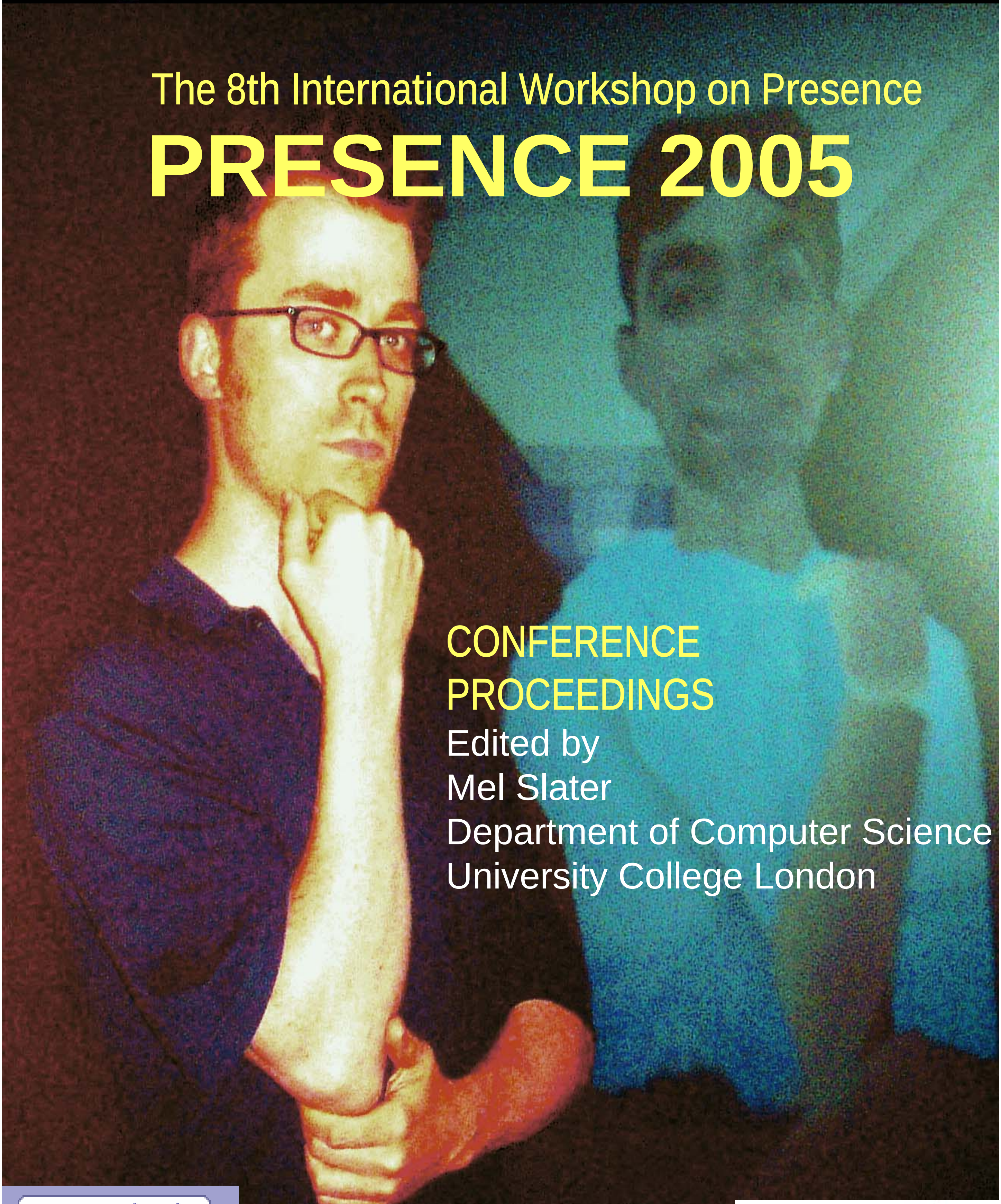
International Society for Presence Research

The 8th International Workshop on Presence

PRESENCE 2005



CONFERENCE PROCEEDINGS

Edited by
Mel Slater
Department of Computer Science
University College London

21-23rd September, 2005
University College London
UK

Information Society Technologies

PRESENCIA

International Society for Presence Research

The 8th Annual International Workshop on Presence

PRESENCE 2005
21-23rd September, 2005
University College London

Annual Conference Proceedings
ISBN 0-9551232-0-8

Edited by:
Mel Slater
International Programme Committee Chair

Local Committee

Administration:
Ms JJ Giwa, Department of Computer Science, University College London

Demonstrations Chairs:
Vinoba Vinayagamoorthy and Doron Friedman, Department of Computer Science, University College London, UK

Posters Chair:
Daniela Romano, Department of Computer Science, University of Sheffield, UK

International Society for Presence Research:

President:
Matthew Lombard, Temple University, USA

Published by
University College London – Department of Computer Science
Gower Street
London
WC1E 6BT
UK

International Programme Committee

Mariano Alcaniz, Universidad Politécnica de Valencia, Spain
Carlo Alberto, Avizzano, Scuola Superiore S. Anna, Italy
Jeremy N. Bailenson, Stanford University, USA
Rosa Baños, University of Valencia, Spain
Woody Barfield, Independent Researcher, North Carolina, USA
Edwin Blake, University of Cape Town, South Africa
Cristina Botella Arbona, Universitat Jaume I, Spain
Frank Biocca, M.I.N.D. Lab, Michigan State University, USA
Doug Bowman, Virginia Tech, USA
Cheryl Campanella Bracken, Cleveland State University USA
Martin Buss, Technische Universitaet Muenchen Germany
Alan Chalmers, University of Bristol, UK
Jonathan Freeman, Goldsmiths College, University of London, UK
Doron Friedman, University College London, UK
Luciano Gamberini, University of Padua, Italy
Maia Garau, University College London, UK
Marco Gillies, University College London, UK
Ilona Heldal, Chalmers University of Technology, Sweden
Wijnand Ijsselsteijn, Technische Universiteit Eindhoven, Netherlands
Roy Kalawsky, Loughborough University, UK
Rita Lauria, North Carolina Agricultural and Technical State University, USA
Jane Lessiter, Goldsmiths College, University of London, UK
Matthew Lombard, Temple University, USA
Celine Loscos, University College London, UK
Katerina Mania, University of Sussex, UK
Giorgio Metta, Università degli Studi di Genova, Italy
Christa Neuper, University of Graz, Austria
Miriam Reiner, Technion: Israel Institute of Technology, Israel
Albert Skip Rizzo, University of Southern California, USA
Daniela Romano, University of Sheffield, UK
Roy Ruddle, University of Leeds, UK
Maria Victoria Sanchez-Vives, Instituto de Neurociencias de Alicante, Spain
Ralph Schroeder, Oxford Internet Institute, UK
Thomas Schubert, University of Jena, Germany
Melissa E. Selverian, Temple University, USA
Mel Slater, University College London, UK
Anna Spagnolli, University of Padua, Italy
Anthony Steed, University College London, UK
Aleksander Väljamäe, Chalmers University of Technology, Sweden
Walter Van de Velde, European Commission, EU
Paul F.M.J. Verschure, Institute of Neuroinformatics, Switzerland
Vinoba Vinayagamoorthy, University College London, UK
Suzanne J. Weghorst, University of Washington, USA
Mary Whitton, University of North Carolina, USA

Preface

I am very pleased to have organised this 8th Annual Workshop on Presence. I organised the first Workshop in 1998 in collaboration with Peter Lawrence of BT Labs, near Ipswich in England[1]. Since then the field has advanced significantly, not least due to the considerable funding that it has received at the European level through the 5th Framework Future Emerging Technologies (FET) Presence Research Initiative, with projects running from 2002-2005, and with excellent future prospects due to the further round of funding expected in the 6th Framework Programme ‘Presence and Interaction in Mixed-Reality Environments (Presence II)’.

As a result of these advances a more rigorous editorial approach has been taken for this year’s conference. As before each paper was assigned two or three referees (depending on whether it was a short or full length paper). The size of the International Programme Committee was expanded, and members were instructed to give very full reviews of submitted papers, almost up to journal standards. 71 papers were submitted, 22 were accepted as full papers, and 16 as short papers. Full papers could be up to 12 pages and short papers up to 6 pages. In addition a new category has been introduced, referred to as Sketches. These are those papers that were not accepted as full or short papers but where one referee gave a high enough rating that the work was considered interesting enough to include in the Proceedings. Such Sketches will not be presented during the main conference sessions but the authors will be presenting posters during the timetabled Poster Sessions. There are 9 such Sketches. In addition to the posters corresponding to those Sketches there will be several additional posters available for discussion during the Poster Session times.

Two special issues of the MIT Press journal *Presence: Teleoperators and Virtual Envrionments* have been reserved for 2006 to take those papers most suitable for the journal at the time of conference submission. Eighteen papers have been recommended for a further round of reviews for the journal.

This year there are three keynote speakers. Paul Verschure introduces the ADA system, an intelligent entity with which people can interact within a large scale virtual environment. This work is of importance more generally since it illustrates the growing intersection between presence and neuroscience research, an outcome directly traceable to the European FET initiatives. Woody Barfield was a pioneer of presence research in the early 1990s when he was a Professor at the University of Washington near Seattle, USA, and later at Virginia Tech. He has since taken a law degree and has become expert in the area of legal issues relating to virtual environments, mixed reality and virtual personhood. Carolina Cruz Neira is internationally known through her research that led to co-invention of the CAVE system at University of Chicago at Illinois, USA, in the early 1990s. She has since gone on to found a thriving research group on all aspects of virtual environments at Iowa State University and she initiated the annual ‘International Immersive Projection Technology Workshop’ that has been instrumental in presenting the ever expanding research and applications in projection technology systems.

---

[1] http://www.cs.ucl.ac.uk/staff/m.slater/BTWorkshop

The conference itself has 12 sessions each covering a major theme surrounding the concept, application, measurement and experimental aspects, and technical means for the realisation of presence. There is also a Panel Session organised by members of the OMNIPRES (FET) project, which had the role in the first FET ‘Presence Research’ funding of providing an umbrella organisation for the 10 additional funded projects.

This conference forms part of the output of the PRESENCIA (FET) project, which has contributed to its organisation and funding, in addition to the scientific programme. Each of the funded ‘Presence I’ projects organised at least one grand meeting of all the projects, usually in the form of a conference that could be attended by other scientists not involved in those projects. This Presence 2005 conference is the last such meeting, and during this occasion a number of the FET Presence projects will be enduring their final formal evaluations by committees of experts.

I would like to take the opportunity to thank every member of the International Programme Committee for their hard work during the period of the paper reviews and subsequent discussions. More papers were submitted than expected and therefore in spite of the IPC having been considerably expanded, each member had more reviews to do than had been expected. For the first time in any Presence workshop, a Programme Committee meeting was arranged, and it was to be on 8th July, 2005 in London. Unfortunately the tragic attacks in London on 7th July made this impossible. As a result during the subsequent week members of the IPC got together virtually and discussed each paper and made final recommendations. This was additional work for all, but possibly resulted in more thoughtful decisions regarding the papers with probably more discussion than had the physical meeting taken place. The final quality of papers is excellent, and I hope that this is the start for the Presence conference to become the internationally recognised high quality, must ‘publish in there or bust’ annual event for researchers in this field.

Finally I would like to take the opportunity to thank the local organisers, J. J. Giwa, Doron Friedman, Vinoba Vinayagamoorthy and Daniela Romano, and the student assistants (at the time of writing unknown) for their considerable help in organising this event. I would like to thank Matthew Lombard, the President of ISPR for his significant help throughout this process – ranging from providing the web pages through to reassuring me in those dark moments when the Word files for these Proceedings wouldn’t join together!

