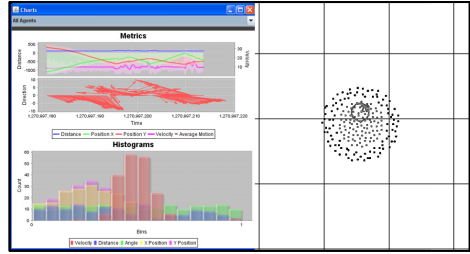


Fig. 3b. Histogram of stable cluster

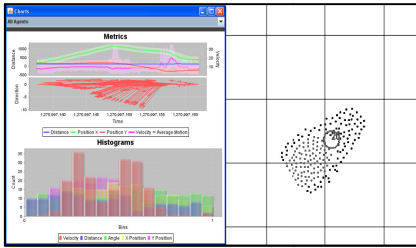


Fig. 3c. Data divides with cluster deformation

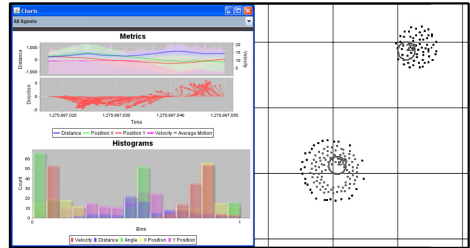


Fig. 3d. Data and clusters separated

4 Methodology: Play Scenario with Sound Model

Wayfaring Swarms extends swarms’ social behavior as a paradigm for playability and players’ interactions with musical play agents. The abstract representation of sounds by swarm graphics is reminiscent of abstract musical notation, not in level of detail as to specify pitch or duration as standard musical notation, but rather in the capacity to represent diverse sounds and transformations. Initial conditions are asserted by tuning swarm recipes to states of sound generators. Swarms’ relationships to sounds are relative to the initial tunings and the corresponding ranges of transformations. Ranges of sound transformations are designed to correspond to player’s actions inducing emergent properties of swarms.

4.1 Sound Model: Synthesis Methods

Sound sources are synthesized using wavetable lookup, physical modeling [17], and formant-modeling techniques [18]. Sound control parameters are derived not directly from graphics, rather from non-visual data of the agents’ states and tendencies. These states are reflected in the graphics and in parallel in the sounds, resulting in multi-modal emergent behavior. Sound authoring is applied to control sounds using the VSS sound server [19].

Sound localization and spatialization are added downstream. Sound sources are distributed to a multi-channel speaker array that surrounds the play area. Swarm clusters’ positions are used to determine sound sources’ localized positions in a

simulated acoustic environment. The simulated auditory field is larger than the physical play area, with proportional dimensions. Clusters in the center of the play area generate sounds located directly above the play surface. Clusters at the periphery of the play area generate sounds heard at simulated distances behind the players. This acoustic model encompasses players within the spatial field of the play area.

4.2 Sound Model: Designing Coherent Transformations

Several techniques were developed to enable reliable correspondences of sounds with highly variable behaviors of clusters. (1) Local deformations of clusters were used uniformly to modify formant characteristics of sounds. This technique may be applied to many classes of sounds. It involves modifying the vowel-like qualities of “openness” and “tightness” of a sound. With this technique the effects of a player’s hand deforming a cluster are immediately reflected in the local tone quality of a sound, without disrupting the composed pitch and rhythmic structure of the sound. (2) Data related to changes of cluster size and velocity is assigned equally to pitch-related and rhythm-related sound properties. This technique is preferable to trivial associations involving isolated audio or musical parameters that co-vary linearly with cluster data. (3) Sound sources are not associated one-to-one with clusters. The number and timing of cluster instances is highly dependent upon local performance actions. Creating a sound source for each cluster would be capricious from a sound modeling standpoint. Clusters are local variations and are not structurally analogous to the composition of sound sources. Instead the designation of sound sources is determined by the scheduling of swarm species in sections of the play sequence. To reflect the local formation of clusters, data from cluster instantiation and termination is applied through the spatialization of sound sources, which emulates a musical technique known as *antiphony*, the exchange of musical ideas from one sound source location to another.

5 Concluding Narrative: Informal Observations of Initial Results

Preliminary informal observations involved groups of one to five players, including college students, technical staff, and faculty, of both genders with diverse cultural and racial backgrounds. Players freely used one or two hands and played for a 20-minute session. Players’ lively social interactions have been noted. Whether shy or disinterested or reluctant to play initially, once they “get the hang of it” players tend to explore diversity of local dynamical patterns. In general players respond to emergent behaviors of swarms. Players verbally refer to agent clusters as individuals, and also refer to single agents as individuals. Utterances attributing personality traits to agents and clusters are regular.

Players exhibit modes of play behaviors that might be roughly grouped in three types: random, exploratory, and ensemble-like modes. The random case is when players either just joined the group or seem lost in the middle of play. When they seem lost, some players tend to step aside and watch others while some players tend to try out random positions on table. The exploratory case is when players tend to

investigate clusters by perturbing and diverting while attending to changes in the sounds. In this mode, play behaviors seem to be uncoordinated as players focus on their individual investigation. The ensemble-like case is when players are attending to cluster merges and separations as a goal. In this mode, play behaviors tend to be coordinated to stated tasks regarding sound as feedback. However it is noted that these three modes are not necessarily progressive in a linear way. Players tend to switch mode regardless of their level of experience with the installation. This may be due to the evolving nature of the interface, suggesting a future research direction to investigate the relationship between the measure of the swarm state statistics and the players mode switching. Orientation of neighboring players also has an influence on play mode.

Listening to sounds appears to add a level of social interaction among players beyond the sharing or stealing of clusters. This observation is noted mostly from body language of largely unspoken cues for exchanging or managing clusters, which suggests a future research direction. Some players appear initially reluctant to interact, and this caution may be heightened by unfamiliarity with the sound textures and uncertainty how the sound may change if the swarm is perturbed. Some players do not wish to be responsible for “breaking something” or “making a noisy sound”.

All observations are subject to formal study. The project presents a promising setup for two kinds of controlled formal studies for playability: 1) Investigation of correlation between swarm states and play mode switching and 2) Investigation of differences in play behaviors when the evolutionary interface is presented with graphics alone and with both graphics and sounds.

Acknowledgement. We wish to thank Hiroki Sayama for sharing his code base to develop the playable implementation. Arthur Peters debugged the capacitive surface and data transmissions and coded the playable simulation version and authoring environment. John McCullough constructed the performance platform. Michael Chladil mounted the projection system and aided the overall implementation, and calibration.

References

1. Sayama, H.: Teaching emergence and evolution simultaneously through simulated breeding of artificial swarm behaviors. In: Proceedings of the Sixth International Conference on Complex Systems, ICCS 2006 (2006), <http://neCSI.org/events/iccs6/proceedings.html>
2. Reynolds, C.: Flocks, herds and schools: A distributed behavioral model. In: SIGGRAPH 1987: Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques, pp. 25–34. Association for Computing Machinery (1987)
3. Vicsek, T., Czirok, A., Ben-Jacob, E., Cohen, I., Shochet, O.: Novel type of phase transition in a system of self-driven particles. *Phys. Rev. Lett* 75, 1226–1229 (1995)
4. Reynolds, C.: <http://www.red3d.com/cwr/boids/> (accessed on April 15, 2011)
5. Sims, K.: Evolving Virtual Creatures. In: SIGGRAPH 1994: Proc. 21st Annual Conference on Computer Graphics and Interactive Techniques, pp. 15–22. Assoc. of Computing Machinery (1994)

6. Sommerer, C., Mignonneau, L., Stocker, G.: *Christa Sommerer and Laurent Mignonneau-Interactive Art Research*. Springer, New York (2009)
7. McCormack, J.: *Impossible Nature*. Australian Centre for the Moving Image (2004)
8. Choi, I.: Interactivity vs. Control: Human-Machine performance basis of emotion. In: Camurri, A. (ed.) *Kansei, the Technology of Emotion (Proceedings of the AIMI International Workshop)*, pp. 24–35. Associazione di Informatica Musicale Italiana, Genoa (1997)
9. Camurri, A., Trocca, R., Volpe, G.: Full-body movement and music signals: an approach toward analysis and synthesis of expressive content. In: *Proc. Intl. Workshop on Physicality and Tangibility in Interaction: Towards New Paradigms for Interaction Beyond the Desktop, CEC-I3, Siena (1999)*
10. Newlove, J.: *Laban for Actors and Dancers: Putting Laban's Movement Theory into Practice*. Nick Hern Books, London (1993)
11. Wanderley, M., Depalle, P.: Gestural Control of Sound Synthesis. In: Johannsen, G. (ed.) *Proceedings of the IEEE Special Issue on Engineering and Music - Supervisory Control and Auditory Communication*, vol. 92(4), pp. 632–644 (2004)
12. Choi, I.: Gestural Primitives and the context for computational processing in an interactive performance system. In: Battier, M., Wanderley, M. (eds.) *Trends in Gestural Control of Music*. IRCAM, Paris (2000)
13. Roads, C., Strawn, J.: *Foundations of Computer Music*. MIT Press, Cambridge (1988)
14. Sayama, H.: Decentralized Control and Interactive Design Methods for Large-Scale Heterogeneous Self-Organizing Swarms. In: Almeida e Costa, F., Rocha, L.M., Costa, E., Harvey, I., Coutinho, A. (eds.) *ECAL 2007. LNCS (LNAI)*, vol. 4648, pp. 675–684. Springer, Heidelberg (2007)
15. Les Ateliers Numériques, <http://www.ateliers-numeriques.net/index-1an.php> (accessed on April 15, 2011)
16. Haken, H.: *Synergetic Computers and Cognition*. Springer, Berlin (1990)
17. Cook, P.: *Real Sound Synthesis for Interactive Applications*. A K Peters, Natick (2002)
18. Chowning, J.: Frequency Modulation Synthesis of the Singing Voice. In: Mathews, M., Pierce, J. (eds.) *Current Directions in Computer Music Research*, pp. 57–64. MIT Press, Cambridge (1989)
19. Bargar, R., Choi, I., Das, S., Goudeseune, C.: Model-based interactive sound for an immersive virtual environment. In: *Proceedings of the 1994 International Computer Music Conference*. International Computer Music Assoc, Aarhus (1994)

Social Interaction in a Cooperative Brain-Computer Interface Game

Michel Obbink, Hayrettin Gürkök, Danny Plass-Oude Bos, Gido Hakvoort, Mannes Poel, and Anton Nijholt

Human Media Interaction, University of Twente, Enschede, The Netherlands
{mobbink,gido.hakvoort}@gmail.com,
{h.gurkok,d.plass,m.poel,a.nijholt}@cs.utwente.nl

Abstract. Does using a brain-computer interface (BCI) influence the social interaction between people when playing a cooperative game? By measuring the amount of speech, utterances, instrumental gestures and empathic gestures during a cooperative game where two participants had to reach a certain goal, and questioning participants about their own experience afterwards this study attempts to provide answers to this question. The results showed that social interaction changed when using a BCI compared to using a mouse. There was a higher amount of utterances and empathic gestures. This indicates that the participants reacted more to the higher difficulty of the BCI selection method. Participants also reported that they felt they cooperated better during the use of the mouse.

Keywords: brain-computer interfaces, social interaction, games, cooperation.

1 Introduction

A brain-computer interface (BCI) is a means of interaction between humans and computers based on neural activity in the brain. It has fascinated people as it could enable whole new ways of controlling objects such as computers or wheelchairs. Since it has come into existence BCI research has mostly focused on helping disabled people, for example by controlling a wheelchair [12] or by helping them to communicate with the outside world through a word speller application [5].

Studies are currently considering applications for healthy users as well. Possibilities are applications such as virtual environment controllers [1] and games [11]. An advantage of games is that when one is integrating BCI into a game one could turn a disadvantage, the lower accuracy that is associated with BCI, into a challenge that the gamer has to master [10]. This challenge could trigger a whole new genre of games where mastering your brain waves is pivotal.

One of the current main problems in BCI research is moving BCI out of the laboratory setting into the everyday environment. For BCI to perform well in

normal situations, it has to perform when there is background noise, for example when the user is engaged in multiple tasks or when the user is collaborating with other people. A drawback of BCI is that equipment for data acquisition, such as electroencephalographs (EEGs), is very sensitive to noise. Muscle movement of the person using the BCI equipment or electrical interference might result in artifacts in the signal. As muscle movements generate artifacts users might be less inclined to interact socially with each other for worry of decreasing BCI performance. This will have consequences for cooperative applications if social interaction between users is proved to be substantially impeded.

This study looks into the influence of BCI control on social interaction in a cooperative game setting. To cooperate with each other, users should be able to interact with each other unimpeded. To study this social interaction, an environment has been setup where a player can use either a BCI or a mouse. The task was comprised of the selection of objects. This means that a BCI could be tested against a normal point and click interface with the mouse. For the BCI selection method the classification method steady-state visually evoked potentials (SSVEPs) [14] is used. This is a method that uses a flickering stimulus to activate the part of the brain where visual information is processed. When showing a group of stimuli, the player can make a selection by looking at one of the stimuli. The different stimuli each flicker on a different frequency, in such a way the stimulus that the player focuses on can be distinguished from the others. By looking at the speech, utterances, instrumental gestures and empathic gestures that players produce while playing the game the influence of BCI on social interaction was analysed.

The second section of this paper describes how to induce and measure social interaction. The SSVEP method that is used during the experiment is explained as well. The third section discusses the methodology and the game. The fourth section presents the results and in the fifth section these are discussed. Section six finishes with the conclusion and possible future work.

2 Background

2.1 Inducing Social Interaction

The first concern in social interaction research is to induce the interaction among users. According to Fowler et al. [6] and Clark [4] language is used as a coordination device, a way by which coordination among two or more individuals can be achieved to reach a common goal, or as Clark calls it: joint actions. According to Fowler et al. several studies have observed that humans have a tendency to cooperate and sometimes even imitate behaviour such as gestures, posture and verbal language. This suggests that while two users work together on a system towards the same goal, they will inherently interact with each other.

2.2 Measuring Social Interaction

Lindley et al. [7] measured the engagement and social behaviour of people playing a game together. The game was Donkey Konga, which could be played with

a conventional controller and with special bongos that required the users to tap the bongos and clap their hands to the beat of the music. They treated a pair of participants as a single unit, as they did not see an individual independent from its partner. They used definitions from the Autism Diagnostic Observation Schedule (ADOS) [9] to code verbal and non-verbal behaviours. Verbal behaviour was either categorized as speech or utterances. They repeated the procedure for non-verbal behaviour, categorizing them between instrumental gestures and empathic gestures. Instrumental gestures are actions that convey a clear meaning, or are used to draw/direct attention. Gestures that could be in this category are: pointing, shrugging, nodding and moving head towards the other person. Empathic gestures are actions that convey emotion, such as placing hands in front of the mouth in shock or resting their chin on a hand. With the bongos the participants produced significantly more utterances, instrumental and empathic gestures. They showed that an alternative game controller such as the Bongos, makes participants produce more social interaction. This research is highly comparable to the current study and therefore comparable measurement methods were used. With the four categories of verbal and non-verbal behaviour all possible events were captured and by looking at the time for speech it provides a method of measuring social interaction.

2.3 Steady-State Visually Evoked Potentials

The SSVEP response is triggered when an user focusses on a stimulus that is flickering at a certain frequency. The SSVEP response is mostly visible between 6 Hz to 18 Hz and is recorded from the occipital region of the scalp [14]. Because the power of an SSVEP response shows only over a very narrow bandwidth that corresponds to the frequency of the stimulus [8], it is detectable with a fast-Fourier transform (FFT). SSVEP is an exogenous event-related potential (ERP), which means that it is an involuntary brain response to an external stimulus and these occur due to internal processing of external events.

An important issue that arose when building the SSVEP system was the set of frequencies that were used and how this was presented to the user. The work of Volosyak et al. [15] present a set of possible frequencies that could be used on an LCD screen. In a small pre-experiment trial performed with 7 participants every combination of their proposed frequencies were tested to select the three frequencies that were used in this study. With an average recall of 84.6% ($\sigma = 11.9$), the set of 7.5, 10 and 12 Hz was selected to be used.

3 Methodology

3.1 Participants

For this study 20 participants divided into 10 pairs, were tested. All participants were asked to bring a friend. If no friend was available they were teamed up with another participant. Pairs did not have to be equal in composition, because all

the pairs performed each selection method and therefore if the composition of a pair had influence on the interaction, it had any influence on all methods and therefore it had no effect on this study. The participants participated voluntarily in this study, and signed a consent form for their participation. To motivate the pairs to do their best a small reward, a pair of cinema tickets, was promised to the pair that completed the experiment in the shortest time. The average age of the participants was 25.25 ($\sigma = 7.20$) with the youngest being 18 and the oldest 54, of the 20 participants 18 were male. Each participant had a normal, or corrected to normal eyesight, used a computer every day and at least some experience with computer games. None of the participants reported a history of epilepsy.

3.2 The Game

The game used in this study consisted of a playground representing a meadow (Figure 1). On this playground there were a few obstacles such as fences and vegetation and a pen. The top-down view gave the participants the ability to plan around the obstacles, and communicate their plans to each other. The playground was populated with three herding dogs and several sheep depending on the task. The goal of this game was to get all the sheep into the pen in the shortest time by giving the dogs movement instructions. By setting a goal that participants had to reach, they had something to work towards together.



Fig. 1. A screenshot of the game containing 10 sheep and 6 dogs controlled by the players

To move the dog, the participant first moves his mouse cursor to the location the dog should move to. The participant presses and holds the left mouse button. From this moment the SSVEP method is active for the dog selection and the dogs are all highlighted with different frequencies. The participant selects the dog that has to move by looking and concentrating on the blinking stimulus of the dog that should move. As the participant holds down the mouse button the SSVEP method continues to acquire more samples over time. SSVEP detection

has a higher accuracy over time, provided the attention of the participant is kept constant. On the other hand the participant might choose to release the mouse button sooner if a quick reaction is needed, but this decreased the chance of the correct dog being selected. So the trade-off between performance and reaction speed is up to the participant to make. If all went successfully, the correct dog moves to the location of the mouse cursor as soon as the button is released, if not a wrong dog moves to the indicated location. During the SSVEP stimulation the participant can still move the mouse cursor, altering the location the selected dog should move to.

The point and click method worked by first clicking the mouse on the dog that the participant wants to use. Once the dog is selected a small circle surrounds it as an indication of the selection. Now the participant can click on the location the dog has to move to and the dog starts moving.

3.3 Experimental Setup

The setup consisted of five computers: two for the participants to play on, two for the BCI acquisition and one for the recording and storing of audiovisual data. The participants were seated next to each other, as can be seen in Figure 2, so non-verbal interaction such as pointing was possible while playing the game. They both looked at their own LCD screens that were placed 50 cm apart from each other. This gave the participants the opportunity to turn their heads and look at each other's screen. As they had some freedom of movement and could move forward or backwards in their chairs there was no fixed distance from participant to the screen. Any movement or speaking might have impaired the accuracy of the SSVEP classifier due to muscle noise which might have lead to artefacts in the data, but it enabled them to communicate more easily at will. The participants were notified in advance that this might be the case, but they had to decide for themselves if they heeded this notification or not. The BCI caps were placed at the start of the experiment and removed at the end of the experiment. A camera and microphone were pointed at the participants as can be seen in Figure 2.

Each pair started with a short training to learn the game and the two different selection methods. Once the training was finished they played two trials of the game, once with the SSVEP selection method and once with the point and click method. Each trial took until they finished the task or a time limit of 20 minutes had passed. Each trial was played on a pre-made map. However, the layout of these maps differed, because if the same map had been used for both trials the pair might have developed a strategy on the first map and deployed it again on the second map without having to discuss this. Thereby the social interaction of the latter trial may be influenced. The maps that were used for both methods therefore differed mainly on layout and obstacles. The combination of map and selection method was selected by counterbalancing each trial. During the whole procedure the experimenter stayed in the same room.



Fig. 2. One participant pointing with one hand and clenching his fist while the other participant is looking on and holding his hand flat on the tabletop

Once the experiment was completed the BCI caps were taken off and the participants were asked to fill in a questionnaire. The questionnaire asked them to think about the cooperation within the pair and rank both selection methods based on how they experienced it. It also asked them if they felt inclined to work together at all, to validate the setup of the experiment and it asked how much difficulty they had selecting a dog with each method. This might provide some correlation between difficulty and certain behaviours that were measured. Finally, the participants were interviewed about their ranking of methods in the questionnaire. By doing an interview with the participants, more information could be gathered than by asking this in the questionnaire.

3.4 Data Acquisition, Processing and Analysis

The SSVEP selection method used EEG signals that were acquired with a Biosemi ActiveTwo system, from five electrodes $PO3$, $O1$, O_z , $O2$ and $PO4$ placed according to the 10-20 international system [13]. This data was digitized at 512 Hz sample rate, re-referenced to electrodes placed on the earlobes and analysed using Canonical Correlation Analysis (CCA) [2]. CCA has advantages over the commonly used power spectral density analysis (PSDA) method introduced by Cheng et al. [3], such as a better signal-to noise ratio and no need for channel selection. CCA tries to correlate the BCI signal to a set of reference signals based on the frequencies that are used. The frequency with the highest correlation to the reference signals is selected.

The videos were annotated manually with the four behaviours that Lindley et al. [7] defined. These were speech, utterances, instrumental gestures and empathic gestures. Speech is the deliverance of formal spoken communication while utterances are all other sounds that were made by participants. Instrumental gestures are gestures that have a deliberate purpose to support cooperation, such as pointing and gazing to the others monitor. Empathic gestures are gestures that may convey the emotional state of a participant. Obvious gestures that could be thought of are gestures such as putting a hand in front of your mouth in shock, or more subtle such as increased repetitive, purposeless movement.

Every speech and utterance component in the audio data was marked from start to finish. The total length of both speech and utterances that participants produced per trial was used for analysis. These values were normalized to a number of seconds of either per minute, because all pairs finished in different times. A pair was considered as a single unit, thus this data was averaged over the pair. The same was done with instrumental gestures and empathic gestures. These were counted after the annotation. The total number of gestures per trial for both was normalized to a number of gestures per minute for each pair. Finally all these values were averaged over all pairs and for each of the selection methods to see the differences.

In the questionnaire participants were asked to rank the selection methods based on the level of cooperation the participants experienced. In a 7-point Likert scale they were asked if they felt the need to cooperate during the experiment to measure if this study was successful in inducing interaction between participants and about the difficulty of selecting the dogs with each method.

4 Results

Before the results are analysed, it is important to see if this study was successful at inducing interaction between participants. An item in the questionnaire asked whether the participants felt inclined to work together. Using a 7-point Likert scale 20 subjects answered with a mode of 7 (9 out of 20 answered with a 7). Testing these results with a Wilcoxon signed-rank test to a neutral result, with an average of 4, yielded $Z = -3.9811, p < 0.001$. Therefore it can be concluded that the experiment was successful in inducing cooperation within the pairs.

Table 1. An overview of all average values, and standard deviation within parentheses, over all the pairs for each of the behaviours for both the selection methods. For speech and utterances these values are in seconds per minute and for instrumental and empathic gestures these values are number of gestures per minute.

	BCI selection	Point and Click
Speech	6.43 (2.92)	7.56 (3.70)
Utterances	1.78 (0.63)	1.18 (0.51)
Instrumental gestures	0.27 (0.28)	0.41 (0.49)
Empathic gestures	1.81 (0.70)	1.21 (0.80)

In table 1 the average values over all the pairs for all the four behaviours and both selection methods are shown. There was a higher number of speech and instrumental gestures during the use of point and click selection, and a higher amount of utterances and empathic gestures during the use of BCI. BCI tasks took on average 9.64 minutes ($\sigma = 5.85$) to finish while point and click tasks took on average 8.12 minutes ($\sigma = 5.07$) in seconds to finish. This was however not a significant difference as the deviation between pairs was very high.

Using a Wilcoxon signed-rank test ($p = 0.0645$) shows that there is a potential trend, but no significant difference between the amount of speech with BCI and point and click, but ($p = 0.0059$) on utterances, it shows that when using BCI significantly more utterances were produced compared to using point and click. There was no significant difference between BCI and point and click for instrumental gestures ($p = 0.3223$). Looking at emphatic gestures, there are clearly significantly more gestures used while playing with BCI ($p = 0.0039$) compared to point and click.

5 Discussion

It was expected that due to the focus that was required for selecting a dog, and the participant's knowledge that speech and movement might disturb the EEG signal during the use of BCI selection, the amount of speech and the number of instrumental gestures would be lower. As cooperation is mostly done by speech and instrumental gestures it was expected that cooperation between participants would be reduced as well. When a wrong dog is selected it causes an unexpected situation, this triggers involuntary reactions from the participants in the form of utterances and empathic gestures. The amount of utterances and the number of empathic gestures were expected to be higher.

The participants indicated in the questionnaire that they found selecting with BCI more difficult than with point and click ($Z = 4.7013, p < 0.001$). However, no significant difference was found between point and click and BCI for neither speech nor instrumental gestures. For speech there was a trend towards significance.

The amount of speech and number of instrumental gestures did not change with the selection methods. In the questionnaire, participants were asked to rank how they thought they cooperated between different selection methods. They answered 17 out of 20 times that they cooperated better during the use of the point and click selection method. This was also supported by some of the participants who voiced this during the interview afterwards. They said that at times they were too busy focusing on selecting the right dog and they did not pay much attention to what the other person was doing. Further research with additional participants could reduce the effect of such an outlier that was found in the speech condition and provide proof with a significant difference.

There was a significant difference between point and click and BCI with utterances and empathic gestures. This shows that some aspects of social interaction do change with different selection methods. There were more laughs, groans, interruptions of speech and other sounds made during a BCI played game and here was a higher number of empathic gestures as well. This increase in the number of empathic gestures and amount of utterances means that more unexpected events happened that the participants reacted on. These events are mostly the selection of a wrong dog and implies the difficulty of the BCI selection. This does not mean that they produced less cooperation, but it was influenced as they had to adapt to new situations when a wrong dog was selected.

The results from the audiovisual data indicate aspects of social interaction are affected by the higher difficulty and effort needed for BCI. The questionnaire and the interview support this, and indicate that the use of BCI noticeably influences the cooperation between participants in such a way that they cooperated better during the use of point and click.

6 Conclusion

This study looked at the social interaction and cooperation during a cooperative multi player game. A comparison was made between BCI selection compared to point and click selection. Measurements were taken from: audiovisual tracks, questionnaires and an interview. The audiovisual tracks were annotated marking the duration of speech and utterances, and the number of instrumental and empathic gestures. This experiment resulted in no significant difference in the amount of speech or the number of instrumental gestures, but there was a trend towards more speech when using point and click. There was a significantly higher amount of utterances and number of empathic gestures when using BCI compared to using point and click. This indicates that aspects of social interaction are affected by the use of BCI. The information provided by the questionnaire indicate this is caused by the difficulty of BCI selection and influences the cooperation in such a way that participants cooperated better during the use of point and click.

For future work it would be interesting to look deeper into the annotation and label each utterance and empathic gesture individually. This could provide more information on what kind of utterances and empathic gestures are more common during BCI. This would show for example if participants laugh or groan more during BCI.

Acknowledgements. The authors gratefully acknowledge the support of the BrainGain Smart Mix Programme of the Netherlands Ministry of Economic Affairs and the Netherlands Ministry of Education, Culture and Science. The authors would also like to thank Michiel Hakvoort for his technical support on the game and Lynn Packwood for improving the language of this paper.