Finally I hope that this conference will be remembered as an enjoyable and intellectually stimulating event for everyone involved, and that it will have some positive impact on all our lives in the future.

Mel Slater
August, 2005.

About ISPR and the PRESENCE Workshops

The PRESENCE Workshops began with the BT Presence Workshop at BT Labs in Suffolk, England on June 10 and 11, 1998, organized by the chair of this 2005 conference, Mel Slater. After three more productive but informal gatherings -- at the University of Essex in Colchester, England in April 1999, Eindhoven University of Technology in Delft, The Netherlands in March 2000 and Temple University in Philadelphia, Pennsylvania in the U.S. in May 2001 -- the International Society for Presence Research (ISPR) was founded to coordinate future annual conferences, sponsor other conference panels and events, and provide a variety of resources to those who conduct research, develop theory, write about, or simply are interested in, the concept of presence. The ISPR web site, at http://ispr.info, provides many of these resources (as well as a list of the organization's current Board of Directors). ISPR also sponsors the presence-l listserv (see ispr.info for details).

With outstanding local organizers, ISPR coordinated the 5th successful PRESENCE Workshop in October 2002 at Universidade Fernando Pessoa in Porto, Portugal, the 6th Workshop in October 2003 at Aalborg University in Aalborg, Denmark, and last fall’s conference at Polytechnic University of Valencia, in Spain. The Workshops have retained their single track format and productive but informal and very pleasant character, and thanks to the hard work and dedication of everyone involved, I’m confident that PRESENCE 2005, the 8th Annual International Workshop on Presence here at University College London will provide all of us another excellent experience as we explore the latest work in a fascinating field.

The ISPR Board is planning future international presence conferences, enhancing the resources it provides online, developing procedures to offer organizational memberships, and considering the development of a journal; look for announcements about all of these at ispr.info and on the presence-l listserv. ISPR exists to serve the presence community and we always welcome questions, comments and suggestions at help@ispr.info (or directly to me at lombard@temple.edu).

Enjoy the conference!

Matthew Lombard, Ph.D.
Temple University
Philadelphia, Pennsylvania

President, International Society for Presence Research
http://matthewlombard.com
http://ispr.info

Contents

Keynote 1

Session 1 Interacting with a Brain

Session 2 Body, Space and Motion

Session 3 Interaction with Avatars

Keynote Talk 2

Session 4 Meta Presence

Session 5 Individual and Cultural Effects

Session 6 Talking Faces and Social Collaboration

Session 7 Presence Theory and Experimentation

Keynote Talk 3

Session 8 Interactivity and Usability - Theory and Practice

Session 9 Multimodal Presence

Session 10 Panel Session

Session 11 Therapy

Session 12 Gaming and Connected Presence

Sketches

Experimenting with Ada: Towards collective mixed-reality environments

Paul F.M.J. Verschure
Institute of Neuroinformatics, University of Zurich & Swiss Federal Institute of Technology (ETH)
Zurich, Switzerland

Abstract

We have constructed a shared mixed reality environment that can support simultaneous interaction with dozens of humans. This space, called Ada, was presented at the Swiss national exhibition Expo.02 and was visited by 553.700 people during the 6 months of this event. Ada raises a number of questions concerning the methods and technologies that facilitate the construction of real-world systems consisting of very dense sensor and effector networks, and the approaches that allow for effective interactions between such a space and its visitors. Ada's design was based on a neuromorphic approach where the artefact itself was conceived as a sentient organism, its central control systems were based on large-scale neuronal models and its modes of interaction as behaviours subserving specific allocentric needs. In this presentation I will describe the key components of Ada and present a quantitative and qualitative analysis of its performance and impact on human behaviour and experience.

Session 1

21st September, 2005

Interacting with a Brain

11.30-12.00 *An investigation of collective human interaction with a large-scale, mixed-reality space*
Kynan Eng[1], Matti Mintz[2] and Paul F.M.J. Verschure[1]
[1] Institute of Neuroinformatics, University/ETH Zurich, Switzerland
[2] Department of Psychology, Tel-Aviv University, Israel

12.00-12.30 *Walking from thoughts: Not the muscles are crucial, but the brain waves!*
Robert Leeb[1], Claudia Keinrath[1], Doron Friedman[2], Christoph Guger[3], Christa Neuper[1,4], Maia Garau[2], Angus Antley[2], Anthony Steed[2], Mel Slater[2] and Gert Pfurtscheller[1,5]
[1] Laboratory of Brain-Computer Interfaces, Institute for Computer Graphics and Vision, Graz University of Technology, Austria
[2] Department of Computer Science, University College London, United Kingdom
[3] g.tec - Guger Technologies OEG, Austria
[4] Department of Psychology, University of Graz, Austria
[5] Ludwig-Boltzmann Institut für medizinische Informatik und Neuroinformatik, Graz University of Technology, Austria

12.30-12.45 *Reliving VE Day With Schemata Activation*
Phil Turner[1], Susan Turner[1] and Dimitrios Tzovaras[2]
[1] School of Computing, Napier University, UK
[2] Informatics and Telematics Institiute, Thermi-Thessaloniki, Greece

An Investigation of Collective Human Behavior in Large-scale, Mixed Reality Spaces

Kynan Eng[1], Matti Mintz[2] and Paul F. M. J. Verschure[1]
[1]Institute of Neuroinformatics, University/ETH Zurich, [2]University of Tel-Aviv, Israel
kynan@ini.phys.ethz.ch, mintz@freud.tau.ac.il, pfmjv@ini.phys.ethz.ch

Abstract

Future mixed reality systems will need to support large numbers of simultaneous, non-expert users at reasonable per-user costs if they are to be widely deployed within society in the short to medium term. We have constructed a prototype of such a system, an interactive entertainment space called Ada that was designed to behave like a simple organism. Using Ada we conducted two studies: the first assessing the effect of varying the operating parameters of the space on the collective behavior and attitudes of its users, and the second assessing the relationship between user demographics, behavior and attitudes. Our results showed that small changes in the ambient settings of the environment have a significant on both user attitudes and behavior, and that the changes in user attitudes do not necessarily correspond to the environmental changes. We also found that individual user opinions are affected by demographics and reflected in overt behavior. Using these results, we propose some tentative guidelines for the design of future shared mixed reality spaces.

Keywords--- mixed reality, shared space, interactive space, presence.

1. Introduction

1.1. Background & Motivations

Most current virtual reality and augmented reality systems use multi-channel video displays and sound output to give the user a sense of being present at an alternate location or in a complex data space. Existing work in computer graphics has elucidated some of the relevant scene rendering parameters for maximizing presence [1], as well as various enhancements of both visual and sound modalities to enhance the sensation of presence. Such enhancements include immersive multi-screen surround projection, 3D stereo point-of-view projection [2] (sometimes combined with head-mounted position and orientation tracking), head-mounted or hand-held displays with superimposed data representations [3], spatialized audio output, etc. Some systems add the modality of touch to further improve both the subjective sensation of presence and task performance [4, 5].

While these methods are known to provide benefits for the sensation of presence, it is not known if collectives of non-expert users can use such systems. Most of the above-mentioned methods for enhancing presence are not yet suited for such large-scale deployment in populations of non-expert users, as they typically have high per-user costs, and they require each user to wear or carry an (expensive) object such as a head-mounted display, a tracking device, a joystick or a personal digital assistant (PDA). Each user may also be required to be alone in a purpose-built room with high-resolution tracking systems and surround projections. In the short to medium term it will only be possible to expose large numbers of non-expert users to immersive virtual or augmented reality environments that require an absolute minimum of specialised hardware for each individual person. To minimise the cost per user, such multi-user systems will be characterised by:

- Shared physical space
- Shared sensors and output hardware
- Multi-purpose environment

We created such a space, an exhibit called Ada, which ran for 5 months during the Swiss National Exposition in 2002 and received 553,700 visitors from the general public. Ada is a multi-purpose interactive space, conceived as an artificial organism, that is designed to engage visitors in entertaining interactions [6, 7]. The space has also been used as an auditorium for an awards ceremony and as a disco. Ada's input modalities include visitor tracking using pressure-sensitive floor tiles [8] over a 160 m^2 floor area (Fig. 1), detecting and localizing handclaps and simple sounds such as the spoken word "Ada" using two sets of three microphones, and capturing video in real-time using ten pan-tilt cameras called *gazers*. Output is provided by local and global speakers, pan-tilt "light fingers" for illuminating selected visitors, colored neon lamps in the floor tiles, and BigScreen: a 360° projection surrounding the space that can show dynamic 3D objects and live video on a single virtual display. The system is controlled by a distributed mix of agent-based software, simulated neural networks and procedural code on a computer cluster (30 AMD Athlon 1800+).

The visitor flow (Fig. 2) was controlled to guarantee a certain quality of visitor experience and throughput. Visitors queued for up to 90 min. at the entrance, viewing a 10-minute video about Ada called *Brainworkers* [9] and reading a leaflet explaining the exhibit. They entered in groups of about 25-30, passing first through the *conditioning tunnel* with several interactive stations that introduced Ada's sensory and motor components. They then waited behind one-way mirrors in the *voyeur area*, observing the group in front of them in Ada's main space. Once it was their turn, they interacted with Ada before heading into the *brainarium* – an area where visitors could view real-time displays of Ada's internal states and see how they correlated with the actions of the following group. Finally, visitors entered the *explanatorium* area, featuring artistic and social discussion elements. Visitors spent about 5 minutes in each section, for a total stay of about 25 minutes.

Although Ada could support many interaction scenarios, the contractual exhibit requirements dictated a minimum level of "normal" Ada functionality. Hence the user interactions in the main space were built on six behavioral modes (Table 1) presented in a fixed visitor cycle of about 5-6 minutes in length, with interaction-dependent timing variations.

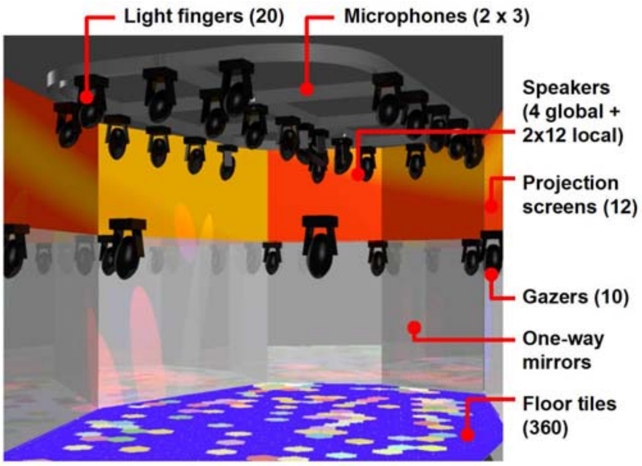


Figure 1: Layout of components in Ada main space. From [6].

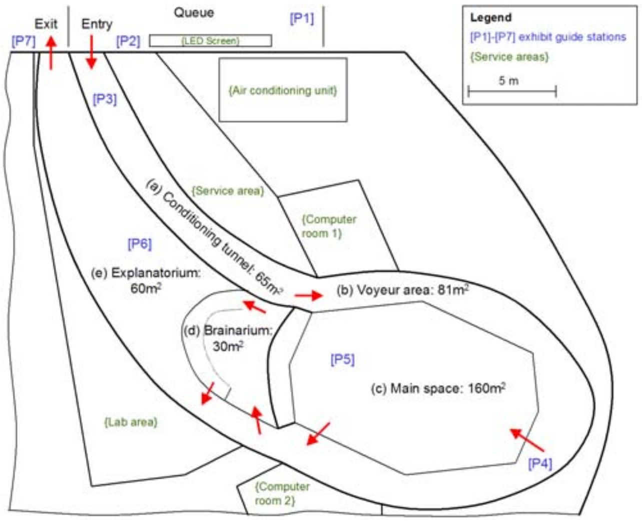


Figure. 2: Floor plan of the Ada exhibit. Arrows indicate visitor flow. Stations for exhibit guides are indicated by P1 through P7. The total publicly accessible floor area of the exhibit was 396 m^2. From [10].

Mode, length	Ambience & Interactions
Sleep 35±3s	*Floor*: Blue, slowly pulsating *Screen*: Dark blue, slow upwards-drifting texture *Sound*: Soft, low-pitched soundscape *Interactions*: Blue transient pulses on floor and "splashing" noises in response to walking. Transition to *Wake* when visitors clap their hands while Ada is in "light" sleep.
Wake 24±1s	*Floor*: Rapid change to bright yellow, then slow fade *Screen*: Bright yellow texture *Sound*: Higher volume, pitch *Interactions*: Tracking assigns colored tile to each visitor. Handclaps localised by briefly drawing halo around each clap source
Explore 103±8s	*Floor*: Light gray *Screen*: Light brown rock texture *Sound*: Bright, open *Interactions*: Periodic "compliance" testing by deploying blinking white tile cues in front of visitors; those visitors that followed the cues for long enough were rewarded with a pulsating ring of tiles, light fingers and a gazer that followed them around. In addition, live video, still snapshots and the tracked path of the visitor were displayed on BigScreen in a position aligned with their direction of motion to maximise their chance of being seen. This simple viewpoint registration method is broadly similar to those found in augmented reality systems [11].
Group 33±6s	*Floor*: Black *Screen*: Dark gray/black background, bright green highlights *Sound*: Foreboding *Interactions*: As for *Explore* Cue learning experiments were occasionally carried out during this mode, in which Ada tried to learn the most effective cues for influencing visitors' positions [12]
Game 64±10s	*Floor*: Green, red borders *Screen*: Yellow/orange texture *Sound*: Lively, loud *Interactions*: Visitors try to step on a bouncing animated tile "ball"
End 35±2s	*Floor*: Red travelling waves towards exit *Screen*: Dark red/orange downwards-drifting texture *Sound*: Soft, sad *Interactions*: Animated white tile "bullet" cues direct visitors to the exit. Switch back to *Sleep* when last visitor leaves.

Table 1: Summary of Ada's behavior modes, ambiences and interactions. Mode length ranges are shown as mean ± standard deviation.

1.2. Ada, Mixed Reality and Presence

Ada is a large-scale, multi-user interactive space, but where does it fit in the taxonomy of virtual reality systems? While Ada does fit one classical definition of a virtual reality system as being a means for humans to visualise, manipulate and interact with computers and extremely complex data [13], it does not represent an external reality. Rather, it communicates with its users using what we call *reality-based metaphors* that match what humans are accustomed to in real life: persistence of objects represented on the floor and the screen, Newtonian mechanics to provide “wave-like” effects and visual object motion and collision, co-location of visual events and their related sound effects, etc. It does provide basic reality augmentation – visitors can see their tracked path project on the screen in a display window that moves to match their walking direction – but since Ada is a physically immersive space, there is no “everyday” reality to augment. It was not possible to visit Ada remotely in a telepresence mode [14]; however, “virtual visitors” could be generated within the system to inhabit the floor space. These attributes of Ada mean that it does not seem to fit easily into the three-dimensional “transportation”, “artificiality” and “spatiality” taxonomy of shared spaces proposed by Benford et al. [15]; in this scheme Ada would rate as highly artificial (no representation of external reality) and moderately spatial (some representation of relative positions in space), but the level of transportation is unclear since Ada is representing an abstract data space rather than a “real” environment. The concept of “spatiality” is also somewhat problematic in Ada, since the physical space of Ada is obviously highly “spatial” as a representation of itself. We suggest that the best available definition of Ada is that of a mixed reality space: a merging of real and virtual worlds to produce a new environment where physical and virtual objects can co-exist and interact [16].

The measurement of presence in a space like Ada can be expressed as the extent to which users acknowledge that Ada is a computer-based social actor [17, 18], where the medium of Ada as a whole behaves as a social actor [19]. The design of Ada explicitly expresses the idea of a social actor, with innate goals and an internal emotional model [10]. Users express this acceptance of the presence of Ada implicitly through their behavior [20] and explicitly via their responses to questionnaires.

The experimental part of this paper is divided into two sections. The first section examines the extent to which Ada’s visitors acknowledged it as a social presence, and explores the relationship between Ada’s operating parameters and the attitudes and behavior of its visitors. The second section investigates the effects of visitor demographics on their attitudes and probes the extent to which individual attitudes are reflected in behavior. This is followed by a discussion of the implication of the results for the design of future large-scale, mixed-reality spaces.

2. Experiments I: Collective Human-Space Interaction

Generating the sense of presence of Ada as a medium requires coherence between the modalities being presented to the user. This is achieved using its emotional model and the reality-based metaphors described earlier. For users to discover this coherence, they need to actively explore the environment. This leads us to generate the following hypotheses:

- Reductions in Ada’s output coherence by reducing the level of one of Ada’s output modalities should decrease reported presence levels and affect activity levels.
- Reduction or removal of an output modality should reduce the reported effectiveness of that modality.
- Excessively high user density should decrease activity and reported presence, since the density of other users will affect both the visibility of Ada via occlusion and the space available for individuals to interact with Ada. Conversely, low visitor density should increase activity and reported presence levels.

2.1. Methods I

Experiments were based on a standard control case and a set of small deviations from this case, in order to minimise any disruption to normal exhibit operation. Floor tracking and floor occupancy data, audio processing signals and MPEG-4 video data were recorded from Ada (up to 5 GB/hour). A timeserver synchronised timestamps across the cluster to within 100 ms. Analysis was performed using Matlab 6.1 (Mathworks, MA, USA) and SPSS 11 (SPSS Inc., IL, USA).

Public pre-exposure to Ada consisted mainly of a mass-media advertising campaign (TV, print, web), in which prospective visitors were told that Ada was an “intelligent” space with a distinct identity. This framing may have influenced some visitors to expect to encounter a kind of entity, but anecdotal evidence suggests that almost all visitors had very little idea of what to expect from Ada before arriving at the exhibit.

Questionnaires were distributed to specific groups of visitors in their choice of German, French and English (the first or second language of virtually all visitors) as they exited the main space. Because our audience consisted of the general public of any age, we opted for a very general phrasing of the questions rather than a specialised presence questionnaire (eg. [21]). Participation was voluntary, and a majority of visitors agreed to participate. Observers were on hand to ensure that visitors did not discuss or copy each other’s answers. Almost all visitors completed the questionnaire within 5 minutes. The first section of the questionnaire collected demographic information (Table 2A): the gender, age, first language, education level and main education type of the participants. The second section required visitors to respond to a set of statements about their interactions with Ada, with responses given on a ten-point scale (Table 2B). The statements fell into four categories: Ada’s sensory abilities (Q1-3), Ada’s reactions to visitors’ actions (Q4-7), perceptions of visitors’ own reactions to Ada’s actions (Q8-12) and overall impressions of Ada (Q13-16). Question 15 explicitly assessed users’ opinions of Ada as a socially active medium (“I felt that Ada is a kind of creature”). Children under 10 years of age were excluded from taking the questionnaire. Questionnaires with more than 4 unanswered questions were discarded. Every valid questionnaire was included in the analysis, except where a single clear response to a particular question could not be ascertained. The visitor demographics in each test session were balanced as necessary by discarding randomly selected questionnaires, or by pooling results from multiple sessions with equivalent operating conditions. For analyzing the responses to the statements, the boxes ticked by the visitors were converted into integers from 1 to 10, with 1 corresponding to most “disagree” with the statement and 10 corresponding to most “agree” with the statement. The question items were analyzed separately without creating combined scales in order to allow.

The experiments themselves were divided into three sections:

1. **Control case:** group behavior and attitudes under normal operating conditions;
2. **Effect of Ada operating parameter variations:** various operating parameters of the space were manipulated to gauge their effect on visitor behavior and attitudes. Visitors did not have the opportunity to observe “nominal” Ada behavior beforehand.
3. **Effect of group size variations:** on behavior and attitudes was also investigated.

2.2. Results I

2.2.1 Control Case

Visitor behavior during the control case followed characteristic patterns that were affected by Ada’s behavior modes (Figures 3 & 4). During Sleep, new visitors entered the space from the lower left corner while the previous group exited via the upper left corner. Handclaps and spoken “Ada” events were detected at around 0.5 Hz and 0.015 Hz, respectively. In the subsequent modes, *Wake*, *Explore* and *Group*, the visitor occupancy tended towards a uniform spatial distribution, but visitors avoided the entrance area. They increased their handclap rates to ~1.4 Hz and their spoken “Ada” rates to ~0.02 Hz, with a drop-off during Group mode. During *Game* mode they spent more time in the half of the space closest to the exit compared to the other half, handclap activity dropped to below 0.5 Hz and spoken “Ada” detection increased. The distributions of tile on/off events during this mode was very different to that in Explore and *Group*, with two tile event rate peaks in the top and bottom half of the space. These peaks correspond to the playing fields for the games where people were moving very fast, while those standing at the borders did not generate as many tile events. During *End* mode, visitors accumulated at the exit and they all but stopped making noise.

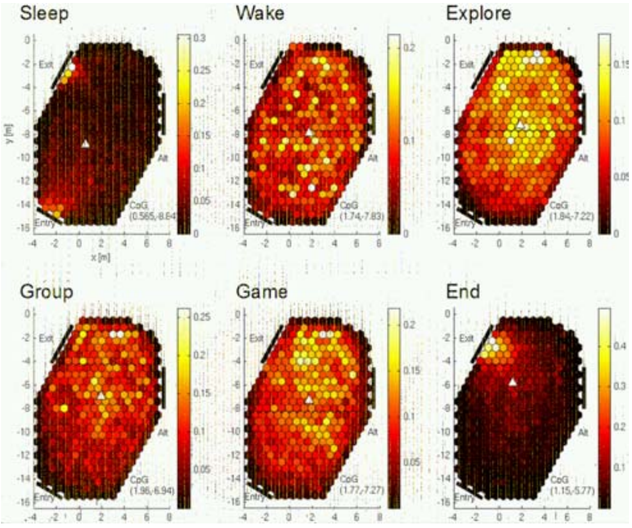


Figure 3: Control case mean floor occupancy for each behavior mode. Averaged over 12 visitor cycles. White triangle = center of gravity of distribution. From [22].

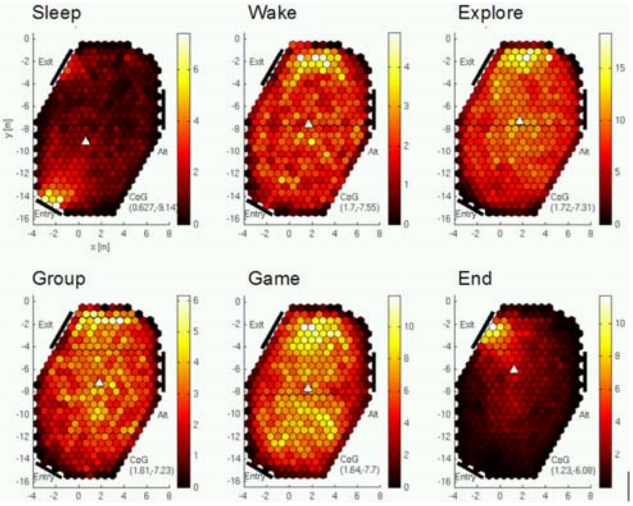


Figure 4: Control case mean event rate for each behavior mode. Averaged over 12 visitor cycles. White triangle = center of gravity of distribution. From [22].