References

1. Bayliss, J.: Use of the evoked potential P3 component for control in a virtual apartment. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 11(2), 113–116 (2003)
2. Bin, G., Gao, X., Yan, Z., Hong, B., Gao, S.: An online multi-channel SSVEP-based brain-computer interface using a canonical correlation analysis method. *Journal of Neural Engineering* 6(4), 46002 (2009)
3. Cheng, M., Gao, X., Gao, S., Xu, D.: Design and implementation of a brain-computer interface with high transfer rates. *IEEE Transactions on Biomedical Engineering* 49(10), 1181–1186 (2002)
4. Clark, H.H.: *Using Language*. Cambridge University Press, Cambridge (1996)

5. Farwell, L., Donchin, E.: Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and Clinical Neurophysiology* 70(6), 510–523 (1988)
6. Fowler, C., Richardson, M., Marsh, K., Shockley, K.: Language use, coordination, and the emergence of cooperative action. In: *Coordination: Neural, Behavioral and Social Dynamics*, pp. 261–279. Springer, Heidelberg (2008)
7. Lindley, S.E., Le Couteur, J., Berthouze, N.L.: Stirring up experience through movement in game play: effects on engagement and social behaviour. In: *CHI 2008: Proceeding of the Twenty-Sixth Annual SIGCHI Conference on Human Factors in Computing Systems*, pp. 511–514. ACM, New York (2008)
8. Lopez, M., Pelayo, F., Madrid, E., Prieto, A.: Statistical characterization of steady-state visual evoked potentials and their use in brain-computer interfaces. *Neural Processing Letters* 29(3), 179–187 (2009)
9. Lord, C., Risi, S., Lambrecht, L., Cook, E.H., Leventhal, B.L., DiLavore, P.C., Pickles, A., Rutter, M.: The autism diagnostic observation schedule-generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders* 30(3), 205–223 (2000)
10. Nijholt, A., Reuderink, B., Oude Bos, D.: Turning Shortcomings into Challenges: Brain-Computer Interfaces for Games. In: Nijholt, A., Reidsma, D., Hondorp, H. (eds.) *INTETAIN 2009. LNICST*, vol. 9, pp. 153–168. Springer, Heidelberg (2009)
11. Plass-Oude Bos, D., Reuderink, B., Laar, B., Gürkök, H., Mühl, C., Poel, M., Nijholt, A., Heylen, D.: Brain-computer interfacing and games. In: *Brain-Computer Interfaces*, pp. 149–178. Springer, London (2010)
12. Rebsamen, B., Burdet, E., Guan, C., Zhang, H., Teo, C.L., Zeng, Q., Ang, M., Laugier, C.: A brain-controlled wheelchair based on P300 and path guidance. In: *The First IEEE/RAS-EMBS International Conference on Biomedical Robotics and Biomechatronics*, pp. 1101–1106. IEEE, Piscataway (2006)
13. Reilly, E.L.: EEG recording and operation of apparatus. In: *Electroencephalography: Basic Principles, Clinical Applications and Related Fields*, pp. 122–142. Lippincott Williams & Wilkins, Baltimore (1999)
14. Ruen Shan, L., Ibrahim, F., Moghavvemi, M.: Assessment of steady-state visual evoked potential for brain computer communication. In: *3rd Kuala Lumpur International Conference on Biomedical Engineering*, pp. 352–354. Springer, Heidelberg (2007)
15. Volosyak, I., Cecotti, H., Gräser, A.: Impact of Frequency Selection on LCD Screens for SSVEP Based Brain-Computer Interfaces. In: Cabestany, J., Sandoval, F., Prieto, A., Corchado, J.M. (eds.) *IWANN 2009, Part I. LNCS*, vol. 5517, pp. 706–713. Springer, Heidelberg (2009)

LUCIA: An Open Source 3D Expressive Avatar for Multimodal h.m.i.

G. Riccardo Leone, Giulio Paci, and Piero Cosi

Institute of Cognitive Sciences and Technologies – National Research Council
Via Martiri della libertà 2, 35137 Padova, Italy
{piero.cosi,riccardo.leone}@pd.istc.cnr.it

Abstract. LUCIA is an MPEG-4 facial animation system developed at ISTC-CNR¹. It works on standard Facial Animation Parameters and speaks with the Italian version of FESTIVAL TTS. To achieve an emotive/expressive talking head LUCIA was built from real human data physically extracted by ELITE optic-tracking movement analyzer. LUCIA can copy a real human being by reproducing the movements of passive markers positioned on his face and recorded by the ELITE device or can be driven by an emotional XML tagged input text, thus realizing true audio/visual emotive/expressive synthesis. Synchronization between visual and audio data is very important in order to create the correct WAV and FAP files needed for the animation. LUCIA's voice is based on the ISTC Italian version of FESTIVAL-MBROLA packages, modified by means of an appropriate APML/VSML tagged language. LUCIA is available in two different versions: an open source framework and the "work in progress" WebGL.

Keywords: talking head, TTS, facial animation, mpeg4, 3D avatar, virtual agent, affective computing, LUCIA, FESTIVAL.

1 Introduction

There are many ways to control a synthetic talking face. Among them, geometric parameterization [1, 2], morphing between target speech shapes [3], muscle and pseudo-muscle models [4, 5], appear the most attractive.

Growing interest have encountered text to audiovisual systems [6, 7], in which acoustical signal is generated by a Text to Speech engine and the phoneme information extracted from input text is used to define the articulatory movements.

To generate realistic facial animation is necessary to reproduce the contextual variability due to the reciprocal influence of articulatory movements for the production of following phonemes. This phenomenon, defined co-articulation [8], is extremely complex and difficult to model. A variety of co-articulation strategies are possible and even different strategies may be needed for different languages [9].

¹ With the collaboration of many students and researchers working at ISTC during these last years, among them: G.Tisato, F.Tesser, C.Drioli, V.Ferrari, G.Perin, A.Fusaro, D.Griioletto, M.Nicolao, G.Sommavilla, E.Marchetto.

A modified version of the Cohen-Massaro co-articulation model [10] has been adopted for LUCIA [11] and a semi-automatic minimization technique, working on real cinematic data acquired by the ELITE optic-electronic system [12], was used for training the dynamic characteristics of the model, in order to be more accurate in reproducing the true human lip movements.

Moreover, emotions are quite important in human interpersonal relations and individual development. Linguistic, paralinguistic and emotional transmission are inherently multimodal, and different types of information in the acoustic channel integrate with information from various other channels facilitating the communicative processes. The transmission of emotions in speech communication is a topic that has recently received considerable attention, and automatic speech recognition (ASR) and multimodal or audio-visual (AV) speech synthesis are examples of fields, in which the processing of emotions can have a great impact and can improve the effectiveness of human-machine interaction.

Viewing the face improves significantly the intelligibility of both natural and synthetic speech, especially under degraded acoustic conditions. Facial expressions signal emotions, add emphasis to the speech and facilitate the interaction in a dialogue situation. From these considerations, it is evident that, in order to create more natural talking heads, it is essential that their capability comprises the emotional behavior.

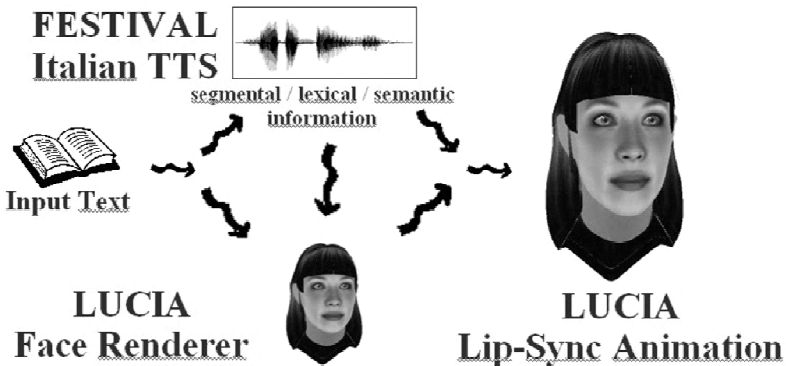


Fig. 1. LUCIA's functional block diagram

In our TTS (text-to-speech) framework, AV speech synthesis, that is the automatic generation of voice and facial animation from arbitrary text, is based on parametric descriptions of both the acoustic and visual speech modalities. The visual speech synthesis uses 3D polygon models, that are parametrically articulated and deformed, while the acoustic speech synthesis uses an Italian version of the FESTIVAL diphone TTS synthesizer [13] now modified with emotive/expressive capabilities. The block diagram of our framework is depicted in Fig. 1.

Various applications can be conceived by the use of animated characters, spanning from research on human communication and perception, via tools for the hearing

impaired, to spoken and multimodal agent-based user interfaces. The recent introduction of WebGL [14], which is 3D graphics in web browsers, opens the possibility to bring all these applications via internet. This software version of LUCIA is currently in the early stage of development.

2 Data Acquisition Environment

LUCIA is totally based on true real human data collected during the last decade by the use of ELITE [15, 16, 17], a fully automatic movement analyzer for 3D cinematics data acquisition [12], which provides 3D coordinate reconstruction, starting from 2D perspective projections, by means of a stereo-photogrammetric procedure which allows a free positioning of the TV cameras. The 3D data dynamic coordinates of passive markers (see Fig. 2) are then used to create our lips articulatory model and to drive directly our talking face, copying human facial movements.

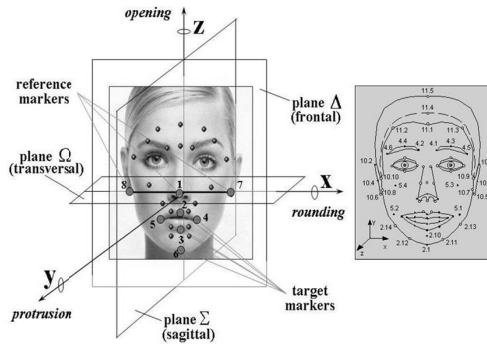


Fig. 2. Position of reflecting markers and reference planes for the articulatory movement data collection (*on the left*), and the MPEG-4 standard facial reference points (*on the right*)

Two different configurations have been adopted for articulatory data collection: the first one, specifically designed for the analysis of labial movements, considers a simple scheme with only 8 reflecting markers (bigger markers in Fig.2) while the second, adapted to the analysis of expressive and emotive speech, utilizes the full and complete set of 28 markers. All the movements of the 8 or 28 markers, depending on the adopted acquisition pattern, are recorded and collected, together with their velocity and acceleration, simultaneously with the co-produced speech which is usually segmented and analyzed by means of PRAAT [18], that computes also intensity, du-ration, spectrograms, formants, pitch synchronous F0, and various voice quality parameters in the case of emotive and expressive speech [19, 20].

In order to simplify and automates many of the operation needed for building-up the 3D avatar from the motion-captured data we developed INTERFACE [21], an integrated software designed and implemented in Matlab©.

3 Architecture and Implementations

LUCIA is a MPEG-4 standard facial animation engine implementing a decoder compatible with the "Predictable Facial Animation Object Profile" [22]. LUCIA speaks with the Italian version of FESTIVAL TTS [13]; we have already seen the overall system illustrated in Fig. 1. The homepage of the project is [23].

MPEG4 specifies a set of Face Animation Parameters (FAPs), each corresponding to a particular facial action deforming a face model in its neutral state. A particular facial action sequence is generated by deforming the face model, according to the specified FAP values, indicating the magnitude of the corresponding action, for the corresponding time instant. Then the model is rendered onto the screen.

LUCIA is able to generate a 3D mesh polygonal model by directly importing its structure from a VRML file [24] and to build its animation in real time.

At the current stage of development, LUCIA is a textured young female 3D face model built with 25423 polygons: 14116 belong to the skin, 4616 to the hair, 2688x2 to the eyes, 236 to the tongue and 1029 to the teeth respectively.

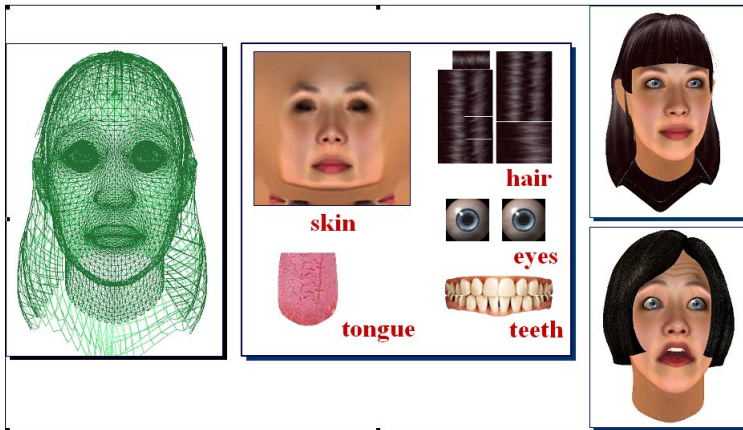


Fig. 3. Lucia's wireframe, textures and renderings

Currently the model is divided in two sub sets of fundamental polygons: the skin on one hand and the inner articulators, such as the tongue and the teeth, or the facial elements such as the eyes and the hair, on the other. This subdivision is quite useful when animation is running, because only the reticule of polygons corresponding to the skin is directly driven by the pseudo-muscles and constitutes a continuous and unitary element, while the other anatomical components move themselves independently, following translations and rotations (for example the eyes rotate around their center). According to this strategy the polygons are distributed in such a way that the resulting visual effect is quite smooth with no rigid "jumps" over all the 3D model.

LUCIA emulates the functionalities of the mimic muscles, by the use of specific "displacement functions" and of their following action on the skin of the face. The activation of such functions is determined by specific parameters that encode small

muscular actions acting on the face; these actions can be modified in time in order to generate the wished animation. Such parameters, in MPEG-4, take the name of Facial Animation Parameters and their role is fundamental for achieving a natural movement. The muscular action is made explicit by means of the deformation of a polygonal reticule built around some particular key points called Facial Definition Parameters (FDPs) that correspond to the junction on the skin of the mimic muscles.

Moving only the FDPs is not sufficient to smoothly move the whole 3D model, thus, each "feature point" is related to a particular "influence zone" constituted by an ellipses that represents a zone of the reticule where the movement of the vertexes is strictly connected. Finally, after having established the relationship for the whole set of FDPs and the whole set of vertexes, all the points of the 3D model can be simultaneously moved with a graded strength following a raised-cosine function rule associated to each FDP.

There are two current versions of LUCIA: an open source 3D facial animation framework written in C programming language [25] and a new WebGL implementation [26]. The C framework allows efficient rendering of a 3D face model in OpenGL-enabled systems (it has been tested on Windows and Linux using several architectures). It has a modular design: each module provides one of the several common facilities needed to create a real-time Facial Animation application. It includes an interface to play audio files, a robust and extendable FAP parser, sample-based audio/video synchronization and an object oriented interface to the components of a face model. Finally, the framework includes a ready to use female face model and some sample applications to play simple animation and to test movements' behavior. The framework's core is built on top of OpenGL and does not rely on any specific context provider (it has been tested using GLUT[27], FreeGLUT[28] and GtGLExt[29]). In order to grant portability, the modules' interfaces are designed so that their implementation details are hidden to the application and it is possible to provide multiple implementation of the same model (e.g., three implementations of the audio module are included using respectively OpenAL[30], Gstreamer[31] and Microsoft MCI[32]). The very recent introduction of 3D graphics in the web browsers (which is known as WebGL [14]) opens new possibilities for our 3D avatar. The powerful of this new technology is that you don't need to download any additional software or driver to access the content of the 3D world you are interacting with. We are currently developing this new software version in order to easily integrate LUCIA in a website; there are many promising functionality for web applications: a virtual guide (which we are exploiting in the Wikimemo.it project - The portal of Italian Language and Culture); a storyteller for e-book reading; a digital tutor for hearing impaired; a personal assistant for smart-phone and mobile devices. The early results can be observed in [26].

4 Emotional Synthesis

Audio Visual emotional rendering was developed working on true real emotional audio and visual databases whose content was used to automatically train emotion specific intonation and voice quality models to be included in FESTIVAL, our Italian

Meaning Semantic	DTD tag names	Abstraction level	Examples	APML
Emotions Expressions	affective	3	<fear>	
Voice Quality	voqual	2	<breathy> ... <tremulous>	VSML
Acoustic Controls	signalctrl	1	<asp_noise> ... <spectral_tilt>	

Fig. 4. APML/VSML mark-up language extensions for emotive audio/visual synthesis

TTS system [33, 34, 35, 36] and also to define specific emotional visual rendering to be implemented in LUCIA [37, 38, 39].

An emotion specific XML editor explicitly designed for emotional tagged texts was developed. The APML mark up language [40] for behavior specification permits to specify how to markup the verbal part of a dialog move so as to add to it the "meanings" that the graphical and the speech generation components of an animated agent need to produce the required expressions (Fig. 4). So far, the language defines the components that may be useful to drive a face animation through the facial description language (FAP) and facial display functions. The extension of such language is intended to support voice specific controls. An extended version of the APML language has been included in the FESTIVAL speech synthesis environment, allowing the automatic generation of the extended phonation file from an APML tagged text with emotive tags. This module implements a three-level hierarchy in which the affective high level attributes (e.g. <anger>, <joy>, <fear>) are described in terms of medium-level voice quality attributes defining the phonation type (e.g., <modal>, <soft>, <pressed>, <breathy>, <whispery>, <creaky>). These medium-level attributes are in turn described by a set of low-level acoustic attributes defining the perceptual correlates of the sound (e.g. <spectral tilt>, <shimmer>, <jitter>). The low-level acoustic attributes correspond to the acoustic controls that the extended MBROLA synthesizer can render through the sound processing procedure described above. This descriptive scheme has been implemented within FESTIVAL as a set of mappings between high-level and low-level descriptors. The implementation includes the use of envelope generators to produce time curves of each parameter.

In order to check and evaluate, by direct low-level manual/graphic instructions, various multi level emotional facial configurations we developed "EmotionPlayer", which was strongly inspired by the EmotionDisc of Zsofia Ruttkay [41]. It is designed for a useful immediate feedback, as exemplified in Fig. 5.

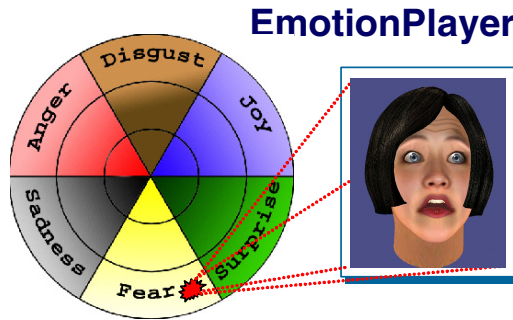


Fig. 5. Emotion Player: clicking on three-level intensity (low, mid, high) emotional disc, an emotional configuration (i.e. high -fear) is activated

5 Conclusions

LUCIA is an MPEG-4 standard FAPs driven OpenGL framework which provides several common facilities needed to create a real-time Facial Animation application. It has high quality 3D model and a fine co-articulatory model, which is automatically trained by real data, used to animate the face.

The modified co-articulatory model is able to reproduce quite precisely the true cinematic movements of the articulatory parameters. The mean error between real and simulated trajectories for the whole set of parameters is, in fact, lower than 0.3 mm.

Labial movements implemented with the new modified model are quite natural and convincing especially in the production of bilabials and labiodentals and remain coherent and robust to speech rate variations.

The overall quality and user acceptability of LUCIA talking head has to be perceptually evaluated [42, 43] by a complete set of test experiments, and the new model has to be trained and validated in asymmetric contexts (VICV2) too. Moreover, emotions and the behavior of other articulators, such as tongue for example, have to be analyzed and modeled for a better realistic implementation.

A new WebGL implementation of the avatar is currently in progress to exploit new possibilities that arise from the integration of LUCIA in the internet websites.

Acknowledgments. Part of this work has been sponsored by PF-STAR (Preparing Future multiSensorial inTerAction Research, European Project IST- 2001-37599, <http://pfstar.itc.it>), TICCA (Tecnologie cognitive per l'Interazione e la Cooperazione Con Agenti artificiali, joint "CNR - Provincia Autonoma Trentina" Project) and WIKIMEMO.IT (The Portal of Italian Language and Culture, FIRB Project, RBNE078K93, Italian Ministry of University and Scientific Research).

References

1. Massaro, D.W., Cohen, M.M., Beskow, J., Cole, R.A.: Developing and Evaluating Conversational Agents. In: Cassell, J., Sullivan, J., Prevost, S., Churchill, E. (eds.) *Embodied Conversational Agents*, pp. 287–318. MIT Press, Cambridge (2000)
2. Le Goff, B.: *Synthèse à partir du texte de visages 3D parlant français*. PhD thesis, Grenoble, France (1997)
3. Bregler, C., Covell, M., Slaney, M.: Video Rewrite: Driving Visual Speech with Audio. In: *SIGGRAPH 1997*, pp. 353–360 (1997)
4. Lee, Y., Terzopoulos, D., Waters, K.: Realistic Face Modeling for Animation. In: *SIGGRAPH 1995*, pp. 55–62 (1995)
5. Vatikiotis-Bateson, E., Munhall, K.G., Hirayama, M., Kasahara, Y., Yehia, H.: Physiology-Based Synthesis of Audiovisual Speech. In: *4th Speech Production Seminar: Models and Data*, pp. 241–244 (1996)
6. Beskow, J.: Rule-Based Visual Speech Synthesis. In: *Eurospeech 1995*, Madrid, Spain, pp.299–302 (1995)
7. LeGoff, B., Benoit, C.: A text-to-audiovisualspeech synthesizer for French. In: *ICSLP 1996*, Philadelphia, U.S.A., pp. 2163–2166 (1996)
8. Farnetani, E., Recasens, D.: Coarticulation Models in Recent Speech Production Theories. In: *Hardcastle, W.J. (ed.) Coarticulation in Speech Production*. Cambridge University Press, Cambridge (1999)
9. Bladon, R.A., Al-Bamerni, A.: Coarticulation resistance in English. *J. Phonetics*, 4, 135–150 (1976)
10. Cosi, P., Perin, G.: Labial Coarticulation Modeling for Realistic Facial Animation. In: *ICMI 2002*, Pittsburgh, U.S.A., pp. 505–510 (2002)
11. Cosi, P., Fusaro, A., Tisato, G.: LUCIA a New Italian Talking-Head Based on a Modified Cohen-Massaro's Labial Coarticulation Model. In: *Eurospeech 2003*, Geneva, Switzerland, vol. III, pp. 2269–2272 (2003)
12. Ferrigno, G., Pedotti, A.: ELITE: A Digital Dedicated Hardware System for Movement Analysis via Real-Time TV Signal Processing. *IEEE Transactions on Biomedical Engineering*, BME-32, 943–950 (1985)
13. Cosi, P., Tesser, F., Gretter, R., Avesani, C.: Festival Speaks Italian! In: *Eurospeech 2001*, Aalborg, Denmark, pp. 509–512 (2001)
14. WebGL- OpenGL for the web, <http://www.khronos.org/webgl/>
15. Cosi, P., Magno Caldognetto, E.: Lip and Jaw Movements for Vowels and Consonants: Spatio-Temporal Characteristics and Bimodal Recognition Applications. In: *Storke, D.G., Henneke, M.E. (eds.) Speechreading by Humans and Machine: Models, Systems and Applications*. NATO ASI Series, Series F: Computer and Systems Sciences, vol. 150, pp. 291–313. Springer, Heidelberg (1996)
16. Magno Caldognetto, E., Zmarich, C., Cosi, P., Ferrero, F.: Italian Consonantal Visemes: Relationships Between Spatial/temporal Articulatory Characteristics and Coproduced Acoustic Signal. In: *AVSP 1997, Tutorial & Research Workshop on Audio-Visual Speech Processing: Computational & Cognitive Science Approaches*, Rhodes, Greece, pp. 5–8 (1997)
17. Magno Caldognetto, E., Zmarich, C., Cosi, P.: Statistical Definition of Visual Information for Italian Vowels and Consonants. In: *Burnham, D., Robert-Ribes, J., Vatikiotis-Bateson, E. (eds.) Proceedings of AVSP 1998*, Terrigal, Austria, pp. 135–140 (1998)

18. Boersma, P.: PRAAT, a system for doing phonetics by computer. *Glott International* 5(9/10), 341–345 (1996)
19. Magno Caldognetto, E., Cosi, P., Drioli, C., Tisato, G., Cavicchio, F.: Coproduction of Speech and Emotions: Visual and Acoustic Modifications of Some Phonetic Labial Targets. In: AVSP 2003, ISCA Workshop, St Jorioz, France, pp. 209–214 (2003)
20. Drioli, C., Tisato, G., Cosi, P., Tesser, F.: Emotions and Voice Quality: Experiments with Sinusoidal Modeling. In: Proceedings of Voqual 2003, Voice Quality: Functions, Analysis and Synthesis, ISCA Workshop, Geneva, Switzerland, pp. 127–132 (2003)
21. Tisato, G., Cosi, P., Drioli, C., Tesser, F.: INTERFACE: a New Tool for Building Emotive/Expressive Talking Heads. In: INTERSPEECH 2005, Lisbon, Portugal, pp. 781–784 (2005)
22. MPEG-4 standard,
<http://mpeg.chiariglione.org/standards/mpeg-4/mpeg-4.htm>
23. LUCIA - homepage, <http://www2.pd.istc.cnr.it/LUCIA/>
24. Hartman, J., Wernecke, J.: *The VRML Handbook*. Addison Wesley (1996)
25. LUCIA Open source project, <http://sourceforge.net/projects/lucia/>
26. LUCIA WebGL version, <http://www2.pd.istc.cnr.it/LUCIA/webgl/>
27. The OpenGL Utility Toolkit,
<http://www.opengl.org/resources/libraries/glut/>
28. The Free OpenGL Utility Toolkit, <http://freeglut.sourceforge.net>
29. GTK+ OpenGL Extension, <http://projects.gnome.org/gtkglext/>
30. OpenAL: a cross platform 3D audio API,
<http://connect.creativelabs.com/openal/>
31. Gstreamer: Open source multimedia framework,
<http://gstreamer.freedesktop.org/>
32. Windows MCI,
http://en.wikipedia.org/wiki/Media_Control_Interface
33. Tesser, F., Cosi, P., Drioli, C., Tisato, G.: Prosodic Data-Driven Modelling of Narrative Style in FESTIVAL TTS. In: 5th ISCA Speech Synthesis Workshop, Pittsburgh, U.S.A (2004)
34. Tesser, F., Cosi, P., Drioli, C., Tisato, G.: Emotional Festival-Mbrola TTS Synthesis. In: INTERSPEECH 2005, Lisbon, Portugal, pp. 505–508 (2005)
35. Drioli, C., Tesser, F., Tisato, G., Cosi, P.: Control of Voice Quality for Emotional Speech Synthesis. In: 1st Conference of Associazione Italiana di Scienze della Voce, AISV 2004, EDK Editore s.r.l., Padova, Italy, pp. 789–798 (2005)
36. Nicolao, M., Drioli, C., Cosi, P.: GMM modelling of voice quality for FESTIVAL-MBROLA emotive TTS synthesis. In: INTERSPEECH 2006, Pittsburgh, U.S.A, pp. 1794–1797 (2006)
37. Cosi, P., Fusaro, A., Grigoletto, D., Tisato, G.: Data-Driven Tools for Designing Talking Heads Exploiting Emotional Attitudes. In: André, E., Dybkjær, L., Minker, W., Heisterkamp, P. (eds.) ADS 2004. LNCS (LNAI), vol. 3068, pp. 101–112. Springer, Heidelberg (2004)
38. Magno Caldognetto, E., Cosi, P., Drioli, C., Tisato, G., Cavicchio, F.: Visual and acoustic modifications of phonetic labial targets in emotive speech: Effects of the co-production of speech and emotions. *J. Speech Communication* 44, 173–185 (2004)
39. Magno Caldognetto, E., Cosi, P., Cavicchio, F.: Modifications of Speech Articulatory Characteristics in the Emotive Speech. In: André, E., Dybkjær, L., Minker, W., Heisterkamp, P. (eds.) ADS 2004. LNCS (LNAI), vol. 3068, pp. 233–239. Springer, Heidelberg (2004)

40. De Carolis, B., Pelachaud, C., Poggi, I., Steedman, M.: APML, a Mark-up Language for Believable Behavior Generation. In: Prendinger, H., Ishizuka, M. (eds.) *Life-Like Characters*, pp. 65–85. Springer, Heidelberg (2004)
41. Ruttkay, Z., Noot, H., Hagen, P.: Emotion Disc and Emotion Squares: tools to explore the facial expression space. *Computer Graphics Forum* 22(1), 49–53 (2003)
42. Massaro, D.W.: *Perceiving Talking Faces: from Speech Perception to a Behavioral Principle*. MIT Press, Cambridge (1997)
43. Costantini, E., Pianesi, F., Cosi, P.: Evaluation of Synthetic Faces: Human Recognition of Emotional Facial Displays. In: André, E., Dybkjær, L., Minker, W., Heisterkamp, P. (eds.) *ADS 2004. LNCS (LNAI)*, vol. 3068, pp. 276–287. Springer, Heidelberg (2004)

The AnimaTricks System: Animating Intelligent Agents from High-Level Goal Declarations

Vincenzo Lombardo, Fabrizio Nunnari, and Rossana Damiano

Dipartimento di Informatica
and CIRMA, Università di Torino,
Virtual Reality and Multimedia Park, Torino

Abstract. This paper presents AnimaTricks, a system for the generation of the behavior of an animated agent from a high level description of its goals. The deliberation component generates a sequence of actions given a set of goals. The animation component, then, translates it into an animation language, leaving to the animation engine the task of generating the actual animation.

The purpose of the system is two-fold. First, we test how deliberation can be effectively tied to the animated counterpart. Second, by generating complex animations from high-level goals, AnimaTricks supports the work of directors and animators in a pre-visualization and re-use perspective.

Keywords: animation, virtual characters, multimedia production.

1 Introduction

The AnimaTricks system addresses the task of realizing the animated behavior of the agents in multimedia production, with the purpose of allowing an author to specify the high-level goals of the agent, leaving the system the task of generating the actual animations that fit a specific situation.

Decision processes and actions are tightly coupled in layered architectures [9,5,7], a paradigm where the declarative description of the agent's behavior drives the generation of animation, in order to craft complex and non-deterministic character behavior. Following this principle, in AnimaTricks, the deliberative component is *integrated* in the system: the agent's deliberative component uses an AI planner to generate the sequence of actions that constitute the agent's behavior. The animation system, then, translates the actions into an animation language. The animation engine, based on the open source OGRE 3D rendering engine, interprets the animation language expressions to generate the visible behavior of the agent.

The structure of the paper is the following. First, we briefly survey the state of the art of declarative languages for behavior description (Sect. 2), then we describe the system architecture (Sect. 3) and the animation system (Sect. 4). Conclusions and future work end the paper.

2 Declarative Languages for Animated Agents

The need to translate the agents' actions into animations, coping at the same time with real time constraints, has led scholars to design ad hoc animation languages that bridge the gap between behavior description, issued by some deliberative component, and the animation layer. The pioneering PAR language [2] introduced the use of templates to represent actions, establishing the design of markup languages for character animation (starting from [1]). One of the most documented and solidly implemented is the Behavior Markup Language (BML), geared to describe the multimodal communicative behavior of an agent [6]. BML acknowledges a set of communicative channels that contribute to the performance of multimodal communicative acts, such as gestures, gaze or posture. It offers tools to describe into detail the behavior of the agent along each channel (for example, the gaze is described in reference to its target, duration, etc.) and to synchronize the channels. BML descriptions can be fed to a realizer that transforms them into the corresponding animations [4]. Finally, EMBR is an animation scripting language underlying the BML realizer [4].

The family of languages mentioned above are intended to model interactive and social behavior, such as gesture, posture and so on. In Animatricks, similarly to [5], we take a neutral stance with reference to the displayed behavior, and provide an animation language geared to low-level authoring of meaningful behaviors. Apart from the differences in design goals, we identify two main differences with respect to BML/EMBR. First, it requires the specification of many temporal constraints. This can be acceptable for its use within the BML architecture, where such constraints are meant to be generated by an automatic solver, but it can be a hard job for a human editor. Differently, in Animatricks we specify animation "speeds", for which it is easier to identify default values (e.g., walk speed), and which are at run-time converted in durations according to actual path lengths. Second, EMBR lacks the syntax to expose parameters for newly defined animations. In Animatricks, the animation language supports the definition of parametrized actions which can adapt to different situations and can be stored for reuse.

3 The AnimaTricks System

The architecture of the AnimaTricks system includes three main components (Figure 1): the Planner (the mind of the "character"), the Executor (the "actor" who plays the character), and the Animation engine (the "body" of the actor).

In the current implementation of the AnimaTricks system, the Planner is given by the JSHOP2 HTN planning system [8]. According to the HTN paradigm, actions can be primitive actions, i.e., directly executable ones ("operators"), or complex actions ("methods"), if they encompass a sequence of simpler actions or abstract over alternatives. When the Planner is invoked, it matches the task to be achieved onto a high-level method and starts refining it into simpler tasks, discarding alternatives that do not fit the given world state. The refinement

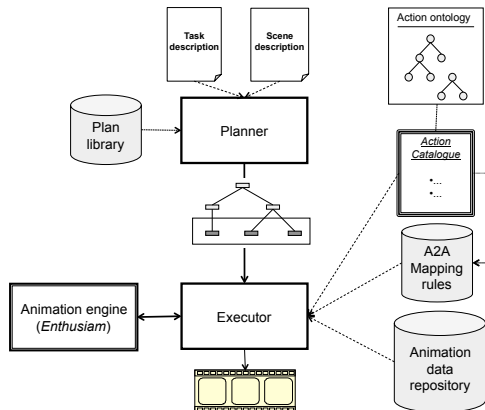


Fig. 1. The architecture of the AnimaTricks system

ends when all high-level tasks have been expanded into one or more sequences of primitive tasks.

The knowledge about plans (i.e., the plan library) is kept separated from the definition of the actions (action catalogue). By doing so, actions can be reused across different projects, independently of the plans they appear into. The actions in the catalog are annotated with reference to an ontology to guarantee that actions are externally defined in a shared, machine-readable form.

The Executor consults the action catalogue to replace each primitive task in the generated plan with the corresponding action. Then, for each action, it applies the matching A2A mapping rule to obtain the animation language expression that describes the animation.

The animation language expressions are then fed to the Animation engine, which generates the animation (possibly using pre-stored animation data from the Animation data repository). The Animation engine, implemented as part of the Enthusiasm project (<http://enthusiasm.sourceforge.net>), relies on the Ogre3D (www.ogre3d.org) rendering engine.

4 Creating Animations: Language and Pipeline

The animation system interprets a list of primitive commands that generate animations edited following the techniques surveyed in [10]. The underlying primitives range from the (parametrized) playback of animation clips, to the generation of movements via IK and the blending of clips. Control structures support the parallel and sequential use of the animation primitives, and the combination of both.

Basic animations can be obtained through the *retrieval* of a clip from a repository of animation clips or through pure procedural animation, e.g., as a bone performing an arc in the space (a *spline*). An object moves through the space

following a *path*, at a certain speed, and rotates on a specified axis, at some turn speed (in degrees per second).

Finally, several animations can be blended sequentially, or, considering that each animation might operate only on a segment of the virtual agent, they can be performed in parallel

For example, the following A2A (action to animation) rule (see below) says that the animation of the action of walking is generated by repeating the walking loop animation (`repeat(clipAnimation "walk_cycle")`, line 3), stored in the animation data repository, while (`par`, line 2) following the path the stretches from the initial to the final location `follow_path(from_location, to_location`, line 4). If not differently specified, the walking speed is the standard one (line 5).

```

1 walk(from_location:String, to_location:String) {
2   par(
3     {repeat(clipAnimation("walk_cycle"))
4       follow_path({from_location, to_location},
5         DEFAULT_WALK_SPEED)
6     },
7     "first")
8 }
```

The pipeline for creating the contents for the AnimaTricks system includes three main phases: the Behavior definition, where the AI expert encodes the behavior of the agent into the format required by the planner and stores it in the Plan library; the Semantic Tagging, where the basic actions are tagged with semantic labels and stored in the Action catalogue; the Action-to-Animation Mapping, where the A2A rules for actions are defined and the animation data are created.

1. **Behavior definition.** The author describes the desired behavior for a certain character and the settings in which the character may be situated. The planning expert designs and implements a plan library that encodes this behavior and tests it.
2. **Semantic tagging.** Given the rigged 3D model of the character, the animator cooperates with the knowledge engineer and the 3D programmer to describe the primitive tasks contained in the plan library, mapping them to existing actions when possible. Currently, the descriptive labels in the AnimaTricks system rely on the IEEE Standard Upper Merged Ontology (SUMO) ontology and on the WordNet lexicon [3].
3. **Action-to-animation mapping.** If an action must be produced from scratch, the animator and the 3D programmer translate it into a A2A mapping rule. First of all, the structure of the action is broken down into its components and the 3D programmer evaluates if the action can be procedurally generated. Finally, the animation data are produced and stored in the repository.

In December 2009, the AnimaTricks system was employed to conduct an experiment on the creation of extra characters in serial productions (TV series, video games, etc.). We asked an author to write down a script by inserting actions that typically recur in TV series (for example, an interior location, some pieces of furnitures, actions such as answering the phone, walking, etc). Then, the animator animated the story following the traditional working methodology.

The same animation was produced in real time by using the procedural animation functionalities supported by Animatricks for validation purposes. The plan library included actions such as entering, sitting at the desk to accomplish several tasks, like doing or receiving phone calls, hand-writing letters and notes, getting up to take objects (pen, sheets, etc.) when necessary (Fig. 2). The plan library contained 17 complex actions (methods) and 21 primitive actions (operators). The planner was tested on 20 different scenarios and produced as many different plans, that contained from 16 to 32 actions; 5 scenarios were selected to run the evaluation.

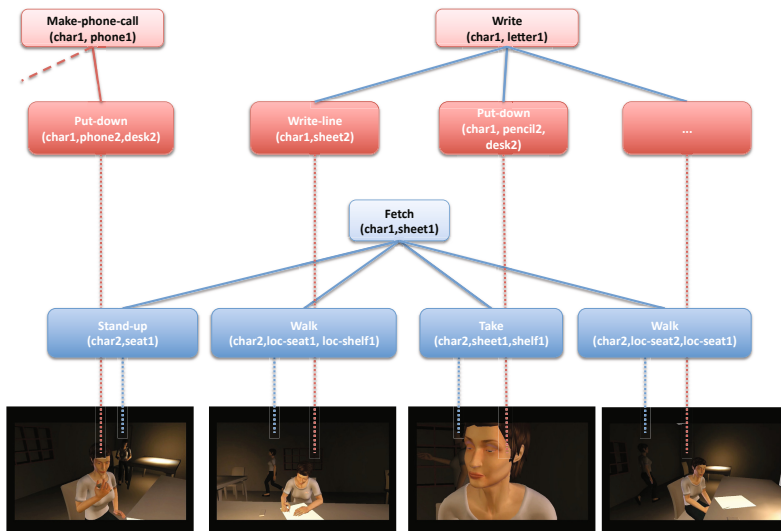


Fig. 2. Snapshots of the video generated by the AnimaTricks system. The upper part of the figure shows the plan generated by the planner (dark boxes are primitive tasks); in each shot, the characters are connected by the dashed lines to the plan actions they are executing.

The resulting video, together with the production pipeline of the AnimaTricks system, was presented to a focus group of animation producers, researchers, and trainers. The experts gave a positive evaluation to the expressiveness of the animation language and to its potential for use in the animation industry.

5 Conclusions and Future Work

In this paper we described the AnimaTricks system for animating artificial agents from a high-level specification of their behavior. The core of the system is a set of rules that map the agent's actions, generated by the AI component, to procedurally generated animations, thus tying deliberation to physical behavior in a virtual environment. With AnimaTricks, each production phase is supported

by declarative languages, in order to make the behavior design more explicit and to promote the reuse of animations.

The current system does not support interactivity. Since the paradigm of HTN planning can be straightforwardly adapted to replanning, we are planning to expand the system to interactive animations.

References

1. Arafa, Y., Mamdani, A.: Scripting embodied agents behaviour with CML: character markup language. In: Proceedings of the 8th International Conference on Intelligent User Interfaces, p. 316. ACM (2003)
2. Badler, N.I., Bindiganavale, R., Allbeck, J., Schuler, W., Zhao, L., Palmer, M.: Parametrized action representation for virtual human agents. In: Cassell, J., Sullivan, J., Prevost, S., Churchill, E. (eds.) *Embodied Conversational Agents*, pp. 256–284. The MIT Press, Cambridge (2000)
3. Fellbaum, C., et al.: *WordNet: An electronic lexical database*. MIT Press, Cambridge (1998)
4. Heloir, A., Kipp, M.: Real-Time Animation of Interactive Agents: Specification and Realization. *Applied Artificial Intelligence* 24(6), 510–529 (2010)
5. Isla, D., Burke, R., Downie, M., Blumberg, B.: A layered brain architecture for synthetic creatures. In: *International Joint Conference on Artificial Intelligence*, vol. 17, pp. 1051–1058. Citeseer (2001)
6. Kopp, S., Krenn, B., Marsella, S.C., Marshall, A.N., Pelachaud, C., Pirker, H., Thórisson, K.R., Vilhjálmsson, H.H.: Towards a Common Framework for Multimodal Generation: The Behavior Markup Language. In: Gratch, J., Young, M., Aylett, R.S., Ballin, D., Olivier, P. (eds.) *IVA 2006. LNCS (LNAI)*, vol. 4133, pp. 205–217. Springer, Heidelberg (2006)
7. Loyall, A.B., Reilly, W., Bates, J., Weyhrauch, P.: System for authoring highly interactive, personality-rich interactive characters. In: *Proc. of the 2004 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pp. 59–68 (2004)
8. Nau, D., Au, T.-C., Ilghami, O., Kuter, U., Munoz-Avila, H., William Murdock, J., Wu, D., Yaman, F.: Applications of shop and shop2. *IEEE Intelligent Systems* 20(2), 34–41 (2005)
9. Perlin, K., Goldberg, A.: Improv: a system for scripting interactive actors in virtual worlds. In: *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 1996*, pp. 205–216. ACM press, New York (1996)
10. Van Welbergen, H., Van Basten, B.J.H., Egges, A., Ruttkay, Z.M., Overmars, M.H.: Real Time Animation of Virtual Humans: A Trade-off Between Naturalness and Control. In: *Computer Graphics Forum*. Wiley Online Library

A Framework for Designing 3D Virtual Environments

Salvatore Catanese¹, Emilio Ferrara²,
Giacomo Fiumara¹, and Francesco Pagano³

¹ Dept. of Physics, Informatics Section, University of Messina, Italy

² Dept. of Mathematics, University of Messina, Italy

³ Dept. of Information Technology, University of Milan, Italy

Abstract. The process of design and development of virtual environments can be supported by tools and frameworks, to save time in technical aspects and focusing on the content. In this paper we present an academic framework which provides several levels of abstraction to ease this work. It includes state-of-the-art components we devised or integrated adopting open-source solutions in order to face specific problems. Its architecture is modular and customizable, the code is open-source.

Keywords: Virtual Environments, Games.

1 Introduction

Commercial games reach production costs up to millions dollars. During the process of development, teams spend a lot of time in prototyping of particular features, and, commonly, building technical frameworks to design the game on top of them. Designing a game from scratch requires a lot of time, fund investments, skills and resources. This process could be shortened by adopting existing platforms. In this paper we describe a fully developed framework, thought to ease the development of 3D virtual worlds, supporting state-of-the-art techniques. It could be adopted if programmers and artists would prefer to focus on aspects of the design process (e.g., gameplay, game mechanics, etc.), instead of technical aspects. In academia, exploiting an existing solution could ease researchers to carry out specific experimentations ignoring other aspects of the development of a platform. In fact, the source code of our framework has been released open-source [1]. This represents a great advantage w.r.t. using different solutions, also freely available on the Web, whose code is closed, architecture is fixed and components are not replaceable, differently from our framework. Moreover, its functionality and potential have already been exploited: demonstrative virtual environments, designed using our platform, have already been presented to the scientific community [2].

¹ <http://informatica.unime.it/velab>

2 Related Work

Literature and academic interest in supporting videogames development is growing. There are some works sharing similarities with ours, e.g., Liu et al. [6] developed a Virtual Reality platform, designing server and client applications, including a rendering engine based on *OGRE*, combined with *RakNet* for networking features, and finally they developed an example application. Also Braun et al. [1] recently presented a platform, for simulating virtual environments where users communicate and interact each other, using avatars with facial expressions and body motions. Some interesting techniques of computer vision and physics effects have been implemented to increase the realism of the simulation. In a broader panorama, Hu-xiong et al. [5] discussed the design and implementation of a game platform capable of running online games, both from client and server perspective. Concluding, Graham and Roberts [4] analyzed the videogame development process from a qualitative perspective, trying to define attributes of 3D games and adopting interesting criteria to help achieving desired standards of quality during the design of academic and commercial products.

3 Architecture of the Framework

In this section we briefly illustrate the components usually integrated in those frameworks supporting the design of virtual environments. An important feature typically provided is the capability of rendering and managing 3D scenes, as 3D environments are usually enriched by a realistic behavior of elements populating them, by reproducing physics and detecting collisions. Input/output management includes, among others, reproducing music and sound, handling devices like keyboard and mouse, pads, etc. Characters populating 3D virtual worlds could be driven by artificial intelligence (another aspect optionally supported), by other humans (requiring networking features for multi-player aspects), or both. In this project we developed components, interfaces and plug-in for the integration of some existing open-source solutions. We devised a framework able of supporting teams in the process of design and development of virtual environments. Its modular nature ensures the possibility of integrating additional components to tackle specific requirements.

Rendering Engine. Our framework integrates an open-source rendering engine, namely *OGRE*. The *rendering* is the process of generation of an image from a model. A *model* is a description of a three-dimensional object defined by data structures containing geometry, textures, lighting, materials, etc. The rendering can be in real-time or not: the first approach is adopted for example in developing videogames, while the latter is typically used in those applications where the primary requirement is photorealism rather than performance (e.g., post-production effects in movies, medical image analysis, etc.). Rendering algorithms usually try to simulate optic phenomena to reproduce 3D environments. These techniques rely on rendering *primitives* (e.g., triangles and polygons for three-dimensional scenes), and not pixels. The process of transforming 3D scenes

into 2D images is implemented by rendering pipelines, supported by graphics hardware. A typical input for a graphics pipeline is a model of scene and its output represents a bi-dimensional raster. *OpenGL* and *Direct3D* are two examples of graphics pipeline implementations. The interaction between graphics pipeline and hardware is possible via direct access to resources or graphics libraries. Usually, rendering engines exploit existing graphics libraries to ease the process of development. Our platform supports both *Direct3D* and *OpenGL*.

Scene Manager. A *scene manager* is included in our framework to manage scenes representing 3D environments. This component organizes the objects on the scene and the relationships among them, in a hierarchical way. A scene manager could be designed in different ways and its implementation could affect the overall performance. Our scene manager adopts a “scene graph” whose nodes represent entities or objects on the scene and edges represent relations among them. Moreover, it manages *bounding volume hierarchies* (BVHs), trees adopted to represent bounding volumes (e.g., spheres, bounding boxes, etc.) containing objects. BVHs are also adopted for speeding up the collision detection among objects on the scene. In order to increase performance, minimizing the number of rendered elements, our framework supports techniques of spatial partitioning, such as the *Binary Space Partitioning* (BSP) [3] and *octrees* [7]. Some areas are not always visible in the frustum (i.e., the region of space visible from the character point of view), thus the scene manager, exploiting spatial relations among nodes, disregards their rendering.

Physics Engine. A valid framework should allow the simulation of the physics on the reproduced virtual environment. Several physics engines have been developed during last years to improve the degree of realism of videogames. *Havok* and *PhysX* are two examples. The latter is a robust solution; we integrated it inside our framework, developing an interface to introduce a good level of abstraction.

Collision Detection. The problem of detecting collisions among objects on the scene is complex and involves several aspects of the simulation of a virtual world. *PhysX* provides advanced directives for detecting collisions of characters with obstacles, to simulate the impact of objects with other elements on the scene and so on. *PhysX* adopts bounding volumes to surround objects inside shapes and checks for interactions among bounding volumes; it additionally exploits the ragdoll model for collision detection of characters. After detecting a collision, the physics engine simulates effects on involved objects.

Character Controller. One important aspect of the gameplay in a videogame is the strategy of control of the character. Players interact with the virtual world using their avatars. It is fundamental to reproduce a realistic behavior in order to avoid frustration during the game experience. Moreover, the character is central in the perception of the player, thus all the imperfections are extremely visible. An important aspect to be considered is the shape adopted for detecting the collisions. A simple choice is a “capsule”, embedding the character; this because, i) the shape is smooth, so the character could run on irregular terrains without

getting stuck; ii) its symmetry ensures the character could turn around itself without hitting obstacles; iii) no rough-edges could obstacle the character when going through narrow passages. Collateral effects are related to its simplicity. More complex models, such as the ragdoll [8], have been proposed. This technique produces procedural animations, instead of static ones. A ragdoll model is a collection of constrained-rigid-bodies (usually representing bones and joints of characters) adopted as a skeletal animation system. Animations are generated in real-time, without adopting predetermined ones, increasing the degree of realism. Moreover, the rigid-body system is subjected to rules of the physics engine; thus, the interaction between the character and the environment is more accurate. Our framework supports both these two control strategies.

4 Design and Characteristics of Our Platform

Our goal was to design a framework solution to support design and development of virtual environments, providing the following features: 3D rendering capabilities, physics simulation, collision management and character controller, I/O management, scene and camera management (i.e., loading and saving scenes representing virtual environments, with graphics and physics characteristics, designed with a specific application).

4.1 Adopted Tools

There are several open-source solutions which provide state-of-the-art features for many of the required tasks. Thus, the most of the time has been spent for the integration of these components together, in order to obtain a unique framework. All the components, interfaced each other, work as a substrate for developing an application on top of them. A short description of the chosen tools follows.

OGRE. It is a cross-platform 3D rendering engine. This engine is scene-oriented and allows the graphical representation of 3D virtual environments, providing state-of-the-art techniques for visual effects, texturing and lighting. A key aspect is its modular nature. It is possible to extend the provided features by the inclusion of new components. We exploited both these aspects to extend and improve the engine itself. Its architecture supports both the *Direct3D* and *OpenGL* graphics pipelines. *OGRE* is constituted by a collection of classes and libraries. Three main classes (*scene manager*, *resources manager* and *render*) do the most of the work. It is possible to inherit and extend the default scene manager to implement new or different features; for example the “BSPSceneManager” is optimized to represent in-door environments, while the “TerrainSceneManager” better works with out-door scenes; finally, the “OctreeSceneManager” is a good compromise for general purposes.

OGREOggSound. We integrated this component in the framework; it is an audio library which acts as a wrapper for the *OpenAL* API. Its adoption allows to automatize the inclusion of audio features, e.g., to support wave and “Ogg

Vorbis” audio sources, static and dynamic environmental sounds, 2D/3D audio, etc. Sound elements can be included inside the scenes as positional audio sources to reproduce more realistic environments.

PhysX and NxOGRE. *PhysX* is part of our framework. It is in charge of the simulation of physics laws to increase the degree of realism of the reproduced virtual environments. Some key aspects managed by the engine involve a rigid-body and soft-body system simulation, an advanced character controller supporting different shapes (e.g., capsules, boxes, triangular and convex meshes, etc.), simulation of fluid dynamics, field strength, collisions, etc. *PhysX* is integrated via the *NxOGRE* class wrappers, which introduces a useful level of abstraction to ease the access to functionalities provided by the library.

Blender. *Blender* is a cross-platform open-source 3D graphics and modeling application. We adopted *Blender* because of the possibility of customizing this tool to include new features. *Blender* can manage multiple scenes. Each scene is represented by its structure, objects, textures, materials, sound sources, etc. We extended it to support also physics; some physical properties can be defined for each object, in order to better reflect its behavior in the virtual world and the type of interaction between the character and objects.

4.2 The Framework

The basic idea is to use *Blender* to design the scene and then pass its output to the appropriate manager. *Blender* originally deals just with graphics, but we extended and improved its open and powerful architecture. Using our framework, it is possible to add to each component of the scene a set of user-defined attributes and their values. This way, designers can specify values describing the physical properties of characters and objects on the scene. To represent this combination of graphical and physical attributes, we defined a novel XML file format via a DTD (Document Type Definition). Thus, we devised a new *Blender* exporter plug-in. The next step was to import these scenes into the framework. This task is delegated to the loader module which analyzes the imported file. It sends graphical information to *OGRE*, physical information to *PhysX* and sounds to *OGREOggSound*. *OGRE* is responsible to manage the scene and to collect user input from devices. Thus, input is passed to *PhysX*, which manages the movement of characters, determines possible collisions and sends back to *OGRE* the response, in order to update the graphical representation. Another task was the camera management. The final touch was sound management, to add soundtracks and dynamic sounds.

ExDotScene. The first step, adopting our framework, is to create the virtual environment, which could be designed by using *Blender*. Graphics are stored as meshes, scenes are represented adopting the designed scene graph, and physics is additionally included. The standard *DotScene* format does not contain meshes or textures, but just an XML description of elements on the scene. Our framework supports: i) the creation of physical objects with a graphical representation

(bodies) and without (actors); ii) static and dynamic objects instantiation; iii) serialization of multiple 3D scenes. Actually, the last point was not natively supported by the built-in `DotSceneInterface` library. We developed an improved version, namely `ExDotSceneInterface`, to allow loading and saving multiple scenes supporting the following elements: i) nodes of the scene graph; ii) graphical entities (i.e., meshes, textures and materials); iii) lighting (i.e., lights on the scene with properties like positioning, direction, brightness, etc.); iv) camera (i.e., source node and target node pointed by the camera, its positioning and orientation); v) scene attributes (e.g., in/out-door, type of shading, clipping distance, etc.); vi) representation of the physics of the scene (e.g., serialization of static and dynamic actors and bodies, models of physics, etc.). According to requirements of the scene loader, which should support the physics, a DTD has been defined, namely `ExDotScene`, which extends the official `DotScene` DTD. A body element contains a shape element and an “actorParam” element. The shape element must contain one and only one element chosen among a list comprising cube, capsule, sphere, convex and triangular meshes. These elements are described by typical attributes like dimensions, etc. Moreover, it is possible to specify additional parameters for the shapes, using a “shapeParam” tag.

Blender Exporter Plug-in. This component recursively analyzes each node of the graph scene, considering information relative to its position, orientation and scale. Also specific elements, such as lights, cameras or meshes are taken into account during this process. Moreover, the framework gives the possibility to export only the properties of objects that satisfy some specific conditions. This way, designers can choose to export physical properties related to the entire scene, or not. Acting on the “Logic Properties” panel of *Blender*, they can set the following values in order to define physical properties of each object: i) *body*, if set to false the object is an actor; ii) *shape*, suggests the type of shape to use for the object; iii) *static*, sets if the physical object is static or dynamic; iv) *mass*, sets the mass -only for dynamic objects-; v) *skin*, suggests the value of the skin width that the physics engine will use during the simulation; vi) *file*, determines the .nxs file to be adopted as a physical representation of the convex or triangular shapes. At the end of the export phase, the scene is stored as an XML file, using the previously discussed `ExDotScene` DTD, which includes the physics.

Scene Importer. The core of the scene importer with physics support is composed of the `DotSceneProcessor` class, plus a number of other node processor classes. `DotSceneProcessor` contains a list of objects and allows several scenes to be loaded. Formats different from the `DotScene` are supported, albeit overriding the correct methods for loading the scene. Our extension supports the `ExDotScene` format.

Character Controller. This component, which was designed exploiting some key aspects of the physics engine, includes the following features: i) creation

of the principal and secondary characters (i.e., creating and tracking these instances, managing the interaction among them and between characters and the environment); ii) update of the character status (i.e., managing movements and actions, at each frame, and synchronizing the graphical mesh of the character w.r.t. the actual status); iii) character auto-stepping (to avoid characters get stuck in minor terrain bumps); iv) character walkable parts (i.e., definition of those areas of the environment that are not accessible to characters, acting like boundaries); v) modifiable bounding volumes (to simulate the crouching or groveling of the character); vi) character callback (i.e., response to collisions). The control system of the character is composed of three main classes: `GameCharacter`, `GameCharacterController` and `GameCharacterHitReport`. By using *PhysX*, it is possible to divide actors by groups, so as to manage and set different behaviors during the collision detection and response, accordingly.

Game Character Controller. The `GameCharacterController` implements the Singleton pattern so to ensure that only one instance of the class may exist inside the application. This class is interfaced with the *PhysX* library and deals with the instantiation and the management of the character. During the render loop of *OGRE*, the `GameFrameListener` class invokes a function of the `GameCharacterController` in order to execute the “simulate” and “render” methods over all characters recorded in the controller.