The control case questionnaire responses (N = 74, 0 invalid questionnaires) showed that Ada's modalities are rated differently (Table 2B): Ada’s sensory modalities (Q1-3) were assigned varying importance by the visitors [ANOVA, F(2, 234) = 4.29, p = 0.015], with Ada's vision being rated significantly lower than tactile sensing (all post-hocs with Bonferroni correction, p = 0.013), but equal to auditory sensors (p = 0.199). Ada’s effector modalities (Q4-7) were rated differentially [F(3, 297) = 7.43, p < 0.001], due to the high rating of the floor output (Q4; p < 0.012 for all comparisons between Ada’s floor effects and the other modalities); the other modalities were indistinguishable from each other (p > 0.5 in all comparisons).

Visitors claim they are very active and not imitating other visitors: Visitors' ratings of their own reactions to Ada (Q8-12) were significantly different to each other [F(4, 388) = 45.3, $p < 0.001$]: the ratings were all distinct (post-hoc, $p < 0.011$) except the pairs of questions (Q8 & Q9, $p = 1.0$) and (Q10 & Q12, $p = 1.0$). Thus, the visitors' ratings in this section fell into three broad groups: very high (visitor self-motion, Q8 & Q9), medium (making noise Q11), and low (looking at BigScreen and imitation of other visitors, Q10 & Q12).

Some visitors thought that the interactive space had creature-like properties: The overall ratings of Ada (Q13-16) were significantly different to each other [F(3, 325) = 28.6, $p < 0.001$]: only the question pairs (Q13 & Q16) and (Q14 & Q16) were indistinguishable ($p > 0.2$), while all other pairs of questions were significantly different ($p < 0.005$). Visitors stated most strongly that they liked Ada and were happier after being in Ada (Q14, 16). Almost as high was their tendency to say that Ada affected their behavior (Q13), while their ratings were lowest for the statement that Ada was a kind of creature (Q15). However, the high standard deviation for this statement indicates that a considerable minority was amenable to the idea that the space was acting as a unitary entity, rather than as a collection of components.

Demographic measure	Available options			
Gender	Male	Female		
Age [years]	10-15 41-50	16-20 51-60	21-30 61+	31-40
First language	German English	French Other	Italian	
Education level	Basic schooling Apprenticeship University degree	Completed high school Technical training Doctorate		
Education type	Technical Arts Business	Natural sciences Social sciences Other		

Table 2A: Ada questionnaire demographic measures and available options.

Category	#	Question body (response 1…10)	Avg	Std.
Ada sensed me with:	1	Eyes	5.70	2.89
	2	Ears	6.74	2.96
	3	Skin	7.10	2.88
Ada reacted to my actions by:	4	Producing light patterns on her skin	7.25	2.59
	5	Projecting patterns on the Big Screen	5.65	2.94
	6	Projecting my image on the Big Screen	5.18	3.19
	7	Producing sound effects and music	5.50	2.87
I reacted to Ada's behavior by:	8	Moving faster on the floor	7.93	2.13
	9	Following the patterns on the floor	8.48	1.92
	10	Looking more at the big screen	4.23	2.56
	11	Making more noise	6.84	2.92
	12	Trying to imitate the behavior of visitors who seem to have Ada's attention	4.25	3.23
(Overall opinions)	13	My behavior was affected by Ada: (not at all…a lot)	6.66	2.48
	14	Interacting with Ada made me feel: (sad… happy)	7.86	1.97
	15	I felt that Ada is a kind of creature. (disagree… agree)	4.61	2.83
	16	I like Ada: (not at all…a lot)	7.08	2.43

Table 2B: Ada questionnaire text (English version) and control case responses (N = 86). 1 = most disagree, 10 = most agree (except where indicated)

2.2.2. Effects of Operating Parameters on Behavior and Attitudes

To understand the effects of Ada's output modalities on the visitors' behavior and their attitudes, tests were run in which different output components were disabled or reduced in intensity. The questionnaire responses to the cases, and the significant effects compared to the control case (t-tests, $p < 0.05$), were as follows:

Sound and music reduced to barely audible level (N = 136, invalid = 0): visitors' rating of Ada's hearing was lower (Q2, $p = 0.019$), but their rating of Ada's sound output was unchanged (Q11, $p > 0.1$). In addition, their rating of the role of Ada's eyes (Q1, $p = 0.049$) was higher, but they spent less time looking at the BigScreen (Q10, $p = 0.042$). These effects had nothing to do with what actually changed: visitors seemed to be confounding Ada's outputs (the reduced sound level) with the inputs. They seemed to blame Ada's silence on deafness rather than its being mute, while attributing increased visual processing capabilities to Ada as compensation for the perceived reduction in auditory processing capabilities.

BigScreen background plain dark blue (N = 58, invalid = 0): visitors did not score the BigScreen significantly lower (Q5 and Q10), possibly since they did not know that some of its capabilities were not being used (as opposed to the sound and music in the previous case, which were being used fully but at low volume). However, they seemed to think that Ada could not hear well (Q2, p = 0.022). This may be due to visitors relating darker, more static rooms with quietness, which was then confounded with deafness as in the reduced sound case.

No gazer images on BigScreen (N = 55, invalid = 0): in this case, visitors rated the visual output lower (Q5, p = 0.024) and also looked at the output less (Q10, p = 0.008). As in the previous case, they also seemed to think that Ada could not hear well (Q2, p = 0.004). However, their response to Q6 (the "correct" BigScreen image response) was not affected – a surprising result, since they were imagining something that never actually happened! They were also less inclined to think of themselves as imitating other visitors (Q12, p = 0.005), suggesting that observation of "successful" visitors' images/videos on the BigScreen are motivating factors for imitating the actions of those visitors.

No guide instructions (N = 60, invalid = 1): as might be expected, visitors gave generally lower ratings due to the lower level of priming they received. Two of these were significant: visitors' self-assessment of their viewing of BigScreen (Q10, p = 0.023) and their own noise-making (Q11, p = 0.003). These differences imply a reduced knowledge of how to interact with Ada.

To quantify the effect of the different test conditions on the rate of visitor-generated tile events, we define the tile event rate modulation index as follows:

$$\text{Modulation index} = \frac{1}{N}\sum_{i=1}^{N}\frac{R_c^i - R_0^i}{R_0^i}$$

N = number of behavior modes

R_c^i = mean tile event rate for test condition during behavior mode i

R_0^i = mean tile event rate for control case during behavior mode i

In all cases the changes in the operating conditions caused a highly significant reduction in the tile event rates ($p < 0.001$), with the largest decrease for the case with no guide instructions (Fig. 6). However, the rate of detection of handclaps and the spoken word "Ada" was not significantly altered. It is interesting to note that despite this change in behavior, the visitor responses to the overall questions about Ada (Q13-16) were not significantly affected by the different manipulations to the space; i.e. the degraded conditions in Ada were still good enough to elicit positive overall visitor responses.

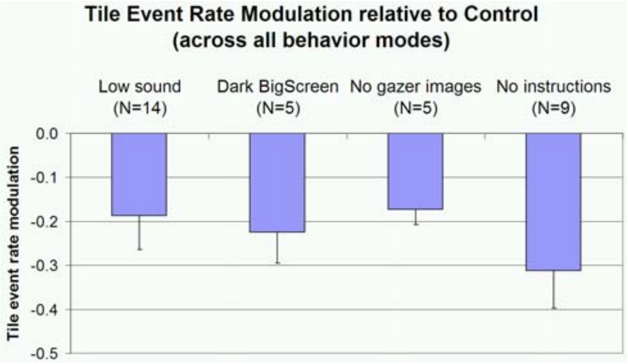


Figure 6: Comparison of tile event rate modulation effects for different experimental conditions, relative to the control case. Error bar = 1 standard deviation. N = number of visitor groups tested (mean 27 visitors per group; i.e. approx. 135-378 visitors per sample).

2.2.3. Effect of Group Size on Behavior and Attitudes

To gauge the effect of visitor density on behavior and ratings, two test cases were created with a smaller (~15 visitors/group; N = 76, invalid = 1) and larger (~32 visitors/group; N = 77, invalid = 0) number of visitors per group (normal group size = 27 visitors). In Game mode, each visitor generated tile events at about 1.3 Hz in small groups, 1.1 Hz in medium-sized groups and 1.0 Hz in large groups. The per-person tile event rates for the cases were distinguishable by group size, with larger groups tending to move slower (ANOVA, p = 0.017). However, the clap rate and spoken "Ada" detection rates were not distinguishable by group size (ANOVA, $p \geq 0.089$ & $p \geq 0.4$). The questionnaire responses confirmed the prediction that small groups would enjoy and understand Ada more: 14 out of 16 questions elicited a higher response (avg. +0.62 points) compared to the control case, while the large group size scored lower in 13 out of 16 questions compared to the control case (avg. -0.56 points). For small groups, significantly increased ratings ($p < 0.05$) were found for Ada's eyes (Q1), Ada's floor output (Q4), and the perception of Ada as an entity (Q15). The responses from the larger group revealed significantly lower ratings for two questions ($p < 0.05$), related to perceptions of Ada's hearing (Q2) and visitors' tendency to make noise (Q11).

2.3. Discussion I

None of the manipulations of Ada's operating parameters significantly affected visitors' perceptions of presence in reporting Ada to be a kind of creature. This could be because the small, uni-modal changes that were made to the operating parameters were not enough to significantly disturb the overall coherence of the space. It is possible that larger uni-modal manipulations or smaller multi-modal manipulations may have resulted in a measurable effect, but there was considerable pressure to avoid this type of experiment due to the adverse affects on the quality of the experience for the paying visitors.

However, in all cases the small manipulations did cause a significant reduction in the mean tile event rate, although this may not necessarily point to an implicit reduction in presence.

Several significant cross-modal effects were found for all operating parameter manipulation cases that did *not* correspond to the output modality that was changed. This result may indirectly support the notion that visitors treat the environment as a whole, rather than as a collection of individual input/output modalities where cross-modal effects would not necessarily be expected.

It is possible that the results observed here mean that the reductions of the output levels did not change the overall coherence of the space by very much. Other methods, such as changing the time lag of the interactions, may have been more effective in introducing severe disruptions of coherence.

A significant change in the reported presence of Ada was found for smaller group sizes. This may reflect the improved visibility of the space due to less occlusion from other visitors and/or improved visitor interactions with Ada in terms of the compliance tests. A corresponding reduction of presence was not found for the larger group, possibly because the group was not much larger than in the control case – a decision motivated, again, by considerations of the quality of the visitor experience. An inverse relationship was found between group size and movement speed as measured by tile event rate, as would be predicted by particle-based models of pedestrian motion (eg. [23]). However, no corresponding relationship was found for vocalizations of the word “Ada” or handclaps for reasons that are unclear, but could possibly be related to the tendency of the sound event detection software to saturate at around 1.5-2 Hz.

3. Experiments II: Individual Effects on Presence

Different users of a shared mixed-reality space will report different levels of presence and differing attitudes in the questionnaire described in Methods I. Two factors that may be important in determining these different responses include the demographics of the individual users and their individual levels of interaction with Ada. This leads us to postulate two hypotheses:

- User attitudes and reported levels of presence will be more positive for cases where the user has extensively interacted with Ada; i.e. when they have successfully completed the compliance testing process and seen their image projected on the screen.
- User attitudes will vary with demographics. In particular, since Ada is a space that emphasises whole-body movements, older (less physically mobile) people will have lower activity levels and thus lower questionnaire responses. Older people may also typically be less receptive to new technologies such as those used in Ada, also leading them to give lower questionnaire responses.