Game Character. The `GameCharacter` class represents the player inside the physical scene. It is possible to set the associated graphical mesh and the scene node in order to synchronize the visual and the physical representation. A specific method deals with setting a new movement direction: a vector representing the three-dimensional components of the velocity of the character is used. At a given moment the position of the character depends not only on the user input, but also on the velocity. During the simulation cycle, the “simulate” method is used to update the physical shape and the “render” method to update the visual representation of the character by setting the position of the mesh according to the position, expressed in global coordinates, of the physical shape.

Character Hit Report. As a consequence of a collision, *PhysX* generates a `HitReport` event. The `GameCharacterHitReport` class allows the setting of customized callback actions as a consequence of the occurrence of a collision, determining the actor with which the collision occurred and the group the actor belongs to. If the actor is a dynamic object and belongs to a group the character may interact with, the impulse that must be applied to the actor is calculated.

Camera. Once a 3D virtual scene is designed and imported inside the framework, developers define the way users, using their characters, explore this world, the so called *camera system*. In general, the core of the camera system is composed of two scene nodes acting as point of view and target point of the camera. We suppose that the node connected to the camera always points to the target node. This way, the movement of the target node produces the movement of the camera node. Moreover, it is possible to move the camera around the target

object directly moving the camera node, thus obtaining particular cameraworks. This simple system allows three cameras, namely *chasing*, *fixed* and *first-person*. In the *chasing camera*, the target node is associated to the point the character is looking at. A more distant view node makes the character as appearing displaced w.r.t. the center of the scene. This way, the character is still visible but we obtain the effect of a side view of the scene. The *fixed camera* is similar. The target node is the point the character is observing, but the camera is fixed and cannot be moved. This system is used in several videogames (e.g., third-person games). In *first-person camera*, the scene node of the character is the camera node. The camera is independent from any other object on the scene, thus the scene coordinates are used to update the position of camera and target nodes.

4.3 Modular Structure

The framework has a modular structure, being composed of a series of packages each dealing with a certain functionality, as shown in Figure 11. The binding process is obtained developing the required interface classes. Sometimes, this introduces a useful level of abstraction in the implementation of the functionalities provided by each component. Each package could be replaced with other components to satisfy particular requirements. In other words, the framework can be customized without re-implementing all the functionalities, if the exposed interface of the replaced package is properly refactored. This is one of the advantages while using our framework w.r.t. other common frameworks (e.g., Unity). A brief description of the main packages follows.

GameSystem. It contains a series of classes dealing with the management of the rendering cycle. Within *OGRE*, it is possible to define some classes which detect changes before and after a frame has been rendered on the screen. To exploit this feature, it is necessary to register the various frame listeners in the object of *OGRE*. The class *GameFrameListener* controls the order of execution of frame listeners, stored as an ordered list. An instance of *GameFrameListener* is then registered in the root object, thus ensuring that each frame listener is executed in the correct order.

GameIO. It manages I/O devices using the Object-Oriented Input System (*OIS*) library. *OIS* provides two modes to manage the input, namely unbuffered and buffered. The latter input mode is safer, because the former could not reveal an event. *GameIO* implements the buffered mode using the *GameInputManager* class. *GameKeyListener* and *GameMouseListener* implement the *OGRE* corresponding interfaces and define the actions executed as a consequence of events produced by the input device(s).

GameAudio. It is developed as an interface with the *OpenAL* library in order to manage audio. To do so, we developed a wrapper class using *OGREOggSound*, which provides methods to integrate *OpenAL* features (see Section 4.1).

GameSceneLoader. It provides those methods needed to load and save a scene. It could be adopted to load *ExDotScene* based scenes, or, eventually, methods could be overloaded to manage different formats (see Section 4.2).

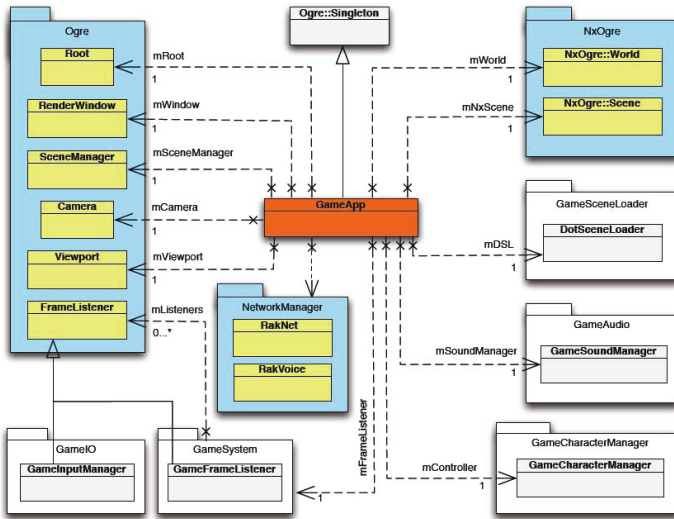


Fig. 1. The modular structure of the framework

GameCharacterController. It is the interface with the control system of the character developed with *PhysX*. It adopts a wrapper class to reproduce physics effects via *PhysX* and to manage the character controller (see Section 4.2).

RakNet. We integrated a cross-platform library for networking, which provides support for TCP and UDP communications. It also includes *RakVoice*, a toolkit for VoIP support and for real-time communications during game sessions which relies on the sound engine to reproduce sounds.

4.4 The Final Touch

The virtual world we earlier described still lacks of some features: first of all, logic has to be completely defined. The framework does not substitute developers in this aspect; this because we believe that a valid product should be designed even in details and the game logic can not be a surrogate of preset patterns. Moreover, the artificial intelligence of not-playing characters must be defined from scratch. The framework supports AI scripts coded in *Python*. The framework provides default implementation for the management of I/O devices but is not necessarily the optimal solution for any requirement. Similar arguments hold for other aspects, like the character control system or the camera system. Concluding, the strength of our platform is its modular nature, which ensures that developers could take the best from each component, but could also replace functions using other components, providing better, or simply different functionalities.

5 Conclusions

In this paper we presented the design and implementation of an academic framework to support the development of 3D virtual environments, helpful in those cases in which teams would like to save time in technical aspects, to focus on the contents. Our contribution can be summarized as follows: we presented a novel approach to create 3D virtual worlds enriched by physics effects. Our solution introduces the definition of physical properties of elements appearing on the scene, within the scene itself. These directives are interpreted by the physics engine, which is strictly interfaced with the rendering engine, reproducing physics effects. We additionally defined an extension of DTD for the *DotScene* format, introducing the support for the physics definition. This work represents the basis for future extensions and has already been adopted to show some techniques of design of virtual environments [2], e.g. the inclusion of environmental effects such as weather, day/night simulation and particle effects, exploiting techniques for terrain generation, realistic management of the water and fluid dynamics and the adoption of new rendering algorithms. Future work includes its adoption in different fields of application such as i) virtual/augmented reality; ii) virtual tours, reconstructions and museums; iii) engineering and architectural simulations.

References

1. Braun, H., Hocevar, R., Queiroz, R., Cohen, M., Moreira, J., Jacques, J., Braun, A., Musse, S., Samadani, R.: VhCVE: A Collaborative Virtual Environment Including Facial Animation and Computer Vision. In: Games and Digital Entertainment, pp. 207–213 (2010)
2. Catanese, S., Ferrara, E., Fiumara, G., Pagano, F.: Rendering of 3d dynamic virtual environments. In: Proceedings of the 4th International ICST Conference on Simulation Tools and Techniques (2011)
3. Fuchs, H., Kedem, Z.M., Naylor, B.F.: On visible surface generation by a priori tree structures. *Computer Graphics*, 124–133 (1980)
4. Graham, T.C.N., Roberts, W.: Toward Quality-Driven Development of 3D Computer Games. In: Doherty, G., Blandford, A. (eds.) DSVIS 2006. LNCS, vol. 4323, pp. 248–261. Springer, Heidelberg (2007)
5. Hu-xiong, L., Guan-dong, X., Zhang, Y.: Solution for developing data communication module on networking game platform. *Comp. Eng. and Design* 11 (2005)
6. Liu, X., Du, H., Wang, H., Yang, G.: Design and development of a distributed Virtual Reality system. In: International Conference on Machine Learning and Cybernetics, vol. 2, pp. 889–894 (2009)
7. Meagher, D.: Geometric modeling using octree encoding. *Computer Graphics and Image Processing* 19(2), 129–147 (1982)
8. Witkin, A.: Physically Based Modeling: Principles and Practice Particle System Dynamics. SIGGRAPH Course notes (1997)

The Mobile Orchestra Explorer

Donald Glowinski, Maurizio Mancini, and Alberto Massari

InfoMus Lab, DIST, University of Genova, Italy
{donald,maurizio}@infomus.org, alby@infomus.dist.unige.it
<http://www.infomus.org>

Active listening is a new concept in Human-Computer Interaction in which novel paradigms for expressive multimodal interfaces have been developed [1], empowering users to interact with and shape the audio content by intervening actively into the experience. Active listening applications are implemented using non-invasive technology and are based on natural gesture interaction [2].

The goal of this paper is to present the Mobile Orchestra Explorer application, developed in the framework of the EU Project MIROR [3].

The Mobile Orchestra Explorer application entails the user to set up the position of a virtual orchestra instruments/sections and then to explore the resulting virtual ensemble by walking through the orchestra space.

This application was tested during during the Festival of Science, a public event hold annually in Genova, Italy. Evaluation was carried out to study system usability and to produce an in-depth description of the user experience.

In the Mobile Orchestra Explorer application the user interacts with the system in two consecutive phases: (i) on the first phase, he/she walks in a sensitive empty space (a theater stage) holding he/she mobile phone in his/her hand and selects orchestra instruments/sections name on the mobile phone screen; when he/she reaches the point of the space in which he/she wants to place an instrument/section he/she press a button on the mobile phone to record its position; (ii) on the second phase he/she is allowed to move in the sensitive space and, as soon as he/she approaches an instrument/section position the corresponding pre-recorded audio track is played back. During both phases the user position is tracked by a fixed infrared camera.

The scenario architecture is represented in Figure 1. A fixed camera grabs frames of the theater stage at 25 frames per second and sends them to the SAME platform on which EyesWeb XMI is running. EyesWeb extracts the user silhouette from the frame background and computes the user barycenter position, relative to the orchestra space. The user, by touching buttons on the screen of his/her mobile phone, sends the following commands to SAME platform:

¹ The work presented in this paper has been partially supported by the EU FP7 ICT Collaborative Project MIROR (Musical Interaction Relying On Reflexion) Grant n°258338, <http://www.mirrorproject.eu>.

command name	possible value	description
<i>mode</i>	<i>0, 1</i>	indicates whether the interaction mode is either <i>setup</i> (0) or <i>explore</i> (1): the user either moves in the orchestra space to arrange the position of instruments/sections or is exploring the orchestra space and listens to the instrument/sections that are close to his/her current position.
<i>instrument</i>	<i>name</i>	(works only in setup mode): the user selects the instrument (or section) indicated by the parameter <i>name</i> .
<i>set</i>	<i>x, y</i>	the user sets the currently selected instrument (or section) position to the current user's position obtained by the camera frame.

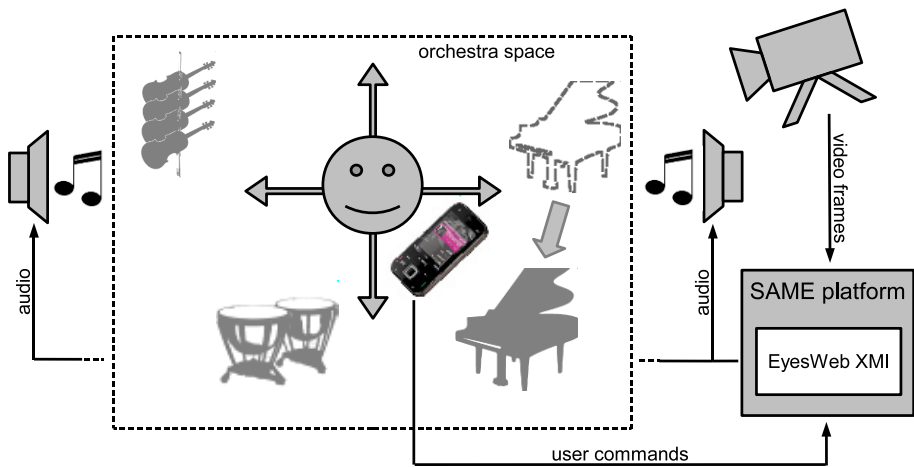


Fig. 1. The Mobile Orchestra Explorer application architecture

References

1. Camurri, A., Canepa, C., Coletta, P., Mazzarino, B., Volpe, G.: *Mappe per Affetti Erranti: a Multimodal System for Social Active Listening and Expressive Performance*. In: Proceedings of the 8th International Conference on New Interfaces for Musical Expression (2007)
2. Camurri, A., Volpe, G., Vinet, H., Bresin, R., Fabiani, M., Dubus, G., Maestre, E., Llop, J., Kleimola, J., Oksanen, S., Välimäki, V., Seppanen, J.: *User-Centric Context-Aware Mobile Applications for Embodied Music Listening*. In: Daras, P., Ibarra, O.M. (eds.) UCMedia 2009. LNICST, vol. 40, pp. 21–30. Springer, Heidelberg (2010)

Realtime Expressive Movement Detection Using the EyesWeb XMI Platform

Maurizio Mancini, Donald Glowinski, and Alberto Massari

InfoMus Lab, DIST, University of Genova, Italy
{maurizio,donald}@infomus.org, alby@infomus.dist.unige.it

In the last few years one of the key issues in Human Computer Interaction is the design and creation of a new type of interfaces, able to adapt HCI to human-human communication capabilities. In this direction the ability of computers to detect and synthesize human *expressivity* of behavior is particularly relevant, that is, computers must be equipped with interfaces able to establish a *Sensitive* interaction with the user (see [3]).

We present a system for realtime analysis of expressivity features in human movement and mapping of these features on a two dimensional space on which we identify four emotions/attitudes: *anger*, *joy*, *sadness* and *relief* [1]. Figure 1 illustrates the structure of our realtime expressivity analysis system:

- we track the body configuration of a user moving in a room using a Kinect controller [2]. This device has been chosen as open drivers are available (<http://www.openni.org>) providing realtime tracking of user's body sections (head, shoulders, hips, arms, hands, legs) in both 2D and 3D coordinates. We analyze 2D data in realtime using the EyesWeb XMI platform [1].
- *smoothness*: from user's left and right hand position we compute smoothness as the correlation between each hand's trajectory curvature (k) and velocity (v).
- *Quantity of Motion (QoM)*: it is an approximation of the amount of detected movement, based on Silhouette Motion Images.

The computed smoothness and QoM are dynamically plotted on a map, as shown in figure 1, on which we highlight some attitudes/emotional states like anger, joy, sadness and relief. The proposed system may be suitable for concrete applications in affective computing, multimodal interfaces, and user centric media applications.

An example of the realtime extraction of expressive features can be downloaded at:

<ftp://ftp.infomus.org/Pub/ftp-user-root/i-search-demo-expressivity.mp4>

¹ This work is partly supported by the EU FP7 ICT Project I-SEARCH (A unified framework for multimodal context Search; n°248296; DG INFSO Networked Media Unit; 2010-2012).

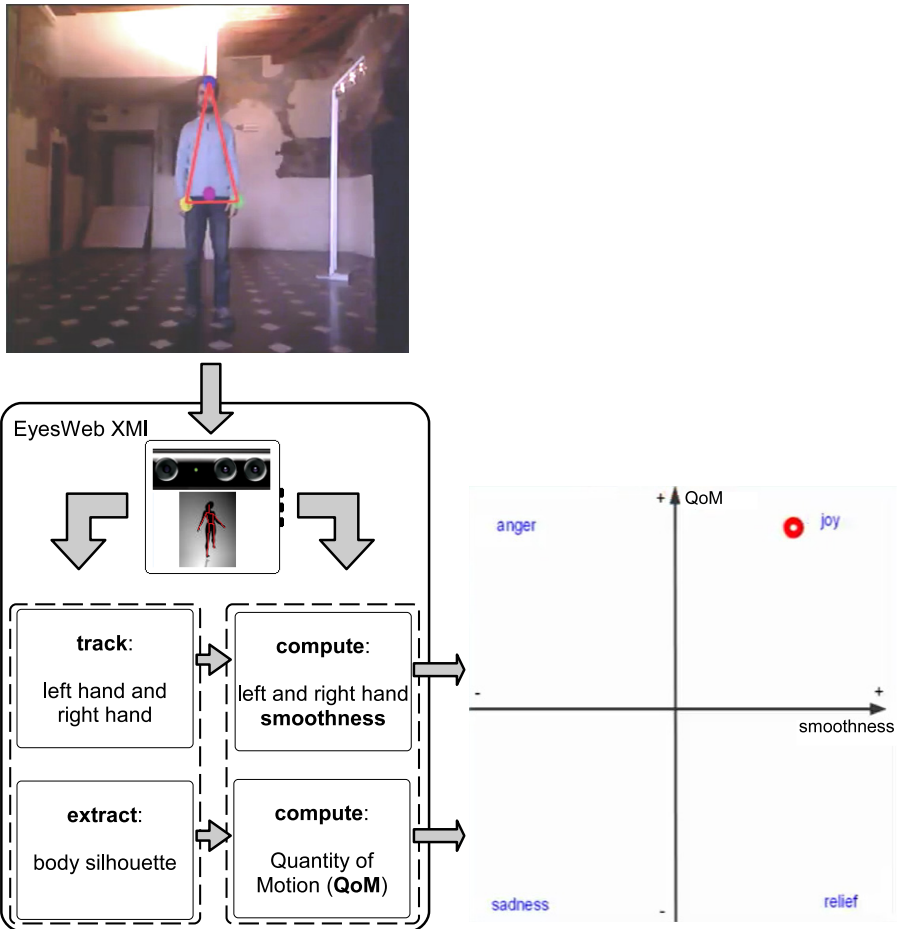


Fig. 1. The realtime expressive movement detection system we present. User image is captured by a camera, user body silhouette is extracted and left/right hand position is tracked. Finally we compute user’s Quantity of Motion and hands smoothness and we draw the resulting values on a plane.

References

1. EyesWeb, <http://www.eyesweb.org>
2. Kinect, <http://www.xbox.com/en-US/kinect>
3. Zeng, Z., Pantic, M., Roisman, G.I., Huang, T.S.: A survey of affect recognition methods: audio, visual and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31(1) (2009)

i-Theatre: Tangible Interactive Storytelling

Jesús Muñoz, Michele Marchesoni, and Cristina Costa

CREATE-NET

Via alla Cascata 56/D Povo, 38123, Trento, Italy

{jesus.munoz,michele.marchesoni,crisrina.costa}@create-net.org

Abstract. Storytelling is fundamental for the cognitive and emotional development of children. New technologies combined with playful learning can be an effective instrument for developing narrative skills. In this paper we describe i-Theatre, a collaborative storytelling system designed for pre-school children: with it, it is possible to use characters and scenarios drawn on paper for creating a digital story, using simple animation techniques and recording voices and sounds. For implementing it, we combined a multitouch surface with a set of tangible objects. This choice allowed lowering the learning effort of a new interface, letting the child to be immersed directly into the storytelling process from the very beginning.

Keywords: Storytelling, multitouch, children, education, tangible technologies, collaboration.

1 Introduction

Pedagogists universally agree on the fundamental role that storytelling plays in children's growth: not just listening to stories in early age is important for oral language development, but also encouraging children to tell their own stories help them to learn how to effectively use the language and to structure their thoughts. New media and interaction technologies can be used as facilitator for eliciting children telling their own stories and developing narration skills.

During the last years, various researchers proposed different approaches to technology-mediated children storytelling. In an early work, Cassell et al. [2] created Rosebud, a computationally-augmented toy to which the child could tell their stories, thus using a familiar mode of interaction for encouraging children to write, edit, collaborate, and share their stories. This work demonstrated the enormous potential of using tangible and natural interfaces for introducing children to technologies.

More recent works by various authors used a great variety of technologies for supporting children's story creation and expression. For example, Sugimoto et al. [1] proposed GENTORO, a system thought for elementary age children based on a robot and a handheld projector, while Cao et al. [4] proposed TellTable, a multitouch surface system. Fails and Guha [3] created Mobile Stories, an

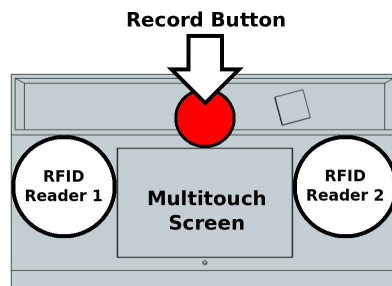
application designed for mobile devices for creating cooperative *shared narratives*. In our previous work [6], we used hand puppets as tangible interface for creating a fully immersive storytelling experience. Finally, Botturi and alt. [5] used various technologies as instruments for motivating children with special needs to participate to creative activities and for expanding children’s expressive choices.

In this paper, we present i-Theatre, a system for storytelling designed for one or two kids who can collaborate for creating, storing and sharing their own stories in a multimedia format. i-Theatre was created using the combination of two main technologies: a multi-touch surface and tangible objects. It looks like a coffee table with a integrated multitouch surface (Fig. 1a). The table embeds also a scanner, two Radio Frequency Identification (RFID) readers Fig. 1b, microphone and speakers and a set of tangible objects (Fig. 2a and Fig. 3a). All this elements provide an abstract layer to use the system thus providing an engaging and easy to use interface.

Our system was conceived for developing narration skills in children in the last year of preschool. While most of the previous works target elementary school children, a smaller number are dedicated to younger children. For this reason a special attention to the needs and skills of our very young users have been taken into account during the system design. The graphical user interface used in the multitouch surface is essential and very intuitive, based on very few elements, while the interaction is mainly achieved through the use of tangible objects. Research conducted by Xu and alt. [7] on the use of Tangible User Interfaces (TUIs) in assisting childrens learning highlighted the little time needed for learning this type of interfaces, due to our inborn ability to manipulate objects tangibly with little cognitive effort. The choice of using a TUI in i-Theatre a as main interaction tool is therefore motivated by our intend of immersing directly the child into the storytelling process, avoiding the initial difficulty of learning a new interface, and at the same time providing an enjoyable environment together with the necessary tools for expressing thoughts and experiences.



(a) i-Theatre prototype



(b) i-Theatre table surface

Fig. 1. The i-Theatre table

2 The i-Theatre Storytelling Process

The aim of the system is to stimulate the child to tell tales using a gradual approach. The first impact with the i-Theatre table is typically very playful: children experiment moving characters around the tangible surface, resizing and rotating them. Then they start creating stories: at the beginning very simple, composed by one single scene; as they get more confident with the system and their own narration skills, stories get more and more articulated, composed by several scenes. A typical story creation process using i-Theatre goes through five stages, each one of them has associated a number of tangible objects:

Stage 1: Creating and saving the story elements

Initially children are invited to create the characters and backgrounds for their story: they can draw on paper their own story elements using various techniques, select them from books, create collages or use small objects. The selected material is then digitalized as pictures that are visualized on the multitouch surface. Eventually, they can be stored for later using their own *Personal Archive*, a tangible object that represents their own personal space where to store all the created content (Fig. 2a). At this stage, children is free to explore the system and get familiarized to it: characters can be moved around the surface, be resized and rotated.

Stage 2: Selecting the elements for the story

When the child is ready, he can select the audio-visual elements that will be used for telling the story from his *Personal Archive*. All the choosen elements will be available later on, during recording.

Stage 3: Telling the story

For starting the registration the children plugs the *Scene Container* (Fig. 3a) into the table: at this point the system is ready to acquire a scene of the story. A *Record Button* is used for starting and stopping the recording session: when it is pressed, all the drawings' movements and animations are grabbed and voices and sounds recorded by the microphone. During the narration *characters* can be dragged in or out the backstage. Then, when the *Record Button* is pressed again, the *Scene Container* will contain a part of the story. Additional scenes can be created by introducing new empty containers.

Stage 4: Select the definitive story order

After that all of the scenes have been recorded, a complete story can created by plugging the *Scene Containers* to each other. The story sequence is defined by the *Scene Containers*' plugging order (Fig. 3b), since there is a direct relation between each tangible object and a piece of story. *Scene Containers* can be plugged into different sequences or substituted to obtain new stories.

Stage 5: Storing the story

When the story is finished the kid can save the created movie inside his *Personal Archive* and watch his story later.

It is not necessary to perform all these phases sequentially from beginning to the end: for example the children can start from already available content stored in the system, such as backgrounds or scenes, or can go back and through from one stage to another for adding a missing part or refining their work.

3 Operation Modes

The i-Theatre system is characterized by different states or operation modes, each of them is related to different moments in the story creation process. A system state is determined by the set of tangible object in use: thought the manipulation of these objects it is possible to switch between different operation modes, thus avoiding the use of a GUI (Graphical User Interface) based menu. These operation modes are listed below:

- *Exploration Mode*: in this operation mode no tangible objects are used. Children can explore and familiarize with the system, moving and playing with the characters on the surface. It is possible acquire new visual content, such as hand made drawings or printed photos, using the scanner embedded into the table.
- *Creation Mode*: children can browse and select stored multimedia material or save new one. The *Creation Mode* starts when at least one *Personal Archive* is placed over the reader (see Fig. 2b)
- *Action Mode*: in this operation mode it is possible register a scene of the story. This mode starts when a kid plugs the *Scene Container* (see Fig. 3) object.
- *Production Mode*: allows children to determine and change the event order on a story. When a kid wants create the story sequence, he must chain some *Scene Containers*.

4 Tangible Interface

For supporting the story creation processing, a set of tangible objects have been introduced and integrated into the system.

- A scanner embedded in the i-Theatre table offers a simple method for digitalizing visual content. A picture can be scanned as a *character* or *backdrop*: in the first case, it can be rotated, moved or scaled, while in the second case the image remains fixed on the background. The selection of the image use is done just before picture acquisition, when the scanning command is given.

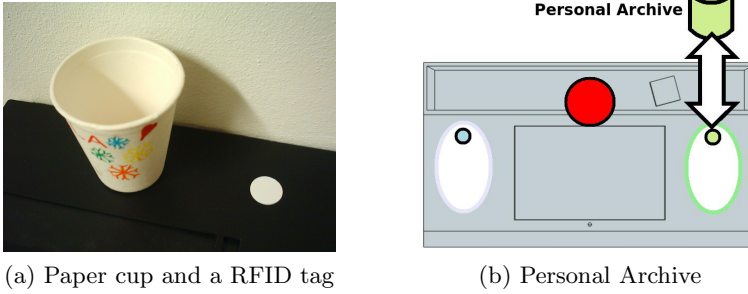


Fig. 2. (a) For this prototype, paper cup has been used like a metaphor of Personal Archive. Each cup is identified by an RFID tag. (b) Placing the Personal Archive over an RFID reader, the screen shows the content.

- The *Personal Archive* serve as interface for entering into the *Creation Mode* and for storing multimedia content into a personal space. It uses RFID (Radio Frequency Identification) technology for identification thus providing a low-cost and contact-less interface for organizing content: each child has her own identifier associated to an RFID tag, that is attached to a personal object that can be an artifact created by the child (e.g. a decorated paper cup) or any other personalized item. The *Personal Archive* is used for associating the child’s identifier to the selected content: when it is put on the reader, the kid can view his own stored images or store new ones. Two RFID antennas installed beneath the table top allow access from any side of the table [Fig. 15](#) and consent up to two kids to store or select multimedia items simultaneously.

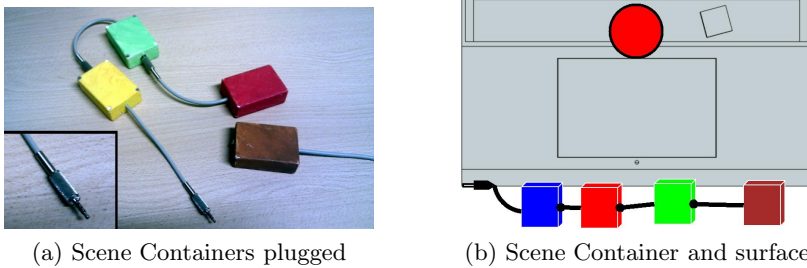


Fig. 3. (a) The Scene Container is a colored box with a male and female jack connectors to allow cascading between them. (b) The cascaded Scene Containers can be connected to the table in order to obtain the desired story order.

- The *Scene Containers* [Fig. 3](#) consent to enter into the *Action Mode* and each of them can be used for storing a single scene of the story. This tangible object contains an unique identification code (ID) that the system associates

to a recorded scene. For the implementation of the *Scene Containers*, customized hardware was designed instead of using RFID technology. This because we needed to obtain the ordered sequence of IDs from the cascaded plugged objects, for being able to read the correct order of story scenes.

5 Conclusion and Future Work

We plan to conduct extensive user studies of the usage of the i-Theatre prototype by preschoolers to test such questions as how i-Theatre usability can be improved, how well i-Theatre elicits storytelling in children, and if its usage in a classroom and in a museum laboratory contexts can be successful.

The first tests conducted are encouraging and showed that in the studied cases the chosen approach is effective in engaging children and motivating them in experimenting different storytelling solutions. Children not only showed interest in creating their own stories but also explored the available tools for improving their narrative. We used these experiences for improving the initial research prototype, in an iterative cycle of development and re-design of system. Finally, thanks to the input from teachers, we identified new features to be implemented for better supporting classroom activities in a flexible way. The i-Theatre is continually in evolution, making technology funny and accessible for the youngest.

Acknowledgements. This paper is based upon work carried out during the i-Theatre project and partially funded by Province of Trento (Italy). We thank Edutech and all the partners involved in the project (MSTN and FBK) for their time and dedication.

References

1. Masaori, S., Toshitaka, I., Tuan, N., Shigenori, I.: GENTORO: A System for Supporting Children's Storytelling using Handheld Projectors and a Robot. In: IDC 2009, Como, Italy (2009)
2. Glos, J.W., Cassell, J.: Rosebud: Technological Toys for Storytelling. In: CHI 1997. Electronic Publications (1997)
3. Fails, J.A., Druin, A., Guha, M.L.: Interactive Storytelling: Interacting with People, Environment, and Technology. In: IDC 2010, Barcelona, Spain (2010)
4. Cao, X., Lindley, S.E., Helmes, J., Sellen, A.: Telling the Whole Story: Anticipation, Inspiration and Reputation in Field Deployment of Tell Table. In: CSCW 2010, Savannah, Georgia, USA (2010)
5. Botturi, L., Bramani, C., Corbino, S.: Stories, Drawings and Digital Storytelling: a Voice for Children with Special Education Needs. In: IDC 2010, Barcelona, Spain (2010)
6. Mayora, O., Costa, C., Papliatseyeu, A.: iTheater Puppets Tangible Interactions for Storytelling. In: Nijholt, A., Reidsma, D., Hondorp, H. (eds.) INTETAIN 2009. LNICST, vol. 9, pp. 110–118. Springer, Heidelberg (2009)
7. Xu, D., Mazzone, E., MacFarlane, S.: Informant Design with Children - Designing Children's Tangible Technology. In: IDC 2007, Aalborg, Denmark (2007)

An Invisible Line: Remote Communication Using Expressive Behavior

Andrea Cera¹, Andrew Gerzso¹, Corrado Canepa², Maurizio Mancini²,
Donald Glowinski², Simone Ghisio², Paolo Coletta², and Antonio Camurri²

¹ Ircam, Centre Pompidou, Paris, France

andreawax@yahoo.it, Andrew.Gerzso@ircam.fr

² InfoMus Lab, DIST, University of Genova, Italy

{maurizio,donald,ghisio}@infomus.org, paolo.coletta@unige.it

An Invisible Line is an installation focusing on the remote communication between 2 human users, based on the analysis of full-body expressivity. It aims at creating shared, networked, social experiences. It is the result of a scientific and artistic collaboration between Casa Paganini - InfoMus Lab (Genova, Italy), IRCAM (Paris, Italy) and The Hochschule für Musik und Theater (Hamburg, Germany).

The center of the experiment is an investigation on the resonance between the non-verbal motoric behaviors of two remote participants, mediated by computers. Instead of focusing on some form of realistic representation of the remote body, *An Invisible Line* studies how this relation could be achieved via abstract displays, expressing the way a computer is interpreting the relation between two people's movements. The visual feedback is purposely lacking: you cannot clearly see your remote partner, and the image of yourself is fragmented too. The audio feedback is a series of auditory displays, sometimes very basic and simple, sometimes more metaphorical, which sonify the machine's analysis of the two body's relations.

We describe the procedure followed by users interacting through the *Invisible Line* installation. Two installation environments are set up in separate locations. Each environment consists of: a camera running at 25 fps, a projection screen, a workstation running EyesWeb XMI (www.eyesweb.org) and Max/MSP (cycling74.com), see Figure [1](#). Let us consider two users, called Andrea and Barbara, acting in each of the locations:

- Andrea enters the first installation environments and places himself in front of the screen: he watches a full body image of himself on the screen; when he moves he hears sounds that respond to his movement's characteristics (e.g., if he moves in a jerky way then he hears discontinuous and harsh sounds). In the remote environment, where Barbara has not yet placed herself on front of the screen, the full body image of Andrea is projected as a white "silhouette".
- Barbara enters the second installation environment: each user watches an image, split into two halves along an invisible vertical line, one half

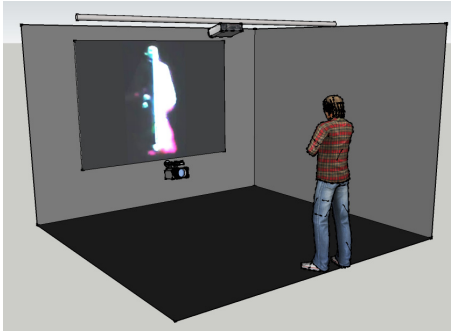


Fig. 1. A representation of the Invisible Line installation environment

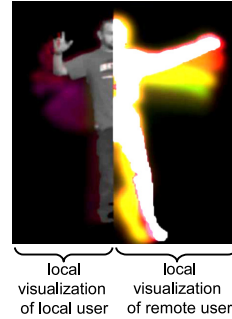


Fig. 2. An example of interaction in the Invisible Line installation

representing the user's mirror image and other half represents the other user's mirror image.

- Users' behavior is analyzed in realtime by specific modules of the EyesWeb XMI platform (the Expressive Gesture and Social Libraries) and the following movement features are computed:
 1. *Smoothness Index of the hand and foot* (SmI-h and SmI-f): distinguishes between continuous (smooth) versus discontinuous (jerky) movement.
 2. *Contraction Index* (CI): indicates whether arms are close to the body or stretched outside.
 3. *Symmetry Index* (SyI): indicates whether user's left and right silhouette halves are superimposable or not.
 4. *Quantity of Motion* (QoM): it represents the energy of movement computed as the pixel-based difference between two consecutive silhouette frames.
 5. *Synchronization Index* (SI) on QoM: it is computed between local and remote user's Quantity of Motion: SI is high when users' movement varies synchronously and SI is low in the opposite case. Only periodicity of movement is considered, ignoring the phase.
- Features 1-4 are mapped to 4 higher level semantic categories of *attitude*: *static*, *sweet*, *agitate*, *violent*. Simple rules have been established to determine attitudes from features: for example high Smoothness and low QoM correspond to the attitude *sweet*. In the near future we aim to further refine these rules.
- If, during a fixed time span, the users' attitude is the same and their movement is *synchronized* then the interaction is considered successful and users receive their prize: Barbara receives a picture of Andrea, and Andrea receives a picture of Barbara. Synchronization is estimated by observing if the two users' QoM varies in time in the same way or not.

Teaching by Means of a Technologically Augmented Environment: The Stanza Logo-Motoria

Serena Zanolla¹, Antonio Rodà¹, Filippo Romano¹, Francesco Scattolin¹,
Gian Luca Foresti¹, Sergio Canazza², Corrado Canepa³,
Paolo Coletta³, and Gualtiero Volpe³

¹ University of Udine

{serena.zanolla,antonio.roda,gianluca.foresti}@uniud.it
romano.filippo@spes.uniud.it
scattolin.francesco@gmail.com

² University of Padova

canazza@dei.unipd.it

³ University of Genova

corrado@infomus.dist.unige.it

{paolo.coletta,gualtiero.volpe}@unige.it

Abstract. The Stanza Logo-Motoria is an interactive multimodal environment, designed to support and aid learning in Primary Schools, with particular attention to children with Learning Disabilities. The system is permanently installed in a classroom of the “Elisa Frinta” Primary School in Gorizia where for over a year now, it has been used as an alternative and/or additional tool to traditional teaching strategies; the on-going experimentation is confirming the already excellent results previously assessed, in particular for ESL (English as a Second Language). The Stanza Logo-Motoria, also installed for scientific research purposes at the Engineering Information Department (DEI) of University of Padova, has sparked the interest of teachers, students and educationalists and makes us believe that this is but the beginning of a path, which could lead to the introduction of technologically augmented learning in schools.

Keywords: Stanza Logo-Motoria, interactive and multimodal environment, augmented reality, augmented environment for teaching, Learning Disability.

1 Introduction

The Stanza Logo-Motoria [1] consists of an empty room equipped with input and output conventional peripheral devices such as a standard webcam, two speakers and a video projector. The user’s body movements and gestures are captured by the webcam positioned on the ceiling and the video signal is processed by a EyesWeb patch to derive low-level features [3], used to define a) how the user occupies the space and b) the quality of gestures the user performs.

Finally, an Adobe Air application is used to render interactive audio-video contents. In the Stanza Logo-Motoria it is possible to subdivide the room in several areas. By entering the various areas and performing previously arranged gestures the user activates the connected auditory and/or visual content. Sounds, segments of spoken language, and complete sentences are used to enhance the motor-auditory experiences through which the user acknowledges non-linguistic communicational tools: body, images, sound and symbols as for their transcultural value. The main applications of the Stanza Logo-Motoria are aimed to support and aid learning in Primary Schools and children who experience severe disabilities or suffer from specific Learning Disabilities (used as a compensatory and dispensation tools in the case of LD) [2]. The system, by using standard hardware and simple strategies of mapping, has sparked the interest of students and teachers in more innovative ways of learning and teaching. The Stanza Logo-Motoria is suitable for the school environment thanks to its easy implementation; moreover, teachers are immediately involved in the design of activities due to the simplicity of mapping, which makes it immediately comprehensible. In fact, for over a year now, the use of the Stanza at school has shown that, by using the same basic scheme, it has been possible to develop in collaboration with teachers a great deal of educational activities involving several school subjects, such as English, History, Science, Music. In the following paragraphs we will explain two applications of the Stanza Logo-Motoria, called Resonant Memory and Fiaba Magica, in detail.

2 Resonant Memory Application

The Resonant Memory application allows the creation of a technologically augmented environment [5] to be used within all the subjects taught at school. The use of body movement associated with the sound widens the range of possibilities to access knowledge from an enactive point of view [6]. Teaching contents, by becoming “physical events”, which occur around the child by means of the child, activate the motor aspect of knowledge [7]. By means of the Resonant Memory application the space, acquired by the webcam, is subdivided in nine areas: eight peripheral and one central. The user’s presence within a specific area triggers the audio reproduction of the corresponding content. The child explores the “reactive space” by freely moving without using sensors:

- Noises, music and environmental sounds are synchronized to the peripheral areas.
- The central area is synchronized instead with an audio reproduction of the contents to be taught that contain the elements to be connected with sounds positioned in the various peripheral areas.

When the user enters the reactive space for the first time the application is in *exploration mode*: whenever the user reaches one of the eight peripheral areas, activating the connected sound content, the application stores this information and, if during this phase, the user reaches the central zone no sound event is triggered. When the user reaches the central area the Resonant Memory application

triggers the *story mode*: the system starts the audio reproduction of a story. The child listens to the content reproduced in the central area, reaches the different peripheral areas experimenting the sounds and finally, enjoying the game, “composes the soundtrack” of the lesson. If, during the listening, the user widens their arms the system pauses the playback of the story, which is interrupted until the user lowers their arms.

The Resonant Memory application creates a sound augmented space: in agreement with the teachers, who have work with the system, we think that, in general, children need to improve their listening ability experiencing the interactive space without being distracted by any visual references connected to the sound. The main aim is to achieve heightened awareness of the body movement in space. The sounds synchronized with the peripheral areas can be words, syllables, sentences, noises, music, etc.; these sounds do not change collocation in space: the same sound always corresponds to the same spatial area. Every sound is triggered by the presence of the user in the corresponding area. The child is motivated to carefully listen in order to insert the proper sound at the right time. The mere movement of the body in space, performed to activate the sound, stimulates the child to spontaneously mime situations, actions, characters and feelings. School children love to use the Resonant Memory application in pairs: together they decide where to go, how to move and what to say. It is also possible to use the application with a group: one child is the “explorer”, the others are the sounds; the explorer performs the exploration phase and, during the story phase, swaps his/her position with that of the children standing near the peripheral area containing the sound.

At school we have been testing the Stanza Logo-Motoria in Resonant Memory modality since February, following a quasi-experimental design (between subjects) with two comparable classes: two Third Classes. The quasi-experimental design requires a pre-test (February) and post-test (June) for a treated (experimental) and comparison (control) group. We intend to verify (experimental hypothesis) if students, by using the Stanza Logo-Motoria as a listening tool for learning English as a second language, improve significantly in word recognition and language comprehension than those who use passive listening by means of headphones or the 5.1 sound system.

3 Fiaba Magica Application

The aim of Fiaba Magica is to support a strengthening path of gestural intentionality of children with multi-disabilities. These children often are able to express communicative intentionality only by means of simple gestures and vocalizations, which can be enhanced and extended thanks to technology. Fiaba Magica is the opportunity to augment gestures with visual and sound stimuli bringing out all the intentional features. Fiaba Magica responds to the need to communicate of a child with limited speech abilities associated with motor impairments.

The Fiaba Magica application allows the user to reconstruct the sound and images of a tale by performing a simple gesture such as widening arms and moving within a specific space. Using Fiaba Magica a child enters a specific area of the Stanza and activates a) the audio reproduction of the first part of the story and b) the projection on the screen of the corresponding image (for example two characters). Once the audio reproduction of the first part of the story is played, by widening their arms, the user can play with the characters projected, animating them. Moving to the next area the child activates a) the projection of a new sequence of the story which includes another set of characters and b) the audio reproduction of the narrative sequence itself. The third advancement in the story is performed in the same way as the other two.

In this particular instance, it is mainly important to have the child understand that the audio/visual feedback is triggered by his/her movement. In a further evolution of the system the opportunity to make a choice could be decided: the gesture of lifting either the right or the left arm could be synchronized to two different developments of the story. The application could take the choice in account by offering two different continuations of the story within the second and the third area. In this way it would be possible to offer a, limited but important, option to change the plot of the story.

4 System Architecture

The Resonant Memory application is based on a software patch developed in the EyesWeb environment; the EyesWeb patch performs the video analysis and sound rendering tasks. In the input stage the signal from the webcam is processed in order to extract several low-level features related to the user's movements. Features extracted include the trajectory of the centre of mass, the motion index, and the contraction index. Background subtraction is achieved via a statistical approach: the brightness/chromaticity distortion method [4]. In the mapping stage the patch analyses these features and runs transitions among four states: exploration, story, pause, and reset. Finally, the output stage controls the playback of a set of pre-recorded audio files. In Fiaba Magica the real-time control and processing of the audio/video material is performed by an Adobe Air application that also provides a user GUI to configure the system. A Flosc server allows the communication between EyesWeb and Adobe Air.

The setup of the Stanza Logo-Motoria (fig. 1) consists of an empty room measuring a minimum of 5x5 metres; a webcam is installed in the centre of the ceiling at a height of minimum 4 metres. The webcam is connected to the computer by means of a USB 2.0 cable. The computer runs the EyesWeb applications, such as Resonant Memory or Fiaba Magica, which, analyzing the video signal coming from the webcam in real time, shape the space by means of sounds and images. The audio feedback is provided by two loudspeakers while the video feedback is given by means of a video projector. The environment has to be lighted by diffused lighting in order to avoid shadows on the floor.

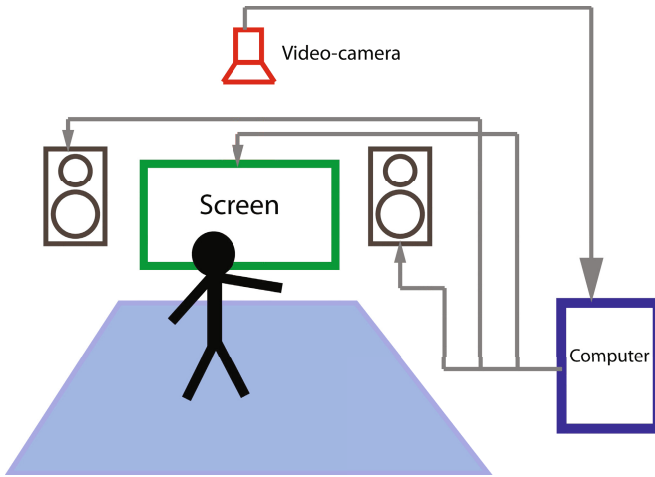


Fig. 1. The setup of the Stanza Logo-Motoria in Fiaba Magica modality

5 Future Developments

The great interest aroused by the Stanza as a teaching tool, testified also by the first results of the on-going experimentation, motivates us to further develop the system a) by introducing sound spatialization techniques in order to enhance the ability of sound localization, in particular for children with visual impairments and b) by integrating the Pittore Vocale [8] in the system, as a tool for increasing the knowledge of sound features starting from voice itself.

References

1. Camurri, A., Canazza, S., Canepa, C., Foresti, G.L., Rodà, A., Volpe, G., Zanolla, S.: The ‘Stanza Logo-Motoria’: an interactive environment for learning and communication. In: Proceedings of SMC Conference, Barcelona, vol. 51(8) (2010)
2. Gardner, H.: *Frames of Mind: The Theory of Multiple Intelligences*. Basic (1983)
3. Camurri, A., Moeslund, T.B.: Visual Gesture Recognition. From motion tracking to expressive gesture. In: Godøy, R.I., Leman, M. (eds.) *Appears as chp. 10 in the book Musical Gestures. Sound, Movement, and Meaning*. Routledge (2010) ISBN: 9780415998871
4. Horprasert, T., Harwood, D., Davis, L.: A Robust Background Subtraction and Shadow Detection. In: 4th ACCV, Taipei, Taiwan, vol. 1 (2000)
5. Azuma, R.: A survey of augmented reality. *Presence: Teleoperators and Virtual Environments* 6(4), 355–385 (1997)
6. Bruner, J.: *Processes of cognitive growth: Infancy*. Clark University Press, Worcester (1968)
7. Leman, M.: *Embodied Music Cognition and Mediation Technology*. The MIT Press (2007)
8. de Götzen, A., Marogna, R., Avanzini, F.: The voice painter. In: Proc. Int. Conf. on Enactive Interfaces, Pisa (November 2008)

INSIDE: Intuitive Sonic Interaction Design for Education and Entertainment

Alain Crevoisier and Cécile Picard-Limpens

Haute Ecole de Musique de Genève (HEM)
Rue de l'Arquebuse 12, CP 5155,
CH-1211 Genève, Switzerland
alain.crevoisier@hesge.ch, ccl.picard@gmail.com

Abstract. The project INSIDE - Intuitive Sonic Interaction Design for Education and Entertainment, aims at offering children and adults without previous musical experience the possibility to create sounds and make music on a very intuitive and playful manner. We develop a concept of tangible interaction using objects that can be placed on any conventional surface, like a table. The objects can be fitted with meaningful icons representing various aspects and functions related to sound and music, such as sound sources, sound modifiers, or mixers.

Keywords: interface, interaction, sound, education, entertainment.

1 Intuitive Sonic Interaction Design

We define as object any element that can be grasped and put on a surface. The project explores new ways of making music and creating sounds, by placing objects on the surface and combining them together. Each object is set to a meaning and function related to sound and music: sound sources (sounds of nature, animals, musical instruments, etc), sound modifiers (echo, reverb, filtering, etc.), players, mixers, controllers, etc. Playing sounds, mixing different sound sources, applying sound effects are performed by simply moving objects on the surface. We focus on inventing rules and interactive scenarios that are playful and intuitive, based on the way objects are combined together. For this purpose, we define three object relations: no relation, neighborhood, and object-in-object. Chains are created when two or more objects enter in relation and can be seen as sequences of objects ordered in the 2D space, as shown in Figure 2, bottom-right.

We developed an ad-hoc application, the Surface Editor¹, to facilitate the design of the interaction, by offering the possibility to set the behavior of objects very easily.

¹ <http://www.surface-editor.net>

The Surface Editor receives the object information via TUIO, using a simple webcam hooked to ReacTIVision². This way, various interactive scenarios can be tested and evaluated very conveniently. However, this application is transparent for users. Once an interactive scenario is loaded, the application is running in the background and no intervention from users is required, except for playing with the objects.

Optionally, multi-touch sensing can be added to the setup by using the Airplane controller³, a device developed in previous projects for transforming any ordinary surface into a multi-touch interface. This is offering extended interactive possibilities. For instance, an object representing a sound source can be played simply by touching it. However, since the Airplane controller is detecting touch by watching fingers crossing a plane of IR light placed a few millimeters above the surface, objects must be thin enough for not interfering with the plane.

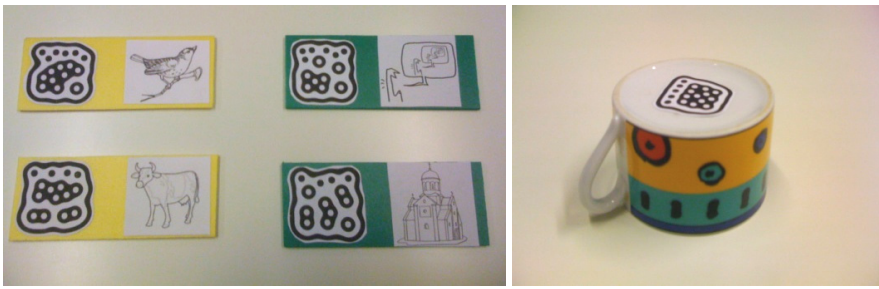


Fig. 1. Right : Examples of icons for objects involved in sound interactions scenarios for novice users: sound sources (animals) and sound modifiers (echo, reverb). Left : Everyday objects, such as a cup, can be used for the interaction.

Figure 1 shows examples of icons for objects involved in sound interactions scenarios. One interesting aspect is that everyday objects can be used in the interactions.

2 Description of the Demo

Our aim is to introduce sound design and music creation in a playful manner. For simplicity, we will not interact with touch gestures for the demo presented at INTETAIN. This offers also more freedom to choose thicker objects that are easier to grasp for young users. The interactive scenario we have defined for this demo involves three aspects related to sound and music, with a pedagogical perspective: playing sounds, creating sounds and recording sounds.

² <http://reactivision.sourceforge.net/>

³ www.future-instruments.net/fr/airplane.php

Playing sounds can be performed in three different ways, manually, with the ‘Loop Player’ object or with the ‘Circular Sequencer’ object. To play sounds manually, users have to hold their hand briefly above any ‘Sound’ object to hide its visual tag. When they remove the hand and the tag become visible again for the camera, then the sound is triggered. The volume of the sound is adjusted by rotating the object. It takes 180° to go from min to max, so there is never any abrupt jump in the value of the volume. If ‘Sound’ objects are put in the ‘Loop Player’ object, then they are repeated continuously. The ‘Loop Player’ object is a larger, thin object that can contain several smaller objects. It illustrates the object-in-object relation (see Figure 2, top-right). Finally, the ‘Circular Sequencer’ object allows constructing sequences by arranging ‘Sound’ objects around it. The timeline is similar to a clock hand doing a circular movement.

Creating sounds is performed by taking a base sound and adding ‘Modifier’ objects (of ‘Effect’ objects) next to it. This is creating a sound chain and multiple sound modifiers can be added. A special object, the ‘Transmitter’ object, can be added to the chain in order to be able to represent a whole chain by a single object. This is very practical to play the chain in one of the three ways described above. The emitter is put in the chain and the receiver then behaves like any other ‘Sound’ object. Thus it can be put in the ‘Loop Player’ object or used with the ‘Circular Sequencer’ object. ‘Modifier’ objects usually have a single parameter that users can adjust by rotating the object (for instance the feedback in the ‘Delay’ object).

In order to record sounds, users need to put a blank ‘Sound’ object in the ‘Record’ object, which is looking similar to the ‘Loop Player’ but with a red color. Any sound produced will be recorded in the blank ‘Sound’ object. Silence is not recorded, which means that if a sound is produced, followed by some silence and another sound, then only the last sound will be recorded.

All variables and information events are sent through OpenSoundControl⁴ (OSC) using the Surface Editor. The data are collected in Processing⁵, using the oscP5⁶ library, an OpenSoundControl (OSC) implementation for Processing, and analyzed to generate and control the corresponding sound generators and effects with the use of the Minim⁷ audio library.

3 Requirements

For our setup, we will bring a webcam, a support to hold it, and a set of objects. We will only need a table and power. However, a loudspeaker would be appreciated. Setup time is approximately of 30 minutes.

⁴ <http://opensoundcontrol.org/>

⁵ <http://processing.org/>

⁶ <http://www.sojamo.de/libraries/oscP5/>

⁷ <http://code.compartmental.net/tools/minim/>

4 Setup

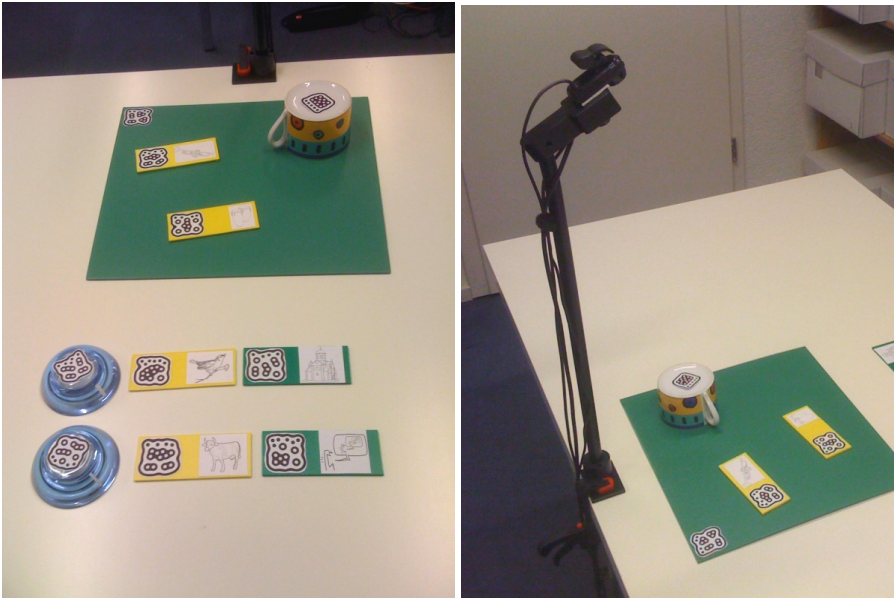


Fig. 2. Right : Example of object-in-object relations (top) and chains (bottom). Left : Webcam fixed on a support to detect objects.

5 Interest in Participating to Kids' INTETAIN

One of our objectives is to make sound design and music accessible to a wider audience, including children starting from 5 or 6 years old. Participating to Kids' INTETAIN could be the occasion for us to have useful feedback about the playability of the scenario we developed and possibly test some variations. The hope is that the research conducted within this project will be useful for music pedagogy in schools as well as entertainment at home. In future work, we could imagine letting users design their own interactive scenarios. For this, we would need to introduce them to the Surface Editor platform first, which could be possible for teachers or older users.

My Presenting Avatar

Laurent Ach¹, Laurent Durieu¹, Benoit Morel¹, Karine Chevreau²,
Hugues de Mazancourt², Bernard Normier², Catherine Pelachaud³,
and André-Marie Pez³

¹ Cantoche, France

² Lingway, France

³ CNRS – LTCI, TelecomParisTech, France

name.lastname@cantoche.com,

name.lastname@lingway.com,

name.lastname@telecom-paristech.com

Abstract. We have developed an application that offers to users the possibility to transmit documents via a virtual agent.