3.1. Methods II

The data collected for the control case was used for the analysis of the demographic effects (see Methods I). To see if individual behavioral differences were correlated with attitudes to Ada, a few subjects (2-4) in randomly selected visitor groups during normal operation were classified as most active or passive by two psychologists, and asked to complete a questionnaire (Active: collected = 60, valid = 60; Passive: collected = 58, valid = 57). Active visitors were defined as those who followed Ada’s cues in *Explore* mode and noticed their own image projected on the BigScreen. Passive visitors were those who remained largely static. All selected visitors were asked what their preferred language was (French, German or English) before being given the questionnaire in that language. Almost all of the active visitors approached agreed to complete the questionnaire; about one quarter of the passive visitors refused to participate. The demographics of the active and passive visitor pools were selected to be well balanced.

3.2. Results II

The average ratings of the active visitors were higher than those of the passive visitors in 11 out of 16 questions, with a mean overall rating difference of +0.36 points. On a question-by-question basis, only one of the modality-specific questions related to Ada’s 360° surround projection BigScreen (Q6, $p = 0.006$) elicited significantly higher responses from the active group. However, the differences were very clear for the overall questions: active visitors said that Ada influenced their behavior more (Q13, $p = 0.037$), they were happier as a result of their experience (Q14, $p < 0.001$) and they liked Ada more (Q16, $p = 0.003$). In addition, there appeared to be a trend for active visitors to report that Ada was a creature (Q15, $p = 0.087$), although this effect was not significant at the 5% level. On the other hand, passive visitors were more likely to say that they were imitating the actions of others (Q12, $p = 0.005$). Hence we can conclude that the extent of interaction with Ada is predictive of the levels of visitor attitudes and reported sense of presence.

Many demographic effects were found in the control case data. The effects were:

Visitor ratings of the floor decrease strongly with age. The rating of the floor input and output (Q3, N = 706 & Q4, N = 639, respectively) were highly age-dependent (ANOVA, $p < 0.001$ for both questions). Younger age groups rated the floor significantly higher than older age groups (post-hocs, $p < 0.05$ with Bonferroni correction), with a peak for both questions in the 16-20 year-old age group (Fig. 5).

Reported presence decreases with age. The responses to the statement that Ada is a kind of creature (Q15, Fig. 5) were found to be age-dependent (ANOVA, N = 637, $p = 0.04$). The 31-40 year old age group gave significantly lower ratings than the 10-15 year old age group (post-hoc, $p = 0.019$); the other post-hoc comparisons were not found to be significant.

Cultural differences are evident in visitor attitudes. German speakers gave lower question ratings than other language groups. In 9 out of 16 questions, people who nominated German as their first language gave significantly lower responses compared to French speakers ($p \leq 0.046$ for Q11, $p \leq 0.009$ for Q1, Q2, Q5, Q7, Q8, Q9, Q12, Q15). In Q12 German speakers also gave significantly lower responses than English speakers ($p = 0.005$). No other language-dependent effects were found.

Females report higher engagement with Ada than males. Females reported increasing their movement speed more than males (Q8, $N = 682$, $p = 0.026$), and were happier as a result of being in Ada (Q14, $N = 671$, $p = 0.014$). No other gender-specific effects were found.

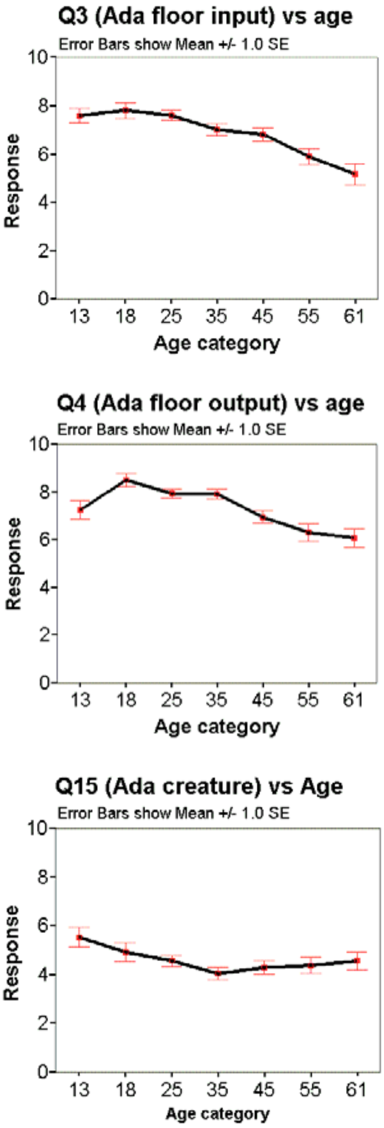


Figure 5: Age-dependence of Ada questionnaire responses. Error bars show mean ± 1 standard error.

3.3. Discussion II

The major finding in this section is the enhanced sense of presence reported by active and young visitors. Active involvement with an artefact has previously been considered as a prerequisite for enhancing a sense of presence [20]. Consistent with the *law of effect* of operant learning [24], Ada was programmed to select, test and finally engage only the most active and responsive visitors. Thus, only active visitors (covered minimum distance), with low social tameness (arrived at the center of the space) and high drive for interaction (responded to the cues on the floor) were rewarded with the "personal" interaction with Ada. The active visitors scored higher on all questions of overall attitude to Ada; they were influenced more by Ada, were happier after the experience, liked it more, and showed a trend of increased perception of Ada as a creature (i.e. higher acceptance of Ada's presence). On the other hand, passive visitors reported more imitation of the other visitors in the space. This implies a somewhat non-direct interaction with Ada, possibly explaining their lower appreciation of Ada, and lower sense of presence of Ada.

The fact that active visitors were identified by interacting with Ada means that it is possible for Ada to deduce something about the internal state of the visitor through this interaction. In other words, the compliance test can also be seen as a simple kind of "personality" test. This type of test-based interaction is important for developing interactive spaces since the results of the tests can be used as signals for allocating system resources and customising interactions to individual users.

The demographic group that stood out in appreciation of Ada as an entity consisted of youngsters. In fact, the youngest age group of children 10-15 years old showed the highest ranking on this question. This readiness to acknowledge the presence of Ada may be inherent in youngsters (perhaps similar to believing in Santa Claus) and thus they may be particularly suited to experiences in mixed-reality space. Alternatively, their enhanced sense of presence may have been triggered by particularly attractive features of Ada, such as the floor-based interactions in *Game* mode and other modes, which may have been less to the taste of older visitors. This may have been why youngsters showed significantly higher appreciation for the input and output capabilities of the floor than other age groups.

The cultural differences found in the questionnaire responses, showing that German speakers gave many lower responses than French speakers, could be due to at least two reasons. One reason could be simply that German speakers tend to give lower responses to questionnaires in general, regardless of questionnaire content. The other possible reason is that there was some sort of genuine cultural bias in the visitors' opinions of Ada. The causes of this bias, if it exists, could be related to many factors including the aesthetic presentation of Ada, the nature of the interactions, the location of the exhibit (in the French-speaking part of Switzerland), etc. However, the exact cause was unclear.

4. Implications, Applications & Limitations

In this paper we have shown that a large-scale, multi-user mixed reality space is able to engage with and sustain at least some sense of presence with a test user set taken from the general population. The reported sense of presence and interactions depend on user demographics, the number of users per unit area and the ambient conditions of the space. In addition, it is possible for the space to deduce user attitudes at a coarse level by interacting with them and observing their behavior. We also observed that user interaction with a mixed reality space leads directly to an enhanced sense of presence. It thus seems reasonable to suggest that, in order to be accessible to a large number of people with widely varying backgrounds, shared mixed reality spaces need to offer interactions at several different levels of sophistication in order to draw as many users as possible into experiencing the space. These interactions need not be extremely complex, but they should provide a graded set of interactions that are suitable for different users.

At the most basic level, the results show that it is possible to judge levels of presence in a mixed reality space which supports large numbers of simultaneous users sharing the same infrastructure. The ability to predict opinions of levels of presence using only observation of overt behavior suggests that it may be possible for an interactive space to deduce levels of presence in real-time without the need for physiological recording, provided that the system is suitably calibrated with groups of test users. Considering this result in the light of our work on learning to deploy maximally effective visitor cues [12], we suggest that it may be possible for an interactive environment to learn the conditions for achieving maximum presence and/or maximally influencing user behavior. Related work has been done on automatically estimating the interruptibility of humans in office work situations [25].

As shown in several other studies of virtual reality systems, eg. [26, 27], varying the operational parameters of Ada affected both user behavior and questionnaire responses. Virtually all of the effects seen served to reduce the level of reported presence and user activity. In addition, several cross-modal effects were seen, suggesting the inherent non-linearity of these effects on behavior and attitudes. This result implies that it is important to consider the overall effect of different strategies for maximizing presence in a virtual environment, as the combined effect of the strategies will not necessarily reflect those of the individual effects. Similarly, the number of physically present users in a given area must also be considered when designing a mixed reality space.

This study is unique in its attempt to bring presence-related applications to the general population. Due to the high cost of such a study and the need for many test subjects, it is necessarily quite limited in terms of the sophistication of the sensors and effectors than can be allocated to each individual user. The partial success in generating sensations of presence within the technological limitations of these studies conducted in 2002 opens the way for future studies, using more sophisticated sensor/effector technologies that are rapidly decreasing in price, to realise more sophisticated, personalised applications with higher levels of subjective presence.

An important question about the results shown here is their generality – will they apply to all large-scale mixed reality spaces, or are they specific to the particular configuration of Ada that we used? We suggest that the effects we have seen are general, although the details of individual effects may differ in magnitude and/or sign. To go further, we speculate that similar effects will be found in all shared spaces, whether interactive or not. Verifying or falsifying these claims will require a large database of observations to be compiled on a wide variety of shared spaces in different urban settings.

5. Future Directions

In future experiments we plan to introduce several enhancements and related investigations to improve the quality of the conclusions that we are able to draw from the data, for example:

- Physiological measures to assess presence in real-time, e.g. [28, 29].
- Investigations of whether the gender-specificity of Ada’s name and outputs affects visitor attitudes and behavior, as reported in other studies [30].
- Investigate the effect of time estimation on presence by adding a question to the questionnaire: how long were you inside Ada? This duration estimation can be correlated with reported presence [31] and demographic variables.

Upgrades to the hardware and software components of Ada will also occur, and a virtual visit component will be introduced to permit interactions between real and virtual visitors.

Acknowledgements

The other members of the Ada development team were: Andreas Bäbler, Ulysses Bernardet, Mark Blanchard, Adam Briska, Marcio Costa, Tobi Delbrück, Rodney Douglas, Klaus Hepp, David Klein, Jonatas Manzolli, Fabian Roth, Ueli Rutishauser, Klaus Wassermann, Adrian Whatley, Aaron Wittmann and Reto Wyss. The on-site exhibit guide team was led by Matthias Erzinger, Miriam Dahme and Thomas Meissner. Stefan Bucherer provided assistance with running the analyses of the floor and audio data. This work was supported by the ETH Zurich, University of Zurich, Manor AG, Expo.02, the Gebert Rüf Foundation and the Velux Foundation.

References

[1] D. Cho, J. Park, G. J. Kim, S. Hong, S. Han, and S. Lee, *The Dichotomy of Presence Elements: The Where and What*, presented at IEEE Virtual Reality 2003 (VR'03), 2003.

[2] G. S. Hubona, P. N. Wheeler, G. W. Shirah, and M. Brandt, *The Relative Contributions of Stereo, Lighting, and Background Scenes in Promoting 3D Depth*

Visualization, ACM Transactions on Computer-Human Interaction, vol. 6, pp. 214-242, 1999.

[3] A. MacWilliams, C. Sandor, M. Wagner, M. Bauer, G. Klinker, and B. Bruegge, *Herding Sheep: Live System Development for Distributed Augmented Reality*, presented at Second International Symposium on Mixed and Augmented Reality (ISMAR 2003), 2003.