Keywords: Virtual agent, linguistic extraction, nonverbal behavior, animation.

1 Introduction

We have developed an application that offers to users the possibility to transmit documents via a virtual agent. The idea is to use avatars to present document so as to increase its diffusion and ease discussion.

The application makes use of a system that combines semantic analysis, nonverbal behavior generation and 3D animation. The system follows a sequential process. First, it extracts from any given documents the pertinent linguistic information; then it computes the multimodal behaviors the agent should use to communicate it; and finally it plays the animation on a web window. An interactive interface has been developed to allow users to go and modify the generated output. This modification can happen either at the linguistic level (refine the extracted information), the communicative and emotional functions level, or even at the animation level. This approach allows for a personalized and expressive communication schema. In this paper we describe the main components of the system.

Demonstration setup: The system works as follow (see Figure 1): at first the user selects a document to be transmitted by the avatar. To communicate the document expressively the avatar needs to know which of its information to highlight. To this aim, a semantic analysis based on crawling technique determines the structure of the document as well as new, pertinent or even contrasting information [1]. It is based on a predefined template written in the format APML-MPA, a variant of the format APML Affective Presentation Markup Language [2], which has been defined for the MyPresentingAvatar project. The template contains variable parts that are instantiated

when a given document is inputted to the system. The analysis step outputs tags value indicating newness, pertinence, contrasting and emotional information. These values are used to compute the animation of the avatar. Once the templates have been instantiated, they are converted in FML-APML format [3]. This file is then sent to the Greta Behavior Engine to compute the sequence of nonverbal behaviors to convey the communicative intentions and emotions described in the input file. It outputs a temporally order sequence of behaviors described with the BML Behavior Markup Language [4]. Finally with the Cantoche module, these BML tags are converted in animation file in the Living Actor format. The output of this last step is used to generate the presentation video. Through an intuitive interface, users can modify the resulting videos. Modifications can happen at 2 levels: high level, ie at the FML-APML level where users can change the value of a communicative function, and at the low level, ie at the BML one, where users can vary the choice of behaviors and its temporal alignment. After any changes made, the application renders a new video.

Technical description and requirements: The application works on a web browser. Internet connection is necessary to access all the modules.

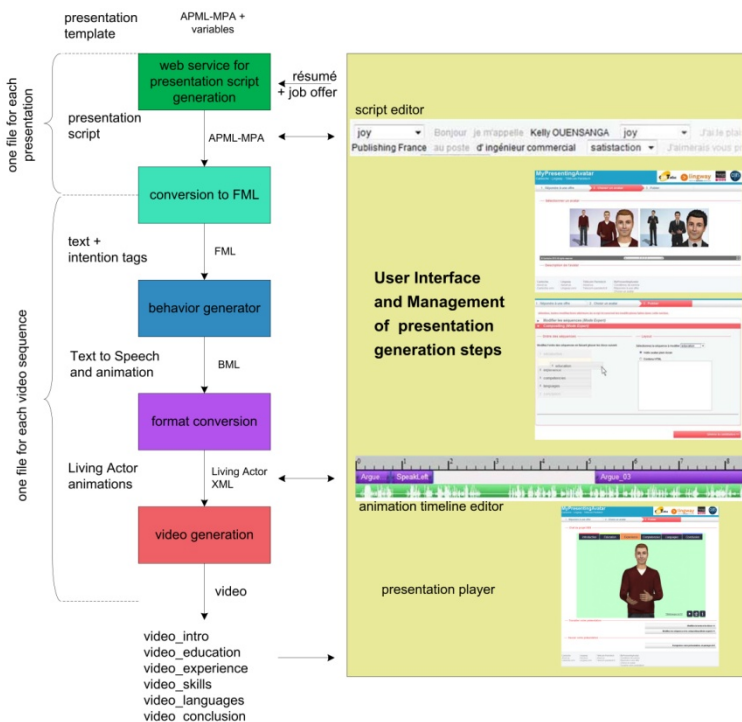


Fig. 1. Overall architecture of MyPresentingAvatar application

Acknowledgments. This research is supported by the ANR Web2.0 MyPresentingAvatar project.

References

1. Coch, J., Chevreau, K.: Interactive Multilingual Generation. In: Gelbukh, A. (ed.) CICLEing 2001. LNCS, vol. 2004, pp. 239–250. Springer, Heidelberg (2001)
2. DeCarolis, B., Pelachaud, C., Poggi, I., Steedman, M.: APMML, a mark-up language for believable behavior generation. In: Prendinger, H., Ishizuka, M. (eds.) *Life-Like Characters, Cognitive Technologies*, pp. 65–86. Springer, Heidelberg (2004)
3. Mancini, M., Pelachaud, C.: The FML-APML language. In: *Proceedings of The First Functional Markup Language Workshop*, pp. 43–47 (2008)
4. Vilhjálmsson, H.H., Cantelmo, N., Cassell, J., E. Chafai, N., Kipp, M., Kopp, S., Mancini, M., Marsella, S.C., Marshall, A.N., Pelachaud, C., Ruttkay, Z., Thórisson, K.R., van Welbergen, H., van der Werf, R.J.: The Behavior Markup Language: Recent Developments and Challenges. In: Pelachaud, C., Martin, J.-C., André, E., Chollet, G., Karpouzis, K., Pelé, D. (eds.) *IVA 2007*. LNCS (LNAI), vol. 4722, pp. 99–111. Springer, Heidelberg (2007)

Interacting with Emotional Virtual Agents

Elisabetta Bevacqua¹, Florian Eyben³, Dirk Heylen⁴, Mark ter Maat⁴,
Sathish Pammi², Catherine Pelachaud¹, Marc Schröder², Björn Schuller³,
Etienne de Sevin⁵, and Martin Wöllmer³

¹ CNRS ParisTech, Paris, France

² DFKI GmbH, Saarbrücken, Germany

³ Technische Universität München, Germany

⁴ Universiteit Twente, The Netherlands

⁵ Université Pierre et Marie Curie, Paris, France

{bevacqua,pelachaud}@telecom-paristech.fr, {eyben,schuller}@tum.de,
{d.k.j.Heylen,maatm}@ewi.utwente.nl,
{Sathish_Chandra.Pammi,marc.schroeder}@dfki.de,
etienne.de-sevin@lip6.fr, woe@mmk.e-technik.tu-muenchen.de

Abstract. Sensitive Artificial Listener (SAL) is a multimodal dialogue system which allows users to interact with virtual agents. Four characters with different emotional traits engage users in emotionally coloured interactions. They not only encourage the users into talking but also try to drag them towards specific emotional states. Despite the agents very limited verbal understanding, they are able to react appropriately to the user's non-verbal behaviour. The demonstrator shows a final version of the fully autonomous SAL system.

Keywords: Embodied Conversational Agents, human-machine interaction.

1 Introduction

The Sensitive Artificial Listener demo shows the final system developed within the European FP7 SEMAINE project. This project aimed at building a Sensitive Artificial Listener (SAL), a multimodal dialogue system which allows users to interact with virtual agents. The system can sustain an emotionally coloured interaction with users for some time reacting appropriately to their non-verbal behaviour. The system can perceive user's verbal and non-verbal behaviours and use this information to plan how to react. Its response is transmitted using an ECA that is capable of communicating through several channels like voice, facial expressions, gestures, head movements and torso shifts. The virtual agent not only encourages the user into talking but also try to pull her/him towards specific emotional states. To achieve such a goal, SAL provides four characters with different emotional traits. Spike is a nasty, angry creature, and his mission in life is to make the user angry too. Poppy is the happy and positive girl and she tries hard to make her interlocutor as happy as she is. Then there's Prudence. She's sensible and pragmatic about everything and she expects the user to be sensible

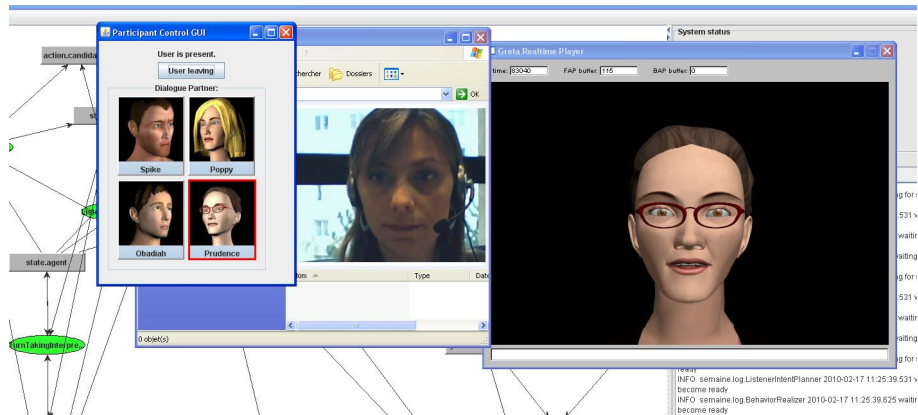


Fig. 1. Demonstration setup

too. Finally, Obadiah is the soul of misery, and he thinks everybody should be miserable, including his interlocutor.

1.1 Demonstration Setup

The demonstration set up is shown in figure 1. The user sits in front of a screen where a SAL character is displayed. The user must wear a microphone for voice analysis and can optionally be recorded by a video camera for facial expression analysis and head movement recognition. The system monitor can be displayed on a second screen; it shows the components and the current data flowing between them. During the interaction the user is the speaker, the virtual agent is mainly the listener who from time to time provides just short sentences to encourages the user into talking. The agent cannot really understand the user's speech, so sometimes the agent's sentences may appear absolutely inconsistent and wild. That is just part of the interaction. It is up to the user to supply all the ideas and the effort to push forward the interaction as much as possible, keeping in mind that there is no point asking the agent questions, or trying to outwit it.

1.2 Technical Description and Requirements

The system uses the SEMAINE API, a distributed multi-platform component integration framework for real-time interactive systems [1]. User's acoustic and visual cues are extracted by analyser modules and then used by the interpreters to derive the system's current best guess regarding the state of the user and the dialogue. This information and the user's acoustic and visual cues are used to generate the agent's behaviour both while speaking and listening.

To run the whole system can be quite cumbersome, so at least a quad core processor machine with 6GB RAM is required. A microphone, web camera and loudspeakers are needed, too. The demo can be easily installed in 10 minutes.

Acknowledgments. This work has been funded by the STREP SEMAINE project IST-211486 (<http://www.semaine-project.eu>).

Reference

1. Schröder, M.: The semaine api: Towards a standards-based framework for building emotion-oriented systems. In: *Advances in Human-Computer Interaction 2010* (2010)

Traditional Shadow Puppet Play – The Virtual Way

Abdullah Zawawi Talib¹, Mohd Azam Osman¹, Kian Lam Tan¹, and Sirot Piman^{1,2}

¹ School of Computer Sciences, Universiti Sains Malaysia,
11800 USM Pulau Pinang, Malaysia

² Department of Business Computing, Faculty of Management Sciences,
Surat Thani Rajabhat University, Surat Thani Province 84100, Thailand
azht@cs.usm.my, azam@cs.usm.my, andrewtankianlam@gmail.com,
sirot.cod07@student.usm.my

Abstract. In this paper, we present a virtual shadow puppet play application that allows real-time play of the puppet and gives the user the impression of being a storyteller or a shadow play puppeteer. Through this tool, everybody can be a puppeteer digitally regardless of the ability to perform the traditional art.

Keywords: Shadow puppet play, virtual puppet, virtual storytelling.

1 Introduction

Creating virtual storytelling application has been greatly simplified with the help of computing and virtual environment technologies, and this has led to a greater interest among researchers and developers to work in this area. The great traditional shadow puppet play called “wayang kulit” in Malaysia and Indonesia [1], is one of the traditional storytelling arts which has given great impacts on the society. The success of each show depends on the storyteller’s or the puppeteer’s ability to attract the audience to watch and enjoy the show until the end of the performance. However, the traditional shadow puppet play is no longer a popular show due to the proliferation of the new media and art, and the lack of interest and the difficulty in mastering the art among the younger generation. Therefore, a virtual tool for shadow puppet play is very much needed to transform the traditional storytelling art into an interactive virtual storytelling environment. Some recent efforts in this area include Lam et al. [2] which introduced a method of modeling that models shadow puppet play using sophisticated computer graphics techniques available in OpenGL and allows interactive play in real-time environment as well as producing realistic animation, and Lam and Talib [3] which proposed a method to improve the interaction of shadow puppet play by applying both the texture mapping and blending techniques.

In this paper, we present a virtual shadow puppet play application based on our previous work (Lam et al. [2], Lam and Talib [3]) that allows real-time play of the puppet and gives the user the impression of being a storyteller or a shadow play puppeteer. Through this tool, everybody can be a puppeteer digitally regardless of their ability to perform the traditional art.

2 Overview of the Virtual Shadow Puppet Play

The main purpose of the Virtual Shadow Puppet Play is to facilitate digital preservation of the traditional art using leading edge technologies. The application enables user to choose the puppet, environment and music. Fig. 1 shows the architecture of the Virtual Shadow Puppet Play which consists of *Puppets* (Puppets and the animation), *Recording* (Recording of the play) and *GUI* (user interface, layout for the music and puppets).

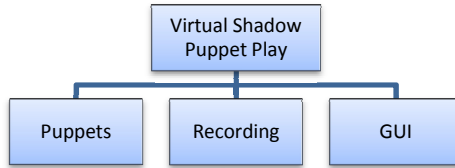


Fig. 1. Architecture of the Virtual Shadow Puppet Play

The Virtual Shadow Puppet Play application provides three different levels of users namely novice, intermediate and expert users. The mode for the novice users is based on a virtual joystick that is used to control the puppets as shown in Fig. 2(a). Four directions of movement are provided namely up, down, left and right. A music button is also provided for the novice users. The mode for intermediate users provides the flexibility in choosing the puppets, and the music, and in controlling the level of brightness of the lighting as shown in Fig. 2(b). The user needs to use the mouse and keyboard to control the puppets instead of the virtual joystick. Besides, this mode also provides several more animations. Expert users can perform with more puppets (with 10 predefined puppets), music (with 7 predefined music) and levels of brightness of the lighting. Expert users can only use the keyboard to control the puppets. Furthermore, in this mode the application also provides a recording function for the show being played. Fig. 3 shows a snapshot of the virtual shadow puppet play.

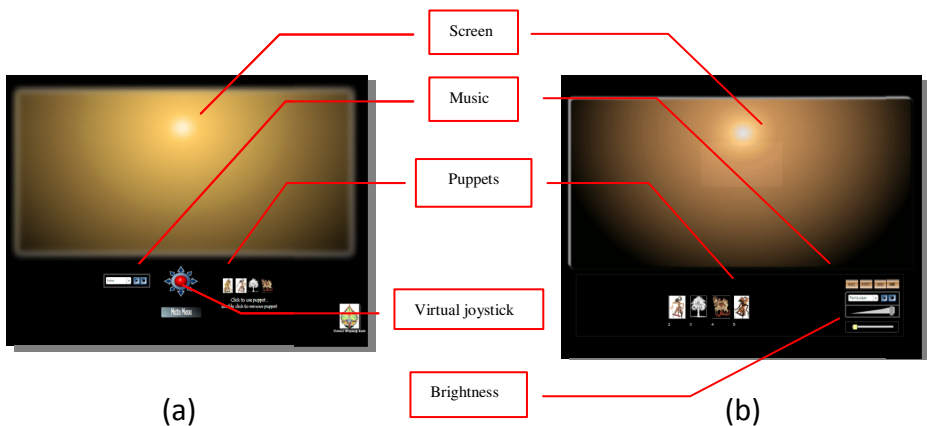


Fig. 2. Interface for (a) Novice Users, (b) Intermediate Users



Fig. 3. The Virtual Shadow Play

3 Conclusion

In our approach to virtual shadow puppet play, the puppets are created as close as possible to the traditional puppets. Our approach has been implemented in a general framework called the Virtual Shadow Puppet Play which allows the users to play the shadow play interactively in real-time anywhere and anytime.

Acknowledgments. The authors would like to thank Ms. Zaidah Juma'at and Ms. Nik Munirah Nik Him for their contributions in the development of this Virtual Shadow Play application.

References

1. Matusky, P.: *Malaysian Shadow Play and Music: Continuity of an Oral Tradition*. Oxford University Press (1993)
2. Tan, K.L., Talib, A.Z., Osman, M.A.: Real-Time Simulation and Interactive Animation of Shadow Play Puppets Using OpenGL. *International Journal of IJCIE* 4, 1–8 (2010)
3. Tan, K.L., Talib, A.Z.: Shadow Image and Special Effects Implementation Techniques for Virtual Shadow Puppet Play. In: *3rd WSEAS International Conference on Visualization, Imaging and Simulation (VIS 2010)*, pp. 80–85 (2010)

The Attentive Machine: Be Different!

Julien Leroy, Nicolas Riche, François Zajega, Matei Mancas, Joelle Tilmanne,
Bernard Gosselin, and Thierry Dutoit

University of Mons (UMONS), Engineering Faculty, IT Research Unit
20, Place du Parc, 7000, Mons

{julien.leroy,nicolas.riche,françois.zajega,matei.mancas,
joelle.tilmanne,bernard.gosselin,thierry.dutoit}@umons.ac.be

Abstract. We will demonstrate an intelligent Machine which is capable to choose within a small group of people (typically 3 people) the one it will interact with. Depending on people behavior, this person may change. The participants can thus compete to be chosen by the Machine. We use the Kinect sensor to capture both classical 2D video and depth map of the participants. Video-projection and audio feedback are provided to the participants.

1 Social Feature Extraction

The main feature which is extracted is the personal space of the participants. Social studies [1] showed that humans have different “ego-spaces”: the public space (from around 3.5 meters in green on Figure 1, left image), the social space where interaction is possible (from around 1.2 meters in blue on Figure 1, left image), the personal space for close interaction (from around 0.45 meters in yellow on Figure 1, left image) and the intimate space (in red on Figure 1, left image). Those measures vary of course depending on cultural and personal contexts. This space is extracted in 3D [2] by using OpenGL and a Microsoft Kinect sensor (Figure 1, right image).

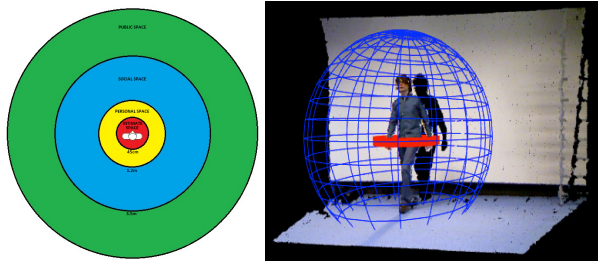


Fig. 1. Right: ego-spaces, Left: 3D extraction of intimate space (red cylinder) and personal space (blue sphere)

The real-time 3D extraction of the inter-personal distances (Figure 2: example with two participants), along with the height of the people and their position and velocity provides a set of both dynamic and static, low-level and mid-level features.

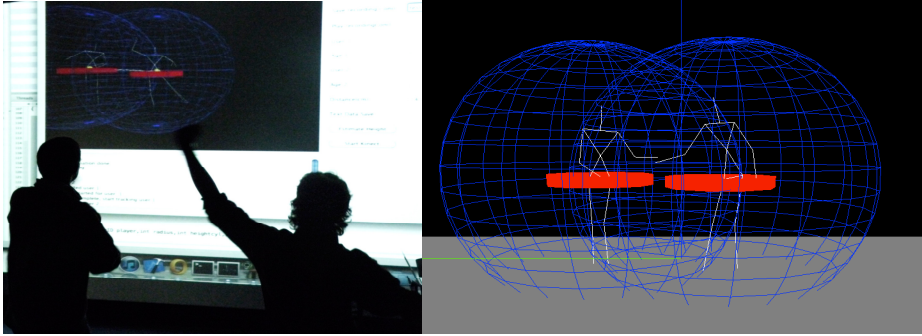


Fig. 2. Two participants' real-time interaction and extraction of social cues

2 Find the Most Interesting People

Those features are then sent to an attention algorithm which will compute feature's relative contrast [3]. As only 3 people are taken into account here, the rarity approach of attention is not relevant, thus we use the global contrast from [3] only. If the features of a person compared with the others are contrasted enough, this person is potentially the most "worthy of interest" for the Machine which will give him the possibility to interact with it. In order to cope with several features in the same time, empirical weights are given and the person with the higher overall weighted contrast is selected.

As a result, while two out of the three people are together, the one who is alone will be selected, if the three people have the same distances, the one in the middle will be selected. If one out of three is sitting on the ground, he will be selected, while if two persons are sitting on the ground, the one standing will be selected. Concerning dynamical features like the 3D speed, they highly attract the system when one person is different (faster, slower), but those dynamic features should be perpetuated by a static feature contrast in order to provide a stable selection of the person.

3 User Feedback

The video feedback which is a wall of HAL's eyes (Figure 3) will change, and all the eyes will focus on the selected person. The idea of several identical objects focusing on a person comes from [4] where people are very excited in trying to be in the mirrors' focus. The selected person will be able to create sounds by using simple gestures. This should push people to try to be different in order to be able to have this interaction. If another person performs an interesting behavior, the Machine will focus

some of the HAL eyes on this second person which is a sign that the interaction could change to the second person. If the first person does not act in order to increase his own interestingness within a given time, he will lose the selection and only the second person will be able to interact and create sounds.

This set-up is of course interesting for its natural user selection in a multi-user scenario, but also because it let us observe the social behavior of the users while interacting with such a machine and the other users.

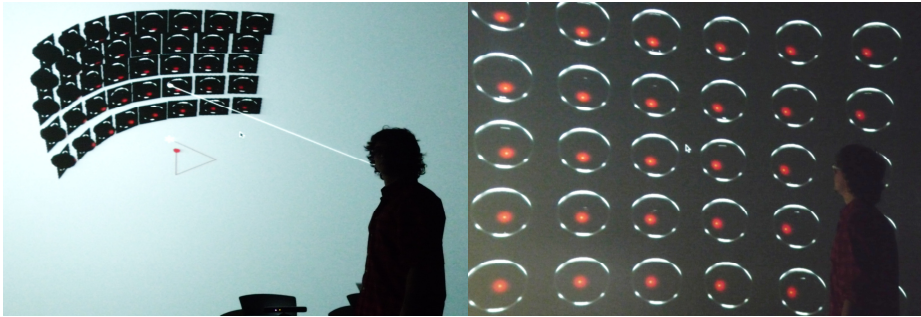


Fig. 3. Left image: 3D structure of the HAL's eyes, Right image: all the eyes focus on the person in front of the screen

4 Requirements

This demo requires a 6x6 meters space and electrical power. If possible, a video projector, a projection screen and a loudspeaker would be appreciated. Half a day should be enough to set up the demo.

Acknowledgments. This demo was created within the NUMEDIART Institute (www.numediart.org) funded by the Walloon region, Belgium.

References

- [1] Hall, E.T.: The Hidden Dimension. Anchor Books (1966)
- [2] Leroy, J., Mancas, M., Gosselin, B.: Personal Space Augmented Reality Tool. In: First joint WIC/IEEE SP Symposium on Information Theory and Signal Processing in the Benelux, Bruxelles, Belgium (2011)
- [3] Mancas, M., Gosselin, B., Macq, B.: A Three-Level Computational Attention Model. In: Proceedings of ICVS Workshop on Computational Attention & Applications (WCAA 2007), Bielefeld, Germany (2007)
- [4] Audience set-up, <http://www.chrisoshea.org/audience>

Towards a Dynamic Approach to the Study of Emotions Expressed by Music

Kim Torres-Eliard^{*}, Carolina Labbé, and Didier Grandjean

Swiss Center for Affective Sciences,
Neuroscience of Emotion and Affective Dynamics laboratory
7, rue des Batoirs, Geneva, Switzerland
{kim.eliard, carolina.labbe, didier.grandjean}@unige.ch

Abstract. The emotions expressed through music have often been investigated by asking listeners to fill questionnaires at the end of a given musical performance or an excerpt; only few studies have been dedicated to the understanding of the *dynamics* of emotions expressed by music in laboratory or in social contexts. Based on a specific model of emotions related to music, the Geneva Emotion Music Scale (GEMS), we tested to what extent such dynamic judgments are reliable and might be a promising avenue to better understand how listeners are able to attribute different kinds of emotions expressed through music and how the social contexts might influence such judgments. The results indicate a high reliability between listeners for different musical excerpts and for different contexts of listening including concerts, i.e. a social context, and laboratory experiments.

Keywords: Emotion, music, dynamic judgment, musical expressiveness.

1 Introduction

1.1 Definition of Emotion

The majority of studies on music and emotion propose to judge musical excerpts in terms of valence and arousal (Vieillard et al., 2008; Chapin, Jantzen, Kelso, Steinberg, & Large, 2010) or in terms of basic emotions (Fritz et al., 2009; Juslin, 2000). However, one might suppose that musical emotions are more complex or subtle and therefore these approaches might not be the best suited to understanding emotions related to music. As noted by Scherer (2004), a major problem in studying music and emotion is the tendency to confuse the terms “emotion” and “feeling”. By adopting a componential approach to emotional processes and by using the component process model (CPM, Scherer, 2001) framework, we define the concept of “emotion” as brief episodes that are important and relevant for the adaptation and well being of individuals. In this context, Scherer (2004) proposed a distinction between “utilitarian emotions” and “aesthetic emotions”. The former being emotions

^{*} Corresponding author.

that humans can experience in everyday life, such as feelings of anger, fear, joy, and sadness. Scherer proposed the term “utilitarian” because this type of emotion has: “[...] *major functions in the adaptation and adjustment of individuals to events that have important consequences for their well being [...]*” (p.241). Regarding aesthetic emotions, the author suggested that there are no appraisals concerning goal relevance or coping potential to the events that elicit them. Therefore, aesthetic emotions might not be governed by vital functions such as bodily needs or current goals. Consequently, the traditional approaches of emotions do not seem appropriate to the study of emotions related to music. In 2008, Zentner, Grandjean and Scherer made a set of experiments enabling them to propose a factorial model of the most relevant emotional terms for the understanding of emotions related to music. These studies gave rise to a nine factorial model of emotions induced by music: the GEMS. These include the dimensions of wonder, transcendence, tenderness, nostalgia, peacefulness, power, joyful activation, tension, and sadness.

1.2 Perception of Emotion vs. Induction of Emotion

The attribution of emotional qualities of music is a complex process allowing humans to represent and explicitly report feelings expressed through music, whereas the induction of emotions is the process of experiencing emotions, i.e. the feelings, as a result of listening to music. In other words, the subjective emotional responses one might have as a result of listening to music (Scherer & Zentner, 2001). This distinction is not always made and this leads to confusion in understanding the mechanisms underlying emotional processes in music which include the attribution of emotion qualities of music *and* the induction of emotion by music (Scherer, 2004). It would be logical to assume that what is expressed by music and what is felt by listeners is the same emotion (Evans & Schubert, 2008). However, several observations may challenge this assumption. Indeed, a piece of music which expresses sadness or melancholy could be listened to in order to provide a feeling of nostalgia in the listener, a state desired and appreciated by the listener (Zentner, Grandjean & Scherer, 2008). Likewise, “agitating music” could have a cathartic effect on the listener and help calm her/him down. Therefore, the correlation between felt emotions and emotions expressed through music is not necessarily positive. The listener does not necessarily feel the same emotions that the music expresses (Evans & Schubert, 2008). Broadly speaking, we can say that recognizing the emotions expressed by music is a process which is based on the perception of acoustical and musical features on which listeners could agree, as evidenced by numerous studies in which this is demonstrated; people have a high reliability on the emotions expressed by music (Hevner, 1935, 1936; Gabrielsson & Juslin, 1996; Fritz et al., 2009; Fabian & Schubert, 2003) whereas felt emotions are more related to subtle subjective and intimate process, as demonstrated in Gabrielsson’s (2001) Strong Experiences in Music (SEM) and music preferences studies (Rentfrow & Gosling, 2003; Rentfrow & McDonald, 2010). And though the GEMS model proposed by Zentner, Grandjean & Scherer (2008) concerns emotions induced by music, it currently represents the most

effective attempt to study the specific emotions related to music, which is why we are using it to investigate emotions expressed by music as well.

1.3 The Dynamic Aspects of Music

An obvious feature of music is that it unfolds over time, as does emotion. In order to effectively apprehend the emotions expressed by music, it is preferable, and probably essential, to base the judgments on continuous measurements. For this purpose, we propose to use an approach called “dynamic judgment”. The works of Emery Schubert (2001; 2004) have been among the first to take into account this characteristic of time and to use continuous measurements. This method allows experimenters to record the judgments of emotions expressed by music in real time and then to follow the changes of perception and attribution over time. As pointed out by Nagel and colleagues (2007), there are different ways of investigating the emotions related to music, such as self report, questionnaires and adjective scales, but all of these approaches are static and therefore unable to demonstrate the complexity of the unfolding of musical emotions. A very old assumption is that music communicates emotion (Gabrielsson & Juslin, 2003). In this context, it’s interesting to consider how people decode this type of “message”, which types of cues they use to build up a representation of emotions expressed by music. Many acoustical parameters and musical features have been identified as being relevant for the understanding of how musicians convey emotions. Among them, those that are most cited are sound intensity (loudness), timbre, intervals, rhythm, articulation, pitch level, accentuation and tempo (for a review, see Gabrielsson & Lindström, 2010). Performers use these different features in order to convey the emotion they want to express (Juslin, 2001). The contexts of the musical performances (e.g. solo versus ensemble) and the contexts of listening might also be important factors playing a role in emotional processes.

2 Method

We conducted a first laboratory exploratory study in order to investigate the dynamics of emotional judgments of the emotions expressed by music through time. More specifically, we wanted to test to what extent, on a given dimension, participants agreed on the emotion expressed by music using the GEMS model.

2.1 Participants

Seventy-one participants (8 men) took part in this experiment, for course credits. The average age is 21.97 (sd = 2.84).

2.2 Materials and Procedure

Based on our musical expertise and our knowledge of the GEMS dimensions, we chose a series of musical excerpts in order to correspond to the 9 dimensions of the

GEMS. We had 36 musical excerpts in total, i.e. 4 musical excerpts per dimension. The mean duration of the excerpts was 2'36'' (range from 2'21'' to 3'18''). The new method of dynamic judgment, using a Flash Interface, allows us to record the dynamic judgments in real-time in a graphical manner. The width of the graph was 1000 pixels (equivalent to 4'16) and the height 300 pixels. Participants had direct visual feedback in the graphic interface of the judgments they were making by moving a cursor up and down as time advances (if necessary the graph-window scrolled). Measurements are made every 250 milliseconds. The x-axis represented time, while the y-axis represented the intensity of the emotion expressed by music (e.g. Nostalgia) through a continuous scale marked by three levels of intensity: low, medium, and high. The main instruction was: “Rate to what extent the music expresses [dimension of interest]”. Before the beginning of the experiment the participants had to achieve a training trial in order to become familiar with the procedure. A description of the GEMS dimensions was provided before the beginning of the judgments. Each participant had to judge 9 excerpts, one excerpt per GEMS dimension.

3 Results

3.1 Reliability of the Emotional Dynamic Judgments

In order to estimate the reliability of the measure we computed the Cronbach alphas for all excerpts and GEMS dimensions across the participants. The Cronbach alphas ranged from 0.84 to 0.98 (Fig. 1).

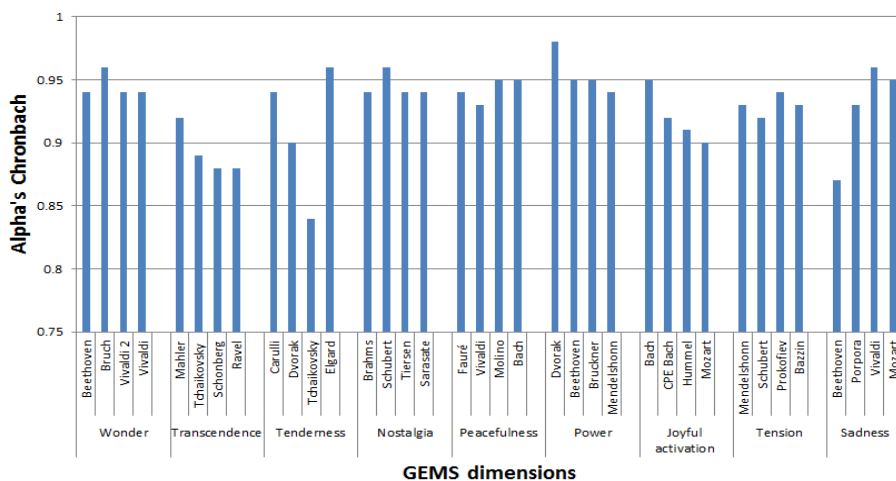


Fig. 1. Cronbach Alphas for the 36 musical excerpts (N=71). The best Cronbach's Alpha lies on the dimension of "Power" for the 4th movement of the New World Symphony by A. Dvorak (.98) and the worst Cronbach's Alpha lies on the dimension of "Tenderness" for the 2nd movement of the Symphony No. 6, *Pathétique*, by P.I. Tchaikovsky (.84).

Figure 2 illustrates the great agreement between participants regarding the emotion expressed by music. The lines represent the participant's judgments and we did a normalization transforming the data into z-scores in order to have all the scores on the same range. The average is represented by the red line.

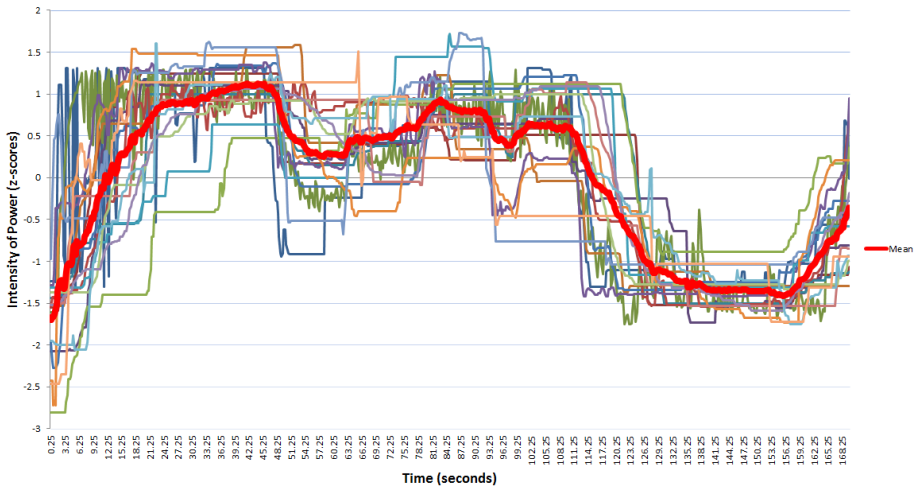


Fig. 2. Individual and averaged z-scores (N=18) for the dynamic emotional judgment of the 4th movement of the New World Symphony by Dvorak judged on the dimension of « Power » (duration: 2'80').

3.2 Dynamic Judgments during Live Performance?

Given the effectiveness of the method, we conducted an experiment with the famous Italian quartet, "Quartetto di Cremona", during a live performance and thus implying a social context at the Saint-Germain Church, in Geneva. We recruited a panel of 14 music lovers and placed them in the audience in front of the musicians. Each participant had a laptop computer and a cursor in order to judge the musical pieces during the concert. Figure 3 presents the Cronbach Alphas for the different movements of the pieces played during the concert. The pieces were the String Quartet n°4 in C major Sz 91 by B. Bartok and the String Quartet n°3 in A major op. 41 by R. Schumann.

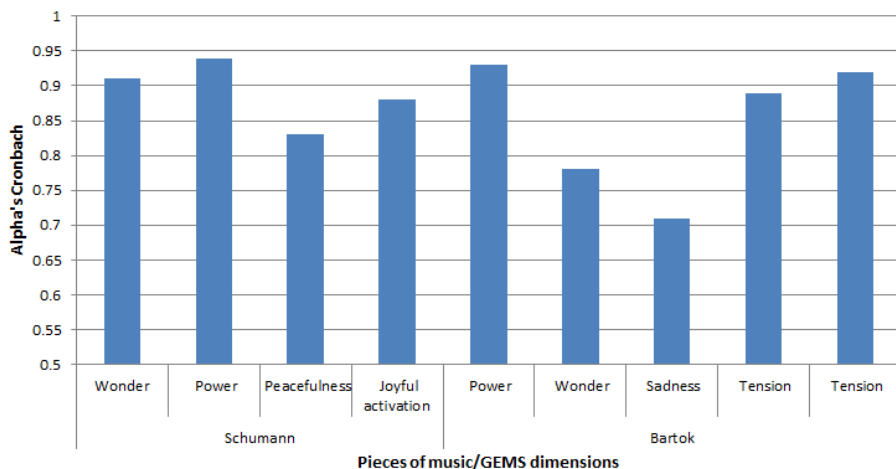


Fig. 3. Cronbach Alphas of the dynamic judgments made during a live performance. The best Cronbach’s Alpha lies on the 2nd movement of the String Quartet n°3 in A major op. 41 by R. Schumann (.94) judged on the dimension of “Power” and the worst Cronbach’s Alpha lies on the 3rd movement of the String Quartet n°4 in C major by B. Bartok judged on the dimension of “Sadness” (.71).

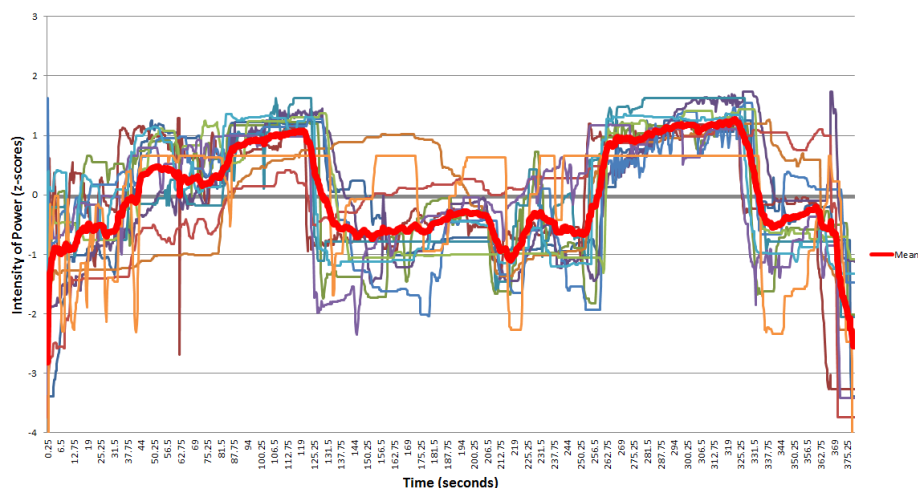


Fig. 4. Z-scores of the dynamic judgment on the dimension of “Power” for the 2nd movement of the String Quartet n°3 in A major op. 41 by R. Schumann (duration: 6’25’’).

4 Perspectives

First, these results show that the method of dynamic judgments is very effective regarding the specific emotional dimensions (i.e. the GEMS dimensions) expressed by music through time, in a laboratory context (and short excerpts) as well as during

live performance implying social interactions (and with longer musical pieces). In a second step, by adopting a Brunswikian (Brunswik, 1956; Grandjean, Baenziger, & Scherer, 2006) perspective, we want to predict the emotional dynamic judgment by the acoustic parameters and musical characteristics in the music score, using a Granger Causality method. In future studies we also plan to systematically investigate the impact of social contexts at different levels including, i) the impact of live musical performances, with the presence of one or several musicians, versus recorded performance, and ii) the context of listening which might be isolated or in a small or big group of people. These kinds of methods will allow researchers to quantify the impact of two different social contexts, i) at the musician level and ii) at the audience level as well as their interactions.

Acknowledgments. The project SIEMPRE acknowledges the financial support of the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET-Open grant number: 250026-2. These studies have also been supported by the Swiss Center for Affective Sciences and the National Center for Swiss National Center of Competence in Research (NCCR) Affective Sciences financed by the Swiss National Science Foundation (SNSF). We would like to thank Professors Bernardino Fantini and Klaus Scherer for helping in the organization of musical event performances.

References

- Brunswik, E.: Perception and The Representative Design of Psychological Experiments, 2nd edn. University of California Press, Berkeley (1956)
- Chapin, H., Jantzen, K., Scott Kelso, J.A., Steinberg, F., Large, E.: Dynamic Emotional and Neural Responses to Music Depend on Performance Expression and Listener Expression. *PLoS ONE* 5(12) (2010)
- Evans, P., Schubert, E.: Relationships between Expressed and Felt Emotions in Music. *Musicae Scientiae* 12, 75–99 (2008)
- Fabian, D., Schubert, E.: Expressive devices and perceived musical character in 34 performances of Variation 7 from Bach's Goldberg Variations (Special issue). *Musicae Scientiae*, 49–71 (2003)
- Fritz, T., Jentschke, S., Gosselin, N., Sammler, D., Peretz, I., Turner, R., Friederici, A.D., Koelsch, S.: Universal Recognition of Three Basic Emotions in Music. *Current Biology* 19, 573–576 (2009)
- Gabrielsson, A.: Emotions in strong experiences with music. In: Juslin, P., Sloboda, J. (eds.) *Music and Emotion: Theory and Research*, pp. 431–429. Oxford University Press, Oxford (2001)
- Gabrielsson, A., Lindström, E.: The influence of musical structure on emotional expression. In: Juslin, P., Sloboda, J. (eds.) *Music and Emotion: Theory and Research*, pp. 223–248. Oxford University Press, Oxford (2001)
- Gabrielsson, A., Juslin, P.: Emotional Expression in Music. In: Davidson, R.J., Goldsmith, H.H., Scherer, K.R. (eds.) *Handbook of Affective Sciences*, pp. 503–534. Oxford University Press, New York (2003)

- Grandjean, D., Baenziger, T., Scherer, K.R.: Intonation as an interface between language and affect. *Progress in Brain Research* 156, 235–247
- Hevner, K.: The affective character of the major and minor modes in music. *American Journal of Psychology* 47, 103–118 (1935)
- Hevner, K.: Experimental studies of the elements of expression in music. *American Journal of Psychology* 48, 246–268 (1936)
- Juslin, P.: Cue utilization in communication of emotion in music performance: relating performance to perception. *Journal of Experimental Psychology: Human Perception and Performance* 26, 1797–1813 (2000)
- Juslin, P.: Communicating emotion in music performance: a review and theoretical framework. In: Juslin, P., Sloboda, J. (eds.) *Music and Emotion: Theory and Research*, pp. 309–337. Oxford University Press, Oxford (2001)
- Nagel, F., Kopiez, R., Grewe, O., Altenmüller, E.: EMuJoy: Software for continuous measurement of perceived emotions in music. *Behavior Research Methods* 39, 283–290 (2007)
- Rentfrow, P.J., Gosling, S.D.: The Do Re Mi's of Everyday Life. The Structure and Personality Correlates of Music Preferences. *Journal of Personality and Social Psychology* 84, 1236–1256 (2003)
- Rentfrow, P.J., McDonald, J.A.: Preference, personality, and emotion. In: Juslin, P., Sloboda, J. (eds.) *Handbook of Music and Emotion: Theory, Research, Applications*. Oxford University Press, Oxford (2010)
- Scherer, K.R.: Appraisal Considered as a Process of Multilevel Sequential Checking. In: Scherer, K.R., Schorr, A., Johnstone, T. (eds.) *Appraisal Processes in Emotion: Theory, Methods, Research*, pp. 92–120. Oxford University Press, New York (2001)
- Scherer, K.R.: Which Emotions Can be Induced by Music? What are the Underlying: Mechanisms? And How Can We Measure Them? *Journal of New Music Research* 33, 239–251 (2004)
- Scherer, K.R., Zentner, M.: Emotion effects of music: Production rules. In: Juslin, P., Sloboda, J. (eds.) *Music and Emotion: Theory and Research*, pp. 361–392. Oxford University Press, Oxford (2001)
- Schubert, E.: Continuous measurement of self-report emotional response to music. In: Juslin, P., Sloboda, J. (eds.) *Music and Emotion: Theory and Research*, pp. 361–392. Oxford University Press, Oxford (2001)
- Schubert, E.: Modeling Perceived Emotion with Continuous Musical Features. *Music Perception* 21, 561–585 (2004)
- Vieillard, S., Peretz, I., Gosselin, N., Khalfa, S., Gagnon, L., Bouchard, B.: Happy, sad, scary and peaceful musical excerpts for research on emotions. *Cognition & Emotion* 22, 720–752 (2008)
- Zentner, M., Grandjean, D., Scherer, K.R.: Emotions Evoked by the Sound of Music: Characterization, Classification, and Measurement. *Emotion* 8, 494–521 (2008)

Mutual Engagement in Social Music Making

Nick Bryan-Kinns

Interactional Sound and Music Group, Centre for Digital Music
Queen Mary University of London, Mile End, London. E1 4NS. UK
nick.bryan-kinns@eeecs.qmul.ac.uk

Abstract. Mutual engagement occurs when people creatively spark together. In this paper we suggest that mutual engagement is key to creating new forms of multi-user social music systems which will capture the public's heart and imagination. We propose a number of design features which support mutual engagement, and a set of techniques for evaluating mutual engagement by examining inter-person interaction. We suggest how these techniques could be used in empirical studies, and how they might be used to inform artistic practice to design and evaluate new forms of collaborative music making.

Keywords: Design, Evaluation, Mutual Engagement, Multi-User, Social Interaction, Human Communication.

1 Introduction

Is music dead? Whilst figures in the commercial music industry bemoan the loss of sense of purpose in contemporary music and the role of the Internet in sidelining music as a political force [9], we contend that new technologies hold the key to reinvigorating music's social role. We accept that music has dropped away from being a driving force behind communication technology innovation, losing out to text based networks such as *Twitter* and *Facebook*, and argue that what is needed are innovative and engaging ways for people to make, share, enjoy, and experience music within the context of the modern, connected, real-time world we live in. To this end, we are exploring ways to understand the role of audio in multi-person interactions from interactive art, music, and performance through to workplace collaborations. We believe that the key to success in this venture will be designing new multi-person audio experiences which are informed by understandings of human communication, and which exploit the unique opportunities offered by new technologies rather than mimicking existing ways of interacting. In short, music *is* dead, long live music!

In this paper we outline our approach to understanding *mutual engagement* in multi-person music making. We first describe a set of design features which we believe will increase mutual engagement in multi-user systems and social music experiences. We then present a set of techniques for identifying mutual engagement in music making by examining the minutiae of the user interface mediated interaction

between participants. Finally we present a few illustrative descriptions of multi-person mutually engaging systems we have designed, built, and evaluated. We believe that our approach is suitable for music making and using music to support work.

2 Mutual Engagement

Sadly, in interface design, sound has been limited to providing alert cues, or ongoing background awareness in collaborative work situations cf. [1]. Some research has examined audio as a medium for collaborative work [13], and even in the field of New Interfaces for Musical Expression, the evaluation of audio centric interfaces tends to focus on parameter manipulation tasks cf. [19] rather than examining the nature of collaborative audio creativity cf. [2] and how these interfaces could support a creative and engaging experience [8]. To address this, we explore the concept of *mutual engagement* – points at which people creatively spark together and enter a state of group flow [5] – and examine how different user interfaces features affect people’s levels of mutual engagement. In this way we hope to identify and develop more socially engaging musical experiences which will return music to the core of human experience. The key distinguishing characteristic of mutual engagement is: “it involves engagement with both the products of an activity and with the others who are contributing to those products” [ibid]. Points of mutual engagement are indicated by the attunement of participants’ actions to each other (e.g. mirroring each others’ contributions, or building on each others’ compositions), and focused interaction with the product at hand (e.g. careful editing and manipulation of parameters for expressive effect). These points of mutual engagement are essentially points of group *flow* cf. [7], similar in context to Sawyer’s ethnographic descriptions of group flow [16], but we concentrate on analysing the minutiae of the group interaction mediated through the interface in order to inform design of engaging collaborative systems. In order to achieve this, Bryan-Kinns and Hamilton [5] drew on models of human communication e.g. [6] and CSCW research e.g. [10] to develop a set of design criteria and evaluation measures for mutual engagement which are outlined in the next section.

2.1 Design Features

We have identified a number of design features [5][3] which we believe are important to supporting mutually engaging interaction:

- **Mutual awareness of action** - highlighting new contributions to the joint product, and indicating authorship has been shown to increase mutual engagement.
- **Annotation** - being able to communicate in and around a shared product, and being able to refer to parts of the product helps participants engage with each other.

- **Shared and consistent representations** - participants find it easier to understand the state of the joint product, and the effect of their own and others' contributions when the representations are shared and consistent.
- **Mutual modifiability** - editing each others' contributions increases engagement with each others' product, and the activity becomes more egalitarian.
- **Spatial organization** - allowing participants to layout elements of the joint product in space increases mutual engagement by supporting fluid and improvised privacy and grouping.

The design question then becomes: *How are these features used to inform design, especially in audio-only interfaces.* Interestingly, in recent studies we found that implementing all the design features could actually reduce mutual engagement, possibly due to cognitive overload.

2.2 Evaluation Techniques

Through our studies we have iteratively refined a set of measures of mutual engagement based on analysis of patterns of participants' interaction, and a robust Mutual Engagement Questionnaire (MEQ) which can be used to compare different interfaces. These measures and questionnaires are suitably generic to be usable across different social music interfaces. Our measures of mutual engagement include:

- Number of **contributions, edits, and deletions** - excessive numbers of contributions in the music domain indicates low levels of mutual engagement.
- Amount of **co-editing** (i.e. editing each others' contributions) - increased constructive co-editing indicates increased mutual engagement.
- **Spatial colocation** - working together in the same part of a virtual space indicates mutual engagement.
- Evidence of **convergence of musical ideas** (i.e. alignment and repetition of musical motifs) indicates mutual engagement.

Measures of musical convergence between participants are problematic. We are currently investigating techniques to reliably identify convergence of musical ideas in social music making including using Music Information Retrieval techniques such as edit-distance and sub-sequence sampling. We believe that although these techniques focus on monophonic sources [11], they could have significant utility in understanding social music interaction in general.

In contrast, our Mutual Engagement Questionnaire (MEQ) is used to compare two or more user interfaces. In this approach participants use a number of user interfaces and then complete the MEQ of twelve questions from four categories (not conveyed to participants): Satisfaction with the product, Feelings of enjoyment or flow cf. [7], Sense of collaboration, and Usability. The comparative nature of the MEQ forces participants into making explicit distinctions between interfaces. Our MEQ would be suitable for comparing different social music interfaces and experiences, and would provide a good indicator of participants' preferences.

3 Explorations of Multi-person Musical Experiences

We have been exploring mutual engagement in social music through a series of studies from interactive art through to group music composition. The main vehicle for this work has been a series of studies of distributed music making applications Daisyphone [3] and Daisyfield (forthcoming). We follow discussion of these systems with a brief description of other social music systems we have been exploring.

3.1 Daisyphone and Daisyfield

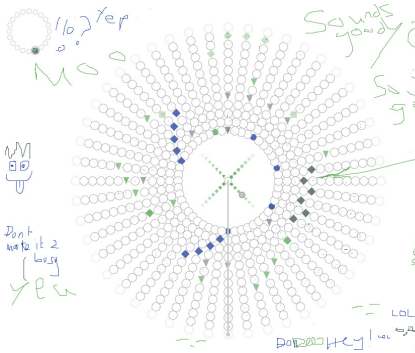


Fig. 1. Daisyphone in use

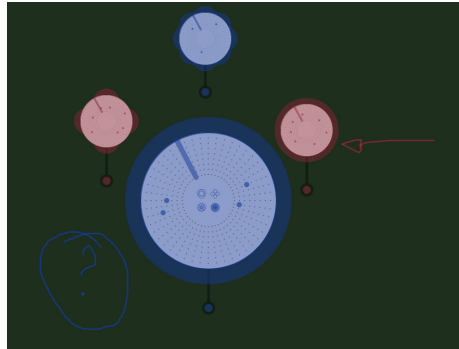


Fig. 2. Daisyfield in use

Daisyphone (figure 1) and Daisyfield (figure 2) share a common underlying distributed, client-server architecture, which allows multiple participants to co-create short loops of music (1 minute) without being in the same physical space. At the heart of the user experience are loops which are shared between participants and can be co-edited at will (there are no ownership controls). Loops are represented as Daisies with the notes of the loop laid out in a circular fashion. Daisyphone provides one shared loop, Daisyfield supports up to 12 shared loops, each represented as a separate Daisy arranged across the shared space. Indeed, Daisyfield is a development of Daisyphone which draws on our studies of naturalistic music improvisation [12] and composition [15]. In keeping with our design features, mutual awareness of action is supported by each participant having a unique colour in the interface, annotation is supported through free graphic drawing, and the whole interface is shared consistently between participants. Lowest notes are on the outsides of Daisies, and highest towards the centre (Persian scale of electro-acoustic sounds). The four shapes in the centre of each Daisy allow for selection of different sounds.

Using Daisyphone and Daisyfield we have explored the role of mutual awareness, persistence of musical contributions, graphical annotations, localization of sounds, and spatial arrangements. We have also used them to revise and validate our MEQ.

Both systems have a set of features which make them particularly amenable to automated analysis of musical convergence: notes are played at a constant speed, and each note has the same duration. However, the interfaces allow several notes to be played at the same time which increases the complexity of applying pattern matching techniques. We shall be exploring these issues in ongoing research.

In studies we confirmed that providing shared annotation mechanisms and mutual awareness of identity of others significantly increased mutually engagement between participants as measured through our MEQ and correlated with our measures of mutual engagement [5]. This was manifested as more focused interaction between participants indicated by fewer contributions and edits of notes [ibid]. We also found that whilst the shared music making can happen without additional communication channels, when present, shared annotation mechanisms are used both for task management and social interaction, and there are more complimentary efforts in composition when shared annotations are provided [3][5]. Also, more annotations about quality of composition seems to indicate more mutual engagement.

Whilst Daisyphone and Daisyfield are unashamedly Graphical User Interfaces, we have also been exploring designing for mutual engagement in musical user interfaces which are not visually oriented. For instance, Stowell et al. [18] showed how rigorous HCI approaches could be used to evaluate people's engagement with and through musical interfaces – a novel beat-boxing synthesizer was evaluated using Discourse Analysis to explore engagement with the technology, and a live beat-tracking system was evaluated using a version of the classic Turing Test to evaluate participants' engagement with each other and the beat-tracking system. In contrast, we focused on the mutual awareness of actions design feature in designing the *Serendiptichord* [14] - a wearable musical instrument whose design considers exploration of musical space, and the engagement of performer with instrument and audience. Similarly, mutual awareness of action and spatial organization were key factors in the design of *uPoi* [17] - a guerilla multi-person interactive audiovisual experience intended to entice and engage participants with each other in unexpected and unusual situations which we observed in use at music festivals. Focusing on designing for mutual awareness of action and shared and consistent representations, *Sensory Threads* [4], is a multi-person mobile experience in which participants sense imperceptible phenomena around them through a responsive real-time soundscape. Our design features informed the design of the soundscape, ensuring that it conveyed the identity of participants clearly, and that there was clear auditory and spatial separation between sounds in the emergent virtual space. In all these non-visual designs we used our mutual engagement design features to drive the design to create a more engaging experience. From the feedback and observations we found that people did engage with each other, but we need to refine our evaluation techniques to work outside the laboratory. In contrast, our research on cross-modal collaborative work has explored the role of audio in workplace group interaction [13], focusing on task efficiency, but using the design features of mutual awareness to create effective collaboration environments. It would be interesting to explore how these systems could be further developed to support social music making through cross-modal interaction.

4 Summary

In this paper we presented our view on mutual engagement as the key to successful multi-person music making. We presented a set of design features and methods of evaluation which we feel could help inform the understanding of social behavior in music, and help to design more mutually engaging musical experiences. The work presented here is a small step in that direction. Future work will test our ideas on mutual engagement in different domains, and explore social music making using a range of modalities through cross-modal interaction.

Acknowledgement. Research supported by EPSRC grants EP/H042865/1, EP/E045235/1, GR/S81414/01.