[4] E.-L. Sallnäs, K. Rassmus-Gröhn, and C. Sjöström, *Supporting Presence in Collaborative Environments by Haptic Force Feedback, ACM Transactions on Computer-Human Interaction*, vol. 7, pp. 461-476, 2000.

[5] C. Basdogan, C.-H. Ho, M. A. Srinivasan, and M. Slater, *An Experimental Study on the Role of Touch in Shared Virtual Environments, ACM Transactions on Computer-Human Interaction*, vol. 7, pp. 443-460, 2000.

[6] K. Eng, D. Klein, A. Baebler, U. Bernardet, M. Blanchard, M. Costa, T. Delbruck, R. J. Douglas, K. Hepp, J. Manzolli, M. Mintz, F. Roth, U. Rutishauser, K. Wassermann, A. M. Whatley, A. Wittmann, R. Wyss, and P. F. M. J. Verschure, *Design for a brain revisited: The neuromorphic design and functionality of the interactive space Ada, Reviews in the Neurosciences*, vol. 14, pp. 145-180, 2003.

[7] K. Eng, A. Baebler, U. Bernardet, M. Blanchard, M. Costa, T. Delbruck, R. J. Douglas, K. Hepp, D. Klein, J. Manzolli, M. Mintz, F. Roth, U. Rutishauser, K. Wassermann, A. M. Whatley, A. Wittmann, R. Wyss, and P. F. M. J. Verschure, *Ada - Intelligent Space: An artificial creature for the Swiss Expo.02*, presented at IEEE/RSJ International Conference on Robotics and Automation (ICRA 2003), Taipei, Taiwan, 2003.

[8] T. Delbrück, R. J. Douglas, P. Marchal, P. F. M. J. Verschure, and A. M. Whatley, *A device for controlling a physical system*, 1999.

[9] J. Zünd, *Brainworkers*, J. Zünd, Ed. Zurich: ETH Zurich, 2002.

[10] K. C. Wassermann, J. Manzolli, K. Eng, and P. F. M. J. Verschure, *Live soundscape composition based on synthetic emotions, IEEE Multimedia*, vol. 10, pp. 82-90, 2003.

[11] G. J. Klinker, K. H. Ahlers, D. E. Breen, P.-Y. Chevalier, C. Crampton, D. S. Greer, D. Koller, A. Kramer, E. Rose, M. Tuceryan, and R. T. Whitaker, *Confluence of Computer Vision and Interactive Graphics for Augmented Reality, Presence: Teleoperations and Virtual Environments*, vol. 6, 1997.

[12] K. Eng, R. J. Douglas, and P. F. M. J. Verschure, *An Interactive Space that Learns to Influence Human Behaviour, IEEE Transactions on Systems, Man and Cybernetics Part A*, vol. 35, pp. 66-77, 2005.

[13] S. Aukstakalnis and D. Blatner, *Silicon mirage: the art and science of virtual reality*. Berkeley, CA, USA: Peachpitt Press, 1992.

[14] M. Minsky, *Telepresence, Omni*, pp. 45-51, 1980.

[15] S. Benford, C. Greenhalgh, G. Reynard, C. Brown, and B. Koleva, *Understanding and Constructing Shared Spaces with Mixed-Reality Boundaries, ACM Transactions on Computer-Human Interaction*, vol. 1998, pp. 3, 1999.

[16] *Mixed reality definition - Wikipedia*, http://en.wikipedia.org/wiki/Mixed_reality, 2005.

[17] C. Nass, J. Steuer, and E. R. Tauber, *Computers are Social Actors*, presented at Computer-Human Interaction, Boston, MA, USA, 1994.

[18] C. Nass, Y. Moon, B. Fogg, B. Reeves, and C. Dryer, *Can Computer Personalities Be Human Personalities?*, presented at Computer-Human Interaction 1995, 1995.

[19] M. Lombard and T. Ditton, *At the Heart of It All: The Concept of Presence, Journal of Computer-Mediated Communication*, vol. 3, 1997.

[20] L. Hallnäs and J. Redström, *From Use to Presence: On the Expressions and Aesthetics of Everyday Computational Things, ACM Transactions on Computer-Human Interaction*, vol. 9, pp. 106-124, 2002.

[21] J. Lessiter, J. Freeman, E. Keogh, and J. Davidoff, *A Cross-Media Presence Questionnaire: The ITC-Sense of Presence Inventory, Presence: Teleoperators and Virtual Environments*, vol. 10, pp. 282-297, 2001.

[22] K. Eng, M. Mintz, and P. F. M. J. Verschure, *Collective Human Behavior in Interactive Spaces*, presented at International Conference on Robotics and Automation (ICRA 2005), Barcelona, Spain, 2005.

[23] D. Helbing, I. Farkas, and T. Vicsek, *Simulating dynamical features of escape panic, Nature*, vol. 407, pp. 487-490, 2000.

[24] E. L. Thorndike, *Animal intelligence*, 1911.

[25] J. Fogarty, S. E. Hudson, C. G. Atkeson, D. Avrahami, J. Forlizzi, S. Kiesler, J. C. Lee, and J. Yang, *Predicting Human Interruptibility with Sensors, ACM Transactions on Computer-Human Interaction*, vol. 12, pp. 119-146, 2005.

[26] M. Slater, M. Usoh, and A. Steed, *Taking Steps: The Influence of a Walking Technique on Presence in Virtual Reality, ACM Transactions on Computer-Human Interaction*, vol. 2, pp. 201-219, 1995.

[27] C. Ware and R. Balakrishnan, *Reaching for Objects in VR Displays: Lag and Frame Rate, ACM Transactions on Computer-Human Interaction*, vol. 1, pp. 331-356, 1994.

[28] M. Meehan, B. Insko, M. Whitton, and F. P. J. Brooks, *Physiological Measures of Presence in Stressful Virtual Environments, ACM Transactions on Graphics*, vol. 21, pp. 645-652, 2002.

[29] M. Slater, A. Brogni, and A. Steed, *Physiological Responses to Breaks in Presence: A Pilot Study,*, presented at Presence 2003: The 6th Annual International Workshop on Presence, Aalborg, Denmark, 2003.

[30] C. Nass, Y. Moon, and N. Green, *Are computers gender-neutral? Gender stereotypic responses to computers, Journal of Applied Social Psychology*, vol. 2Z, pp. 864-876, 1997.

[31] W. IJsselsteijn, I. Bierhoff, and Y. Slangen-de Kort, *Duration Estimation and Presence*, presented at Presence 2001, Philadelphia, PA, USA, 2001.

Walking from thoughts: Not the muscles are crucial, but the brain waves!

Robert Leeb[1], Claudia Keinrath[1], Doron Friedman[2], Christoph Guger[3], Christa Neuper[1,4], Maia Garau[2], Angus Antley[2], Anthony Steed[2], Mel Slater[2] and Gert Pfurtscheller[1,5]

[1] Laboratory of Brain-Computer Interfaces, Institute for Computer Graphics and Vision, Graz University of Technology, Inffeldgasse 16a/II, A-8010 Graz, Austria
[2] Department of Computer Science, University College London, Gower Street, WC1E 6BT London, United Kingdom
[3] g.tec - Guger Technologies OEG, Herbersteinstrasse 60, A-8020 Graz, Austria
[4] Department of Psychology, University of Graz, Universtitaetsplatz 2, A-8010 Graz, Austria
[5] Ludwig-Boltzmann Institut für Medizinische Informatik und Neuroinformatik,
Graz University of Technology, Inffeldgasse 16a/II, A-8010 Graz, Austria
{robert.leeb@tugraz.at, keinrath@tugraz.at, d.friedman@cs.ucl.ac.uk, guger@gtec.at, christa.neuper@uni-graz.at, maia.garau@gmail.com, a.antley@cs.ucl.ac.uk, a.steed@cs.ucl.ac.uk, melslater@gmail.com, pfurtscheller@tugraz.at}

Abstract

Able-bodied participants are able to move forward in a Virtual Environment (VE) by imagining movements of their feet. This is achieved by exploiting a Brain-Computer Interface (BCI) which transforms thought-modulated EEG signals into an output signal that controls events within the VE. The experiments were carried out in an immersive projection environment, commonly referred to as a "Cave" in which participants were able to move through a virtual street by foot imagery alone. Experiments of BCI feedback on a normal monitor, VE experiments with a head-mounted display (HMD) and in the Cave-VE are compared.

Keywords — ***Virtual environment (VE), Brain-Computer Interface (BCI), walking, thoughts***

1. Introduction

“Yes he was walking! The illusion was utterly convincing …” experienced the leading actor from Arthur C. Clark in the book 3001, the final odyssey [1], when he was wearing a “Braincap” connected to the “Brainbox”. Thereby he could experience this science fiction technology and explore different virtual and ancient real worlds. Has this dream gone real? Here we show that participants are able to move forward – “to walk” – in a Virtual Environment (VE) by imagining movements of their feet.

The improvement of seamless and natural human-computer interfaces is an all-the-time necessary task in virtual reality (VR) development. An interesting research problem is to realize locomotion through a VE only by mental activity or "thought". Typically, participants navigate by using a hand-held device, such as a joystick or a wand. Unfortunately contradictory stimuli appear in such situations; on the one hand the world around them is moving, which generates the illusion of walking, but on the other hand the participant is thinking on his index finger, for pressing the button on the joystick. This results in a reduced sense of being present in the VE, and is one of the causes of simulation sickness [2].

A possible next step towards next-generation interfaces could be achieved by exploiting a Brain-Computer Interface (BCI) which represents a direct connection between the human brain and the computer [3]. The electroencephalogram (EEG) of the human brain encompasses different types of oscillatory activities, in which the oscillations in the alpha and beta band (event-related desynchronization, ERD [4]) are particularly important to discriminate between different brain states (e.g. imagination of movements). A BCI transforms thought-modulated EEG signals into an output signal [3] that can control events within that VE [5, 6].

The goal of this work is to demonstrate that it is possible to move through different VEs, e.g. a virtual street, without any muscular activity, when the participant only imagines the movement of both feet and to show the influences of different feedback modalities on the same task.

VR provides an excellent testing ground for procedures that may apply later in reality. One important future application may the use of VE for people with disabilities. If it is possible to show that people can learn to control their movements through space within a VE, it would justify the much bigger expense of building physical devices as e.g. a robot arm controlled by a BCI.

2. Methods

2.1. Graz Brain-Computer Interface

Direct Brain-Computer communication is a novel approach to develop an additional communication channel for human-machine interaction. The imagination of different types of movements, e.g. right hand, left hand, foot or tongue movement, results in a characteristic change of the EEG over the sensorimotor cortex of a participant [4].

The Graz-BCI detects changes in the ongoing EEG during the imagination of hand or foot movements and transforms them into a control signal [7]. Three bipolar derivations, located 2.5 cm anterior and posterior to the electrode positions C3, Cz and C4 of the international 10/20 system [8] were recorded with a sampling frequency of 250 Hz (sensitivity was set to 50µV) and bandpass filtered between 0.5 and 30 Hz. The ground electrode was positioned on the forehead.

The logarithmic bandpower (BP) was calculated for each channel by digitally band-pass filtering the EEG (using a Butterworth filter of order 5) in the upper alpha (10 - 12 Hz) and beta band (16 - 24 Hz), squaring the signal and averaging the samples over a 1-s epoch. The resulting 4 BP features were transformed with Fishers linear discriminant analysis (LDA) [9] into a control signal. Finally the computed control signal was used to control / modify the feedback (FB) and either visualized on the same PC as a bar (see Figure 1a) or sent to the VE as a steering input inside a virtual world (see Figure 1b and 1c) [5].