References

- [1] Ackerman, M.S., Starr, B., Hindus, D., Mainwaring, S.D.: Hanging on the ‘wire’: a field study of an audio-only media space. *ACM Transactions on Computer-Human Interaction* 4(1), 39–66 (1997)
- [2] Barbosa, A.: Displaced soundscapes: a survey of network systems for music and sonic art creation. *Leonardo Music Journal* 13, 53–59 (2003)
- [3] Bryan-Kinns, N.: Daisiphone: The Design and Impact of a Novel Environment for Remote Group Music Improvisation. In: *Proceedings of DIS 2004, Boston, USA*, pp. 135–144 (2004)
- [4] Bryan-Kinns, N., Airantzis, D., Angus, A., Fencott, R., Lane, G., Lesage, F., Marshall, J., Martin, K., Roussos, G., Taylor, J., Warren, L., Woods, O.: Sensory Threads: Perceiving the Imperceptible. In: *Proceedings of 5th International Conference on Intelligent Environments, IE 2009 (2009)*
- [5] Bryan-Kinns, N., Hamilton, F.: Identifying Mutual Engagement. *Behaviour & Information Technology* (2009), doi: 10.1080/01449290903377103
- [6] Clark, H.H., Brennan, S.E.: Grounding in communication. In: Resnick, L.B., Levine, J., Behrend, S.D. (eds.) *Perspectives on Socially Shared Cognition*, pp. 127–149. American Psychological Association, Washington, DC (1991)
- [7] Csikszentmihalyi, M.: *Flow: the psychology of optimal experience*. Harper Collins, New York (1991)
- [8] Dobrian, C., Koppelman, D.: The ‘E’ in NIME: musical expression with new computer interfaces. In: *Proceedings of New Interfaces for Musical Expression (NIME)*, pp. 277–282. IRCAM, Centre Pompidou, Paris, France (2006)
- [9] Graff, G.: Bob Geldof Pleads For Rock’s Future At SXSW Keynote, *The Hollywood Reporter* (March 17, 2011)
- [10] Gutwin, C., Greenberg, S.: A descriptive framework of workspace awareness for real-time groupware. *Computer Supported Cooperative Work* 11, 411–446 (2002)
- [11] Hanna, P., Robine, M., Ferraro, P., Allali, J.: Improvements of Alignment Algorithms for Polyphonic Music Retrieval. In: *Computer Music Modeling and Retrieval 2008*, Denmark (2008)
- [12] Healey, P.G.T., Leach, J., Bryan-Kinns, N.: Inter-Play: Understanding Group Music Improvisation as a Form of Everyday Interaction. In: *Proceedings of Less is More — Simple Computing in an Age of Complexity*, Microsoft Research Cambridge (2005)

- [13] Metatla, O., Bryan-Kinns, N., Stockman, T.: Constructing Relational Diagrams in Audio: The Multiple Perspective Hierarchical Approach. In: Proceedings of ASSETS 2008, Halifax, Canada (2008)
- [14] Murray-Browne, T., Mainstone, D., Bryan-Kinns, N., Plumbley, M.D.: The Serendiptichord: A Wearable Instrument for Contemporary Dance Performance. To appear in Proceedings of the 128th Convention of the Audio Engineering Society, London, UK (2010)
- [15] Nabavian, S., Bryan-Kinns, N.: Analysing Group Creativity: A Distributed Cognitive Study of Joint Music Composition. In: Proceedings of Cognitive Science, pp. 1856–1861 (2006)
- [16] Sawyer, K.R.: Group Creativity: Music, Theater, Collaboration. LEA, New Jersey (2003)
- [17] Sheridan, J.G., Bryan-Kinns, N.: Designing for Performative Tangible Interaction. *International Journal of Arts and Technology*. Special Issue on Tangible and Embedded Interaction 1(3-4), 288–308 (2008)
- [18] Stowell, D., Robertson, A., Bryan-Kinns, N., Plumbley, M.D.: Evaluation of live human-computer music-making: quantitative and qualitative approaches. *International Journal of Human-Computer Studies* 67, 960–975 (2009)
- [19] Wanderley, M.M., Orio, N.: Evaluation of input devices for musical expression: Borrowing tools from HCI. *Computer Music Journal* 26(3), 62–76 (2002)

Measuring Ensemble Synchrony through Violin Performance Parameters: A Preliminary Progress Report

Panagiotis Papiotis, Marco Marchini, Esteban Maestre, and Alfonso Perez

Music Technology Group, Universitat Pompeu Fabra
Tanger 122-140, 08018 Barcelona, Spain
{panos.papiotis, marco.marchini,
esteban.maestre, alfonso.perez}@upf.edu

Abstract. In this article we present our ongoing work on expressive performance analysis for violin and string ensembles, in terms of synchronization in intonation, timing, dynamics and articulation. Our current research objectives are outlined, along with an overview for the methods used to achieve them; finally, focusing on the case of intonation synchronization in violin duets, some preliminary results and conclusions based on experimental recordings are discussed.

Keywords: violin, expressive performance, intonation, ensemble performance, bowing gestures, motion capture.

1 Introduction

Expressive music performance analysis is a wide and interdisciplinary research field, combining elements from signal processing, computational musicology, pattern recognition, and artificial intelligence among others. A thorough presentation of the state of the art in the field can be found in [1].

The problem of expressive performance analysis can be stated as follows: given a score S_j and a recorded performance of that score E_j , performance analysis is carried out by measuring the deviations between S_j and E_j in terms of *timing* (onset times and note durations), *dynamics*, *articulation* (vibrato, tremolo, overall timbre etc.) and, depending on the instrument with which the piece is performed, *intonation*. These deviations can then be seen as the performer's interpretation of the piece, a combination of personal artistic choices as well as implicit musical knowledge.

In the case of ensemble music, where multiple musicians are performing their respective parts simultaneously, performance analysis can be performed on two different levels: on the *intrapersonal* level, thus studying only the relationship between each musician's individual performance and the score it is based on, and on the *interpersonal* level [2], where the relationship between the musicians' performances is also studied; specifically, the influence of one musician's specific performance parameters (either performed live or pre-recorded) on the specific performance parameters of each other member of the ensemble. Naturally, this case is

significantly less studied that the case where the focus is placed on a single musician, although in the past decade a number of researchers have investigated the subject from various viewpoints [3], and utilizing different performance parameters [4].

1.1 Objectives

In studying the influence of one musician over the other (and vice versa) in an ensemble situation, there are three main objectives to be regarded:

- *Detecting* the source and target of the influence mechanism,
- *Analyzing* the nature of the influence mechanism, and
- *Simulating* the influence mechanism by means of a computational model.

It is important to note that the first of these objectives, namely *detecting* the source and target of the influence mechanism, begins with the hypothesis that such an influence actually exists. Therefore, rejecting/validating this hypothesis is the first and most important step of the procedure.

2 Data Acquisition and Pre-processing

In this chapter, we provide an overview of the methods followed for the acquisition of the audio data and instrumental gestures, as well as the computational tools used to perform an alignment between the score and the performance of each musician.

2.1 Recordings Procedure

Several recording sessions were carried out involving motion capture using an EMF commercial device (as detailed in [5]), piezoelectric pick-up microphones fitted on the bridge of each violin, and ambient microphones capturing the overall produced sound. For the moment, these recordings focus on the scenario of two interacting violinists in different experimental set-ups.

The first round of experiments featured two professional violinists recording excerpts from W. A. Mozart's *12 Duets* (KV 487), J. S. Bach's *Concerto for Two Violins* (BWV 1043) and L. Berio's *Duetti per due violini*. The experimental set-ups consisted of:

- a *solo* set-up, where each musician performed their part alone
- a *normal* set-up, where they performed their respective parts together as in a normal duet situation, and finally
- a *switched score* set-up where they performed together, however with the scores switched as in respect to the normal set-up.

A second round of experiments was carried out shortly afterwards involving two amateur violinists, and using the same audio and motion capturing techniques. The subjects were tasked with playing together for the first time pieces of very basic difficulty which they were not familiar with. The experimental set-ups for the second

round consisted of the same set-ups as the previous experiment; however, the pieces were (as mentioned above) simpler, including performing a popular and well-known melody (*Greensleeves*) in unison, and performing one of the duets by L. Berio at a steady, slow tempo.

2.2 Score-Performance Alignment

In order to measure the deviations between the recorded performances and their respective scores, it was necessary to perform a score-performance alignment in the temporal domain.

However, aligning an audio recording to a score is, given the current state of the art in signal processing, a difficult task. This is even more the case for a continuous-excitation instrument such as the violin where the note attacks are varied and smooth; for this reason, we exploited the recorded motion capture data and its derived performance descriptors (bow position, bow transversal velocity, bow pressing force etc.) using the method described in [5]. In this method, bow direction changes as well as more subtle measurements such as an estimation of the applied bow force provide the most probable candidates for note changes, combined with information extracted from the audio (such as the fundamental frequency curve and the root mean square energy of the recorded sound). These features are given as input to an implementation of the Viterbi algorithm, which calculates the temporal alignment between the score and the recorded performance.

3 Towards an Analysis of Intonation Adjustments

In this first phase of this work, our main direction was to study the mechanism through which the violinists adjust to each other's pitch; for the experimental set-ups detailed above, to determine if violin 1 adjusts to violin 2 or the opposite, and how (timing of intonation changes, critical band of acceptable pitch difference, et cetera).

3.1 Temporal Matching of Different Experimental Set-Ups

Since the recordings in the experimental set-ups were performed without a metronome, it was necessary to time-warp the performances in order to compare pitch deviations between violinists 1 and 2; in the *solo* recordings, for example, this comparison is impossible since the two recordings of violinists 1 and 2 were not temporally synchronized.

This was achieved by applying a note-by-note temporal warping algorithm based on resampling the signal between note onsets and restoring its original pitch using an implementation of the phase vocoder algorithm. Besides producing an accurate and non-destructive temporal alignment, this approach can be also very useful in performing user evaluation tests, where subjects can hear the *normal* duet recordings and the *solo* aligned-duet recordings and rate the quality of the duet's intonation; thus investigating whether intonation adjustments alone (i.e. without temporal mismatch) can provide enough information to discriminate between solo and duet recordings.

3.2 Data Post-processing and Score Representation

All data analysis takes part in MATLAB. Fundamental frequency (pitch) estimation is performed with the YIN [6] algorithm, while the score is imported as a MusicXML file, and converted to a time series. All time series are resampled to 1KHz, to facilitate comparisons between time series; finally, pitch estimation errors (such as octave errors) are removed with the use of pitch guides.

Figure 1 presents an excerpt of the J.S. Bach concerto recorded in the first experimental set-up.

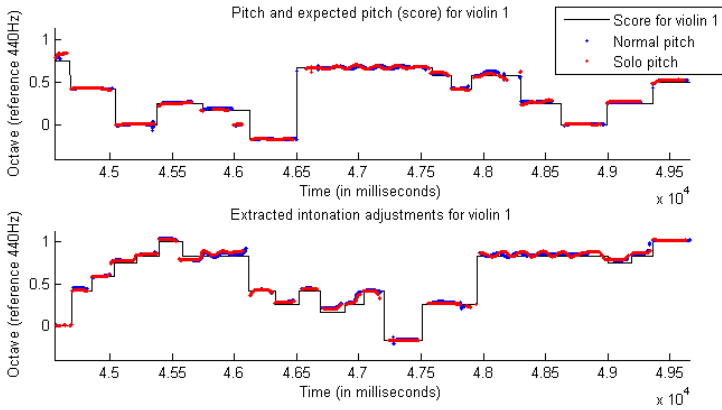


Fig. 1. Expected pitch (score) [black], recorded pitch in the *normal* set-up [blue], and recorded pitch in the *solo* set-up [red], for an excerpt of J.S. Bach’s ‘Concerto for two violins’

4 Preliminary Results and Discussion

Some preliminary results and their implications are already visible from the data processing step. Namely, as it can be observed in Figure 1, the professional musicians employed for the first round of experiments were able to reproduce the same intonation (seen in fig.1 as pitch curves) with impressive accuracy; the mean difference (in pitch cents) between the *solo* and *normal* recordings, in regard to violinist 1 and 2, can be seen in table 1:

Table 1. Mean difference and standard deviation (in pitch cents) between *solo* and *normal*

Piece	Mean difference	Standard deviation
Bach, violinist 1 (professional)	-4.976 cents	25.4 cents
Bach, violinist 2 (professional)	3.149 cents	32.7 cents
Berio n.17, violinist 1 (professional)	-8.165 cents	31.1 cents
Berio n.17, violinist 2 (professional)	-1.404 cents	13.7 cents
Berio n.24, violinist 1 (amateur)	0.610 cents	15.0 cents
Berio n.24, violinist 2 (amateur)	-2.813 cents	12.5 cents
Greensleeves, violinist 1 (amateur)	-4.659 cents	18.8 cents
Greensleeves, violinist 2 (amateur)	-1.340 cents	14.35 cents

If we consider Rossing's [7] estimation that the JND (just-noticeable-difference) for pitch is ~ 5 cents, it is immediately evident that whichever intonation adjustments take place, they are in response to very subtle deviations from the expected pitch. Moreover, these adjustments are also partially obscured by several performance aspects, such as vibratos, glissandi between notes, and of course the shortcomings of pitch estimation algorithms.

To demonstrate this concept more clearly, Figure 2 shows a scatter plot between the pitch deviations of violinists 1 and 2, for the *normal* and *solo* set-ups.

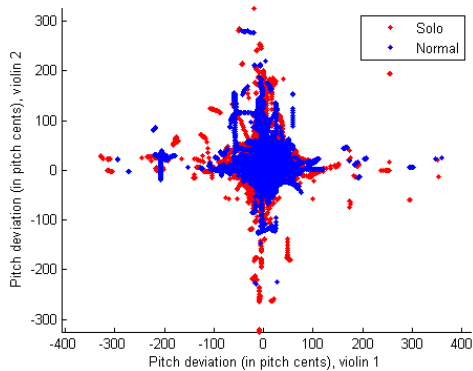


Fig. 2. Pitch deviation for violinist 1 versus pitch deviation for violinist 2, for an excerpt of J.S. Bach's 'Concerto for two violins'

As it can be seen in the above figure, there is no visible correlation between the pitch deviations of the two violinists, and there is little (if any) visible difference between the *solo* and *normal* set-ups. The implications of such an observation lead us to the conclusion that, whichever synchronization phenomenon occurs between the intonation adjustments in a duet, it is not consistent throughout the piece, and cannot be viewed as a process which is either invariant to time or cyclostationary.

To this end, our next step has been to employ several dependence measures between the two time series, such as Mutual Information, Granger Causality and non-linear directional coupling detection algorithms derived from computational neuroscience [8].

The latter coupling measure used (referred to as the *L*-measure) is capable of providing also the *directionality* of the interdependence, in our case the strength with which violinist 1 influences violinist 2 and vice versa; figure 3 displays some preliminary results based on this measure, where we were able to observe an increase of average coupling strength in the *normal* set-up compared to the *solo* set-up. The most evident separation between the two set-ups can be found in the amateur musicians' recordings - since the second violinist was less adept in *prima vista*, the first violinist had to adapt quite a lot in order to balance the performance.

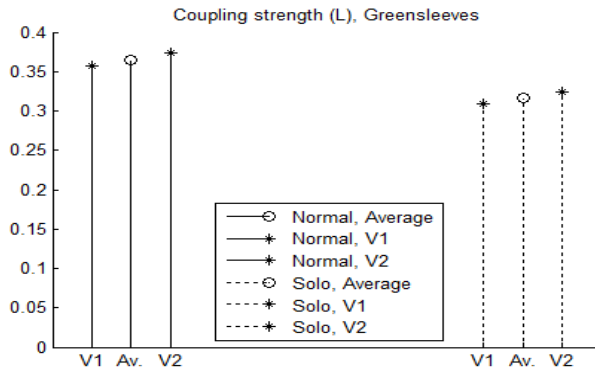


Fig. 3. Coupling strength between the pitch deviations of violinists 1 and 2, for the *normal* and *solo* recordings of one of Greensleeves (amateur musicians). The x-axis shows three different coupling strengths for every recording: V1 \Rightarrow V2, V2 \Rightarrow V1 and average coupling strength.

However, before we can extract clear conclusions from the use of these measures, a refinement is required in the initial features used for this approach; such an effort is currently underway.

Acknowledgements. The work presented on this document has been partially supported by the EU-FP7 ICT SIEMPRE project.

References

1. Widmer, G., Goebel, W.: Computational Models of Expressive Music Performance: The State of the Art. *Journal of New Music Research* 33(3), 203–216
2. Keller, P.E.: Joint action in music performance. *Emerging Communication* 10, 205
3. Moore, G.P., Chen, J.: Timings and interactions of skilled musicians. *Biological Cybernetics* 103(5), 401–414 (2007)
4. Kalin, G.: Formant Frequency Adjustment In Barbershop Quartet Singing. Doctoral dissertation, KTH, Department of Speech, Music and Hearing (2005)
5. Maestre, E.: Modeling instrumental gestures: an analysis/synthesis framework for violin bowing. Doctoral dissertation, Univ. Pompeu Fabra, Barcelona, Spain (2009)
6. De Cheveigné, A., Kawahara, H.: YIN, a fundamental frequency estimator for speech and music. *Journal of the Acoustical Society of America* 111(4), 1917–1930 (2002)
7. Rossing, T.D.: *The Science of Sound*, 2nd edn. Addison-Wesley (1990)
8. Chicharro, D., Andrzejak, R.G.: Reliable detection of directional couplings using rank statistics. *Physical Review E* 80, 026217 (2009)

Communication in Orchestra Playing as Measured with Granger Causality

Alessandro D'ausilio¹, Leonardo Badino¹, Yi Li², Sera Tokay³, Laila Craighero⁴,
Rosario Canto^{1,4}, Yiannis Aloimonos², and Luciano Fadiga^{1,4}

¹ RBCS, Italian Institute of Technology, Genova, Italy

² Department of Computer Science, University of Maryland, USA

³ Şişli Symphony Orchestra, Istanbul, Turkey

⁴ DSBTA - University of Ferrara, Ferrara, Italy

Abstract. Coordinated action between music orchestra performance, driven by a conductor, is a remarkable instance of interaction/communication. However, a rigorous testing of inter-individual coordination in an ecological scenario poses a series of technical problems. Here we recorded violinists' and conductor's movements kinematics in an ecological interactive scenario. We searched for directed influences between conductor and musicians and among musicians by using the Granger Causality method. Our results quantitatively show the dynamic pattern of communication among conductors and musicians. Interestingly, we found evidence that the aesthetic appreciation of music orchestras' performance is based on the concurrent increase of conductor-to-musicians causal influence and reduction of musician-to-musician information flow.

Keywords: communication, action coordination, joint action, neuroscience of music, music performance, movement kinematics, Granger causality, neuroaesthetic.

1 Introduction

Coordinated action is a form of interaction and it has been suggested that it may rely on sensorimotor communication among participants [1]. In fact, action coordination requires the continuous exchange of information among several individuals and the ability to read other's motor intention. Therefore, coordinated action is the accurate negotiation of our own motor output according to sensorimotor messages sent by other participants in the interaction [2-7].

In this context, music orchestras are a particularly interesting instance of sensorimotor coordination between several players and a conductor. More interestingly, ensemble music performance is also a remarkable instance of social interaction aimed at a common aesthetic goal. In fact, musicians train for years in order to acquire a non-linguistic and successful sensory-motor communication. Therefore, orchestra's collective behaviors might be a powerful model of inter-individual non-linguistic communication among highly skilled individuals.

2 Brief Methods

Here we studied music orchestras (a violins section and a conductor) in an ecological rehearsal scenario thus excerpting no particular interference on participant's behavior. We recorded violinists' bows and conductor's baton kinematics via an unobtrusive passive infrared optical system. We searched for directed influences, and modulation thereof, among actions (acceleration profiles) of the participants without imposing any artificial constraint. Eight violinists played five musical pieces they knew and routinely rehearse (Mozart K136-1, K550-1). They played all pieces three times, with their own conductor (OWN) and another professional conductor they never played with (NEW). Directed influences between participants were computed by using the Granger Causality (GC) method [8-9].

In our experiment we explored whether conductors' kinematics were associated to a differential influence on musician's performance (driving force) and if this was able to affect inter-musicians communication (interaction strength). Furthermore, we had independent expert musicians (offline and blind to the scope of the experiment) complete a questionnaire about each audio recording. The questionnaire contained items testing several subjective scales such as their ability to follow the piece (separately for melody and rhythm), the degree of musical entrainment and emotional involvement.

3 Results and Discussions

Results of GC analyses showed that the causal influence between conductors and each player was different across conductors. The New conductor drives (greater GC strength) the orchestra in two pieces out of five - namely piece 3 and 5. Moreover, we also found that players show different pattern of driving forces within them when they play under the direction of the two conductors. Under the direction of the New conductor, each player has less influence on each other in three pieces out of five - namely piece 1, 2 and 3. Questionnaire data revealed a significant interaction between Conductor and Piece but no simple effects. The interaction was further explored with follow-up tests, revealing a difference between conductors in pieces 3 and 5.

Results show that the two conductors exhibit different driving forces strength towards musicians in some pieces, whereas communication strength among players was also modulated by the characteristics of the two conductors in others. Conductors' directed drive differed in two of the pieces (3, and 5), whereas the conductors modulated inter-musician influences in three pieces (1, 2 and 3).

However, we have to take into account that such dynamical network of causal interactions was aimed at producing a pleasurable effect in the listeners. Interestingly, aesthetic evaluations were modulated by piece and conductor. Specifically, two pieces (number 3 and 5) were considered significantly different between conductors. Interestingly enough, piece number 3 was the only one showing a differential influence of one conductor over the other in affecting players and at the same time affecting inter-musicians communication. Therefore, it might be the case that the aesthetic appreciation

of music orchestras' performance is based on the concurrent increase of conductor to musician causal influence and a reduction of musician-to-musician information flow.

In conclusion, the present results add to the growing body of research that considers musicians as a perfect model to study sensory-motor brain plasticity and organization [10]. Here, we used musicians as a model of how effective communication might be based on efficient gestures coordination. In fact, each musician is certainly reading a score, knows perfectly what to play, and can listen and see what other musician do. However, the violinist has to concurrently follow the conductor, providing critical information on how to interpret a given piece. The musician has to wisely balance several external sources of information and mix them up in order to reach the required performance.

References

1. Rizzolatti, G., Craighero, L.: The mirror-neuron system. *Annu. Rev. Neurosci.* 27, 169–192 (2004)
2. Couzin, I.D., Krause, J., Franks, N.R., Levin, S.A.: Effective leadership and decision-making in animal groups on the move. *Nature* 433(7025), 513–516 (2005)
3. Frith, C.D.: Social cognition. *Philos Trans. R Soc. Lond B Biol. Sci.* 363(1499), 2033–2039 (2008)
4. Grafton, S.T., Hamilton, A.F.: Evidence for a distributed hierarchy of action representation in the brain. *Hum. Mov. Sci.* 26(4), 590–616 (2007)
5. Newman-Norlund, R.D., van Schie, H.T., van Zuijlen, A.M., Bekkering, H.: The mirror neuron system is more active during complementary compared with imitative action. *Nat. Neurosci.* 10(7), 817–818 (2007)
6. Rands, S.A., Cowlshaw, G., Pettifor, R.A., Rowcliffe, J.M., Johnstone, R.A.: Spontaneous emergence of leaders and followers in foraging pairs. *Nature* 423(6938), 432–434 (2003)
7. Sebanz, N., Bekkering, H., Knoblich, G.: Joint action: bodies and minds moving together. *Trends. Cogn. Sci.* 10(2), 70–76 (2006)
8. Geweke, J.: Measurement of linear dependence and feedback between multiple time series. *J. Am. Stat. Ass.* 77, 304–313 (1982)
9. Granger, C.W.J.: Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37, 424–438 (1969)
10. Münte, T.F., Altenmüller, E., Jäncke, L.: The musician's brain as a model of neuroplasticity. *Nat. Rev. Neurosci.* 3(6), 473–478 (2002)

Author Index

- Aaltonen, Viljakaisa 1
Ach, Laurent 240
Alofs, Thijs 123
Aloimonos, Yiannis 273
- Badino, Leonardo 273
Bargar, Robin 170
Bevacqua, Elisabetta 243
Bianchi-Berthouze, Nadia 73, 149
Boguslawski, Gemma 73
Bryan-Kinns, Nick 260
- Camurri, Antonio 63, 103, 229
Canazza, Sergio 231
Canepa, Corrado 103, 229, 231
Canto, Rosario 273
Catanese, Salvatore 209
Cavallero, Francesca 103
Cera, Andrea 229
Chevreau, Karine 240
Chiorri, Carlo 63
Choi, Insook 170
Chumerin, Nikolay 28
Coletta, Paolo 63, 229, 231
Combaz, Adrien 28
Cosi, Piero 193
Costa, Cristina 223
Craighero, Laila 273
Crevoisier, Alain 236
- Damiano, Rossana 203
D'Ausilio, Alessandro 273
de Mazancourt, Hugues 240
Dertien, Edwin 38
de Sevin, Etienne 243
Durieu, Laurent 240
Dutoit, Thierry 249
- Eyben, Florian 243
- Fadiga, Luciano 273
Ferrara, Emilio 209
Ferrari, Nicola 103
Fiumara, Giacomo 209
Foresti, Gian Luca 231
- Gerzso, Andrew 229
Ghisio, Simone 63, 229
Glowinski, Donald 63, 93, 219, 221, 229
Gosselin, Bernard 249
Grandjean, Didier 252
Gürkök, Hayrettin 183
- Hakvoort, Gido 183
Henao Ramirez, Eduardo Andres 48
Heylen, Dirk 243
- Keränen, Jaakko 1
- Labbé, Carolina 252
Lamberti, Fabrizio 48
Leone, Giuseppe Riccardo 193
Leroy, Julien 249
Li, Yi 273
Lombardo, Vincenzo 203
- Mader, Angelika 38
Maestre, Esteban 267
Mancas, Matei 249
Mancini, Maurizio 63, 93, 219, 221, 229
Manuri, Federico 48
Manyakov, Nikolay V. 28
Marchesoni, Michele 223
Marchini, Marco 267
Massari, Alberto 93, 219, 221
Mazzarino, Barbara 103
Minuto, Andrea 57
Morel, Benoit 240
Muñoz, Jesús 223
- Nijhar, Jasmir 73
Nijholt, Anton 12, 57, 129, 139, 160, 183
Normier, Bernard 240
Nunnari, Fabrizio 203
- Obbink, Michel 183
Osman, Mohd Azam 246
- Paci, Giulio 193
Pagano, Francesco 209
Pammi, Sathish 243
Pantic, Maja 160

- Papiotis, Panagiotis 267
 Paravati, Gianluca 48
 Pelachaud, Catherine 240, 243
 Perez, Alfonso 267
 Pez, André-Marie 240
 Picard-Limpens, Cécile 236
 Piman, Sirot 113, 246
 Plass-Oude Bos, Danny 183
 Poel, Mannes 183
 Poelman, Wim 57

 Reidsma, Dennis 38, 83, 129
 Reponen, Erika 1
 Riche, Nicolas 249
 Robben, Arne 28
 Rodà, Antonio 231
 Romano, Filippo 231
 Ron-Angevin, Ricardo 18

 Sancha-Ros, Salvador 18
 Sanna, Andrea 48
 Savva, Nikolaos 149
 Scattolin, Francesco 231
 Schröder, Marc 243
 Schuller, Björn 243

 Sun, Xiaofan 160
 Swartjes, Ivo 123

 Talib, Abdullah Zawawi 113, 246
 Tan, Kian Lam 246
 ter Maat, Mark 243
 Tetteroo, Daniel 129
 Theune, Mariët 123
 Tilmanne, Joelle 249
 Tokay, Sera 273
 Torres-Eliard, Kim 252

 van der Sluis, Frans 139
 van Dijk, Betsy 129, 139
 Van Hulle, Marc M. 28
 van Vliet, Marijn 28
 van Welbergen, Herwin 83
 Velasco-Álvarez, Francisco 18
 Volpe, Gualtiero 63, 103, 231
 Vyas, Dhaval 57

 Wöllmer, Martin 243

 Zajega, François 249
 Zanolli, Serena 231