The complete biosignal analysis system consisted of an EEG amplifier (g.tec, Graz, Austria), a data acquisition card (National Instruments Corporation, Austin, USA) and a recording device running under WindowsXP (Microsoft Corporation, Redmond, USA) on a commercial desktop PC [10]. The BCI algorithms were implemented in MATLAB 6.5 and Simulink 5.0 (The MathWorks, Inc., Natick, USA) using rtsBCI [11] and the open source package BIOSIG [12].

Detailed information about the physiological background of motor imagery and ERD can be found elsewhere [4, 13], also about signal processing, feature extraction and the Graz-BCI [7, 10] and generally about various BCI systems [3, 14].

2.2. Participants and experimental paradigm

Three healthy participants (between 23 and 30 years) took part in these experiments over 5 months. All were right handed and without a history of neurological disease and gave informal consent to participate in the study.

In the first step a number of training runs (TR) were performed with each subject. These data were used to setup a classifier, which can be used in the next step for providing a feedback (FB) to the subject. The visual FB informs the participant about the accuracy of the classification during each imagery task.

The performances of three different FB conditions are compared: first the results of the standard BCI bar-FB with a simple bar (see Figure 1a), secondly using a head mounted display (HMD) as FB device (see Figure 1b) and finally using a highly immersive “Cave” projection environment (see Figure 1c).

Each feedback condition was measured multiple times (called sessions) and the order of recording was condition bar, HMD, Cave, HMD, bar. Figure 3 displays which type of FB has been used in each run and session, respectively. In each session 4 runs have been performed, whereby each run consisted of 40 trials (20 foot and 20 right-hand cues, in random order) based on the standard Graz-BCI paradigm [7]. Each trial lasts about 8 second and between the trials was a randomized interval in the range from 0.5 to 2 seconds. The data of the standard BCI run was used to compute a LDA classifier and the error rates were estimated by a 10 times 10-fold cross-validation LDA-training. The calculated classifier with the best classification accuracy during the imagination period (between second 4.5 and 8, in 0.5 s intervals) was selected for further use in all feedback runs. Further details of BCI training with motor imagery can be found elsewhere [7].

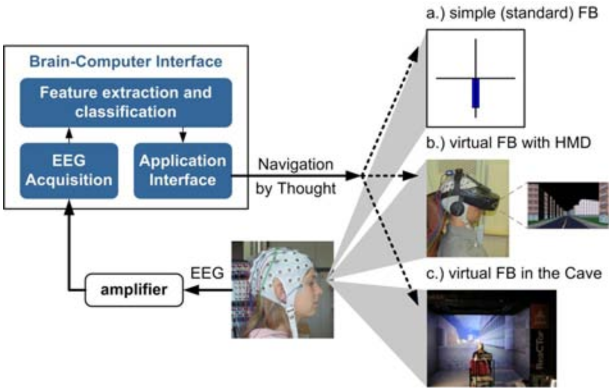


Figure 1: Schematic model of the used BCI-VR system with the participant wearing the electrode cap. Three different visual feedback modalities are displayed: (a) standard feedback whereby a vertical bar is controlled by the BCI output. (b) The participant is wearing a HMD. A screenshot of the virtual environment as seen by the participant is displayed at the far right. (c) Picture of one participant during the experiment in a Cave-like system. The surrounded projected environment creates the illusion of being in a virtual street. (b,c) Navigation through the VE is controlled by the output of the BCI.

2.3. Simple standard BCI feedback

In each run the participant had to imagine feet or right hand movement in response to a visual cue-stimulus presented on a computer monitor, in the form of an arrow pointing downwards or to the right, respectively. In addition to the visual cue an auditory cue stimulus was also given either as a single beep (hand imagery) or as double beeps (feet imagery). A visual feedback in the form of a moving bar (see Figure 1a) was given to inform the participant about the accuracy of the classification during each imagery task (i.e. classification of right hand imagery was represented by the bar moving to the right, classification of foot movement imagery made the bar moving downward).

2.4. Virtual feedback with a HMD

Virtual reality FB was presented with VRjuggler [15] and a Virtual Research V8 HMD (Virtual Research Systems, Inc., Aptos, USA) driven by an ATI Radeon 9700 graphics card (ATI Technologies, Inc., Markham, Canada). The given task of the participant was to walk to the end of the street inside this virtual city, whereby any time the computer identified the participant's brain pattern as a foot movement a motion happened (see Figure 1b). The same BCI paradigm as in the condition above (section 2.3) was applied, only the cue was given just acoustically. Correct classification of feet motor imagery was accompanied by moving forward with constant speed in the projected virtual street and the motion was stopped on correct classification of hand motor imagery (see Table 1). Incorrect classification of foot motor imagery resulted as well in halting, and incorrect classification of hand motor imagery in backward motion [16]. The walking distance was scored as a "cumulative achieved mileage" (CAM), which is the accumulated forward distance covered during feet movement imagination and is used as a performance measurement.

		subject imagined	
		foot movement	hand movement
Cue class	foot movement	forward	stop
	hand movement	backward	stop

Table 1: Dependency between the predetermined cue classes and the movements imagined by the subject and their resulting motions performed in the virtual street.

2.5. Virtual feedback in the Cave

Two sessions were performed in London in a multi-projection based stereo and head-tracked VE system commonly known as a "Cave" [17]. The particular VE system used was a ReaCTor (SEOS Ltd.¸ West Sussex, UK) which surrounds the user with three back-projected active stereo screens (3 walls) and a front projected screen on the floor (see Figure 1c). Left- and right-eye images are alternately displayed at 45Hz each, and synchronized with CrystalEye™ stereo glasses. A special feature of any VE system is that the images on the adjacent walls are seamlessly joined together, so that participants do not see the physical corners but the continuous virtual world that is projected with active stereo [18]. The application implemented in DIVE [19] was a virtual main street with various shops on both sides (see Figure 2). Some of the shops could theoretically be visited but in this experiment the task was to go only straight forward as far as possible. The street was populated with some virtual characters that walked along the street, whereby the characters were programmed to avoid collisions with the participant. The communication between the BCI and the VR was done via the Virtual Reality Peripheral Network (VRPN, [20]).

Figure 2: Participant in the virtual main street with shops and animated avatars during the Cave-FB. The subject wears an electrode cap (connected to the amplifier) and shutter glasses.

3. Results

All participants were able to navigate in the different VE's and the achieved BCI performance in the VR tasks was comparable to standard BCI recordings. The usage of VR as FB was stimulating the participant's performances. Especially in the Cave condition (highest immersion) the performance of 2 participants was excellent (up to 100% BCI classification accuracy of single trials), although variability in the classification results between individual runs occurred (see Figure 3 and 7).

All runs performed consecutively on one day are called one session and most of the time one session contains four runs. In Figure 3 all performed runs over a period of 5 month with simple standard bar-FB, HMD-FB and Cave-FB and the trainings runs without FB (TR) are indicated in each subject. All runs following the indicated date are performed at this day. Each run consisted of 40 trials, 20 trials with a cue for foot imagery and 20 for right hand imagery in randomized order. The duration of a trial is 8 seconds (a random pause of 0.5 to 2 seconds is added between the trials to avoid adaptation), therefore a run lasted approximately 6.5 minutes and one session lasted about 1 hour including the time electrode montage.

Concerning the difference between the various feedback modalities no statistical evaluation of the data was possible, because only three individuals participated in these experiments.

The results are split into two parts: on the one hand the classification accuracy of the BCI is interesting to study the influence of the different FBs on the participants and on the other hand the task performances.

3.1. BCI classification

The BCI classification error is a measure how good the two brain states could be identified in each run. A classification error of 0 % denotes a perfect separation between the two mental tasks (20 examples for right hand movement imagination and 20 examples for foot movement imagination). A random classification would result in a classification error of 50 %. The error varies over the time of the trial (see Figure 4, the exemplarily used runs are indicated in Figure 3 with a black diamond). At second 3 the participant heard the cue (single or double beep) and started to imagine the desired movement. The optimal performance varies over the measurements and between individuals, but is typically at least two seconds after the trigger [21], see Figure 4 for the BCI classification of each participant of one run during the Cave experiments. Especially participant P3 could achieve a long and stable brain pattern over nearly the whole FB time (last row in Figure 4), which directly corresponds to very good CAM in Figure 5.

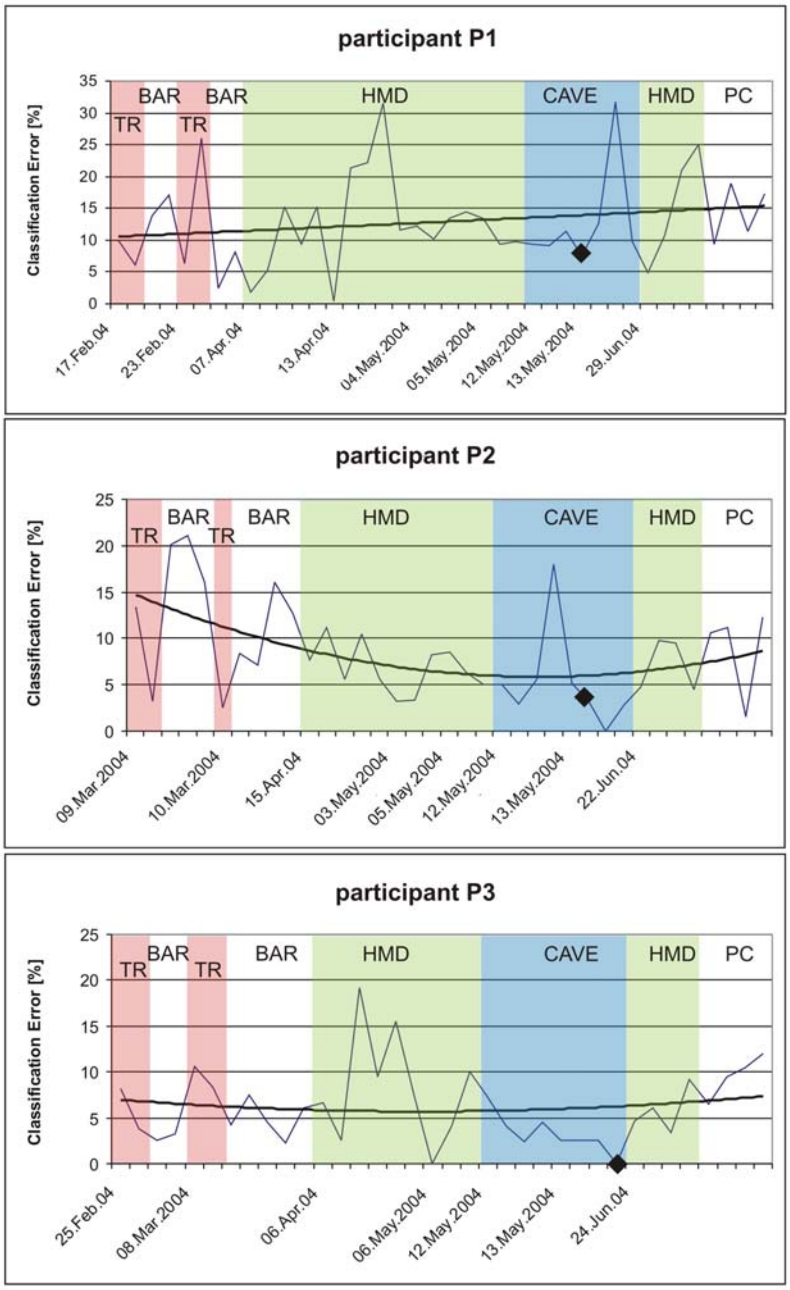


Figure 3: Classification error (in %) for all runs of the 3 participants. Runs with BAR-FB, HMD-FB and Cave-FB and the trainings rungs without FB (TR) are indicated in each subject. An interpolation of 2nd order shows the trend of the classification error over the time (black line). More than one run has been performed on each day, therefore all data points following the indicated date are performed at this day. The runs marked with a black diamond ◆ (one in each subject) are analyzed in detail in Fig. 3 (classification error) and in Fig. 5 (CAM, task performance).

The results of all runs with FB over a period of 5 months are displayed in Figure 3. Separately indicated are the runs with bar-FB, HMD-FB and Cave-FB. An interpolation of 2nd order has been performed to show the trend of the classification error over the time (thick black line). The time-courses of the classification error of the individual participants, on the one side, fluctuate considerably over runs and, on the other side, display different trends in the 3 participants: in participant P1 the classification error shows a slightly increasing trend over runs, in participant P2 a minimum during the Cave experiments and in participant P3 a relative constant level.

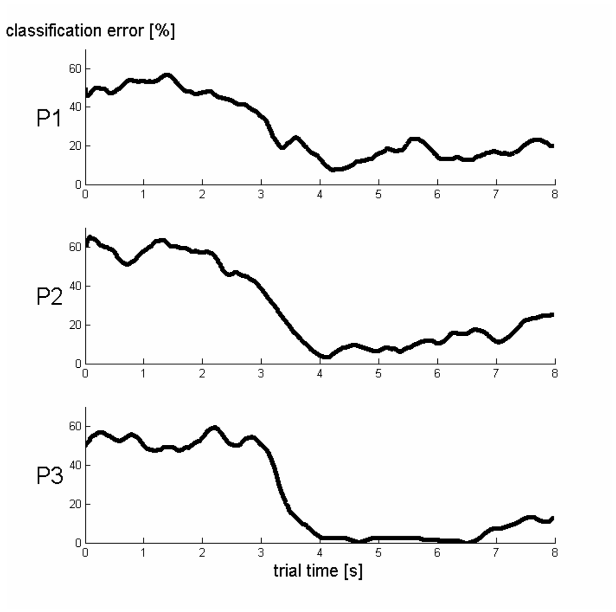


Figure 4: Mean classification error (in %) of one run (marked with a black diamond in Figure 3) over the trial time of all 3 participants. At second 3 the participant heard the cue (single or double beep) and started to imagine the specified movement during the FB period (between second 4.25 and 8).

3.2. Task performance

Some single run results of the first session with the Cave-FB obtained for the 3 participants are exemplary displayed in Figure 5 (this runs are indicated in Figure 3 with a black diamond and are the same runs as displayed in Figure 4). Both the theoretically possible CAM is plotted in dashed and the real achieved CAM as a full line. Because each participant had a different sequence of the 20 foot (F) and 20 right hand (R) motor imageries which were randomly distributed to avoid adaptation, the theoretical pathways are different in all pictures. Nevertheless the numbers of trials for both classes are the same and therefore the maximum possible CAM is the same. Participant P3 achieved the best performance with a CAM of 85.4 %. A CAM of 100 % corresponds to a correct classification of all 40 imagery tasks over the entire feedback time. A random classification would result in a CAM of 0 %. For comparison reasons the CAM performances of the bar-FB experiments have been simulated offline.

In Figure 6 the mean achieved CAM of all participants and condition is plotted. The trend of each participant over the FB conditions is plotted as grey dashed line. Figure 7 displays a detailed analysis of the same data. Each box plot has lines at the lower quartile, median, and upper quartile values. The whiskers are lines extending from each end of the box to show the extent of the rest of the performances. The trend of each participant over the three FB conditions is indicated with a grey dashed line. Two participants' show an increase over the condition, but participant P1 achieved worse results with the HMD.

It is nearly impossible to achieve the maximum gain able CAM of 100%, because every small procrastination or hesitation of the participant results in reduced mileage. For a perfect outcome, a correct classification must happen during the whole FB time of all trials. Therefore the results are not directly comparable to normal BCI performance results.

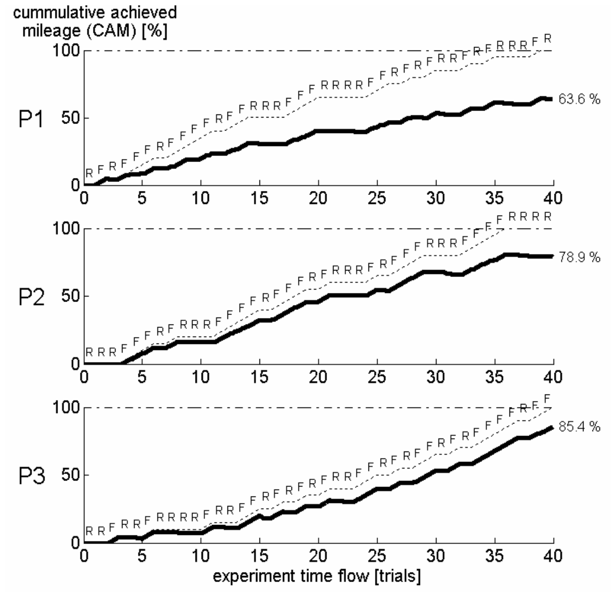


Figure 5: Task performance measures of all 3 participants (P1, P2 and P3) displayed in the theoretical possibility CAM (dashed line) and the real CAM (full line).

3.3. Presence and body representation

After completing the experiments in the Cave, the participants were asked to fill in the Slater-Usoh-Steed presence questionnaire [22] and then a non-structured interview was conducted. The results of the questionnaire and interview data have been evaluated separately [23]. After the standard BCI experiments and after the HMD experiments no presence questionnaires and interviews have been conducted. As a result of that no comparable analysis can be done over the three FB conditions and therefore this topic can not be discussed further in this paper, nevertheless the BCI may be considered as a very unusual extension of the body.

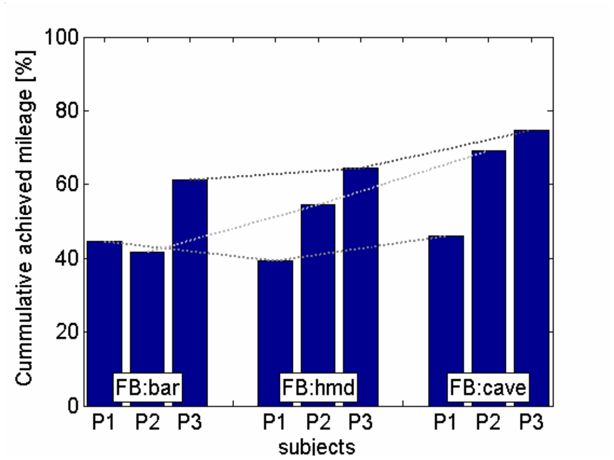


Figure 6: Mean CAM values of all participants and all 3 FB conditions. The trend of each participant over the FB conditions is plotted as grey dashed line.

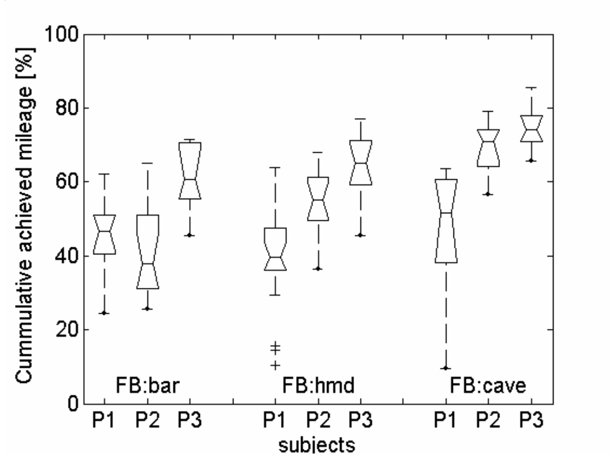


Figure 7: Distribution of the achieved CAM of all participants and all 3 FB conditions. Each plot has lines at the lower quartile, median, and upper quartile values. The whiskers are lines extending from each end of the box to show the extend of the rest of the data.

4. Discussion and conclusion

These data indicate that EEG recording and single trial processing are possible in a HMD or a Cave-like system, and that feet motor imagery is an adequate neural strategy to control events within the VEs. Imagination of both feet movement is a mental task which comes very close to that of natural walking. The next important step in this research is to change the experimental paradigm to eliminate externally-paced cues. In this way the participant could decide to start walking at will. Such an asynchronous BCI system however, is more demanding and more complex for the participant [24].

The participants were able to achieve a grand average CAM of 49.2%. The result of a random session would be a CAM of 0%. Relative good performances are obtained with the virtual FB's (Cave better than HMD), except some outliers. One reason for some inferior classification results of individual runs especially in the Cave condition in Figure 7, e.g. CAM of 9.5 in participant P3) could be the loss of concentration in connection with a moving visual scene, because observation of moving objects can have an impact on neurons in the motor area [25]. Another possible explanation for the problems in the performance results of participant P1 (top row in Figure 5) could be that between trial 14 and 17 and between trial 20 and 25, the same class always should have been performed, that is the "standing class" (right hand movement) in this example, but the participant wasn't able to remain stationary for such a long period. A similar effect can be observed at the end of the run plotted in the middle row of Figure 5. Perhaps a faster alternation between the two classes would achieve better results, but the sequence of cues was randomized automatically through each run. The problem of this long period of "standing" is that during this time no feedback is given to the participant. If the correct movement (right hand motor imagery) is imagined, the participant remains stationary, but if the wrong movement (foot motor imagery) is imagined, then the participant walks backwards. Walking backwards is visual feedback, in contrast to remaining stationary, so the period of giving no information back to the participant is broken. It can also be observed that the way which was walked backwards isn't that steep and long as the path forward.

The task performances (see Figure 6 and 7) and the BCI classifications (see Figure 3) achieved the best values during the Cave-FB. The argument that only the task experience triggered this result can be disproved, because the conditions were recorded in another sequence and unfortunately the classification error increased in participant P1 over the time (see Figure 4), which would be contradictory to that argument. Whether a VE or an immersive VE as feedback has an impact on the performance or can shorten the training time needs further investigation. The number of participants is too small to allow statistical analysis, but the results are consistent. All subjects reported that the Cave was more comfortable than the HMD and both were very much preferred over the BCI training on a monitor.

In principle should it be possible to achieve the same performances in both VE conditions, the HMD and Cave. The limited field of view (FOV) of the HMD and the weight on the head was irritating and bothering. Also the optical resolution of the HMD was less than in the Cave. Therefore the subjects felt less present with the HMD as in the Cave. The Cave was compared to the HMD as a VE-FB much more natural and is hence preferable.

The main reason given for preferring the VR was that it provided motivation. The street was treated as a sort of race course and every subject wanted to get further as the others in the previous sessions. The motivation seems to greatly improve BCI performance, but too much excitement might have a negative impact, as it makes it harder to concentrate on the BCI control. Two subjects had sometimes nearly perfect runs till the last 2 or 3 trials of the run. At that time they already realized that they could achieved a new distance record, but this excitement reduced their concentration and therefore the last trials were performed badly, which reduced the task performance insomuch that no new record could be achieved. The aspect of motivation and the task/goal of the subject during the experiment have a great influence on the BCI performance and must be taken into consideration in all further BCI experiments.

VR provides an excellent training and testing ground for procedures that may apply later in reality. One important application may the use of VE for people with disabilities. If it can be shown that within VE people can learn to control their movements through space, than this justifies the much greater expense of building physical devices (e.g. neuro-prosthesis or a robotic arm) that are controlled by a BCI. Another application of the combined BCI and VR is the use of the VE with the goal to will enhance the classification accuracy and shorten the time needed for BCI trainings session. Feedback presentation by using VR is very powerful and may improve the biofeedback therapy as e.g. to reinforce the rehabilitation in stroke patients.

The research reported in this paper is a further step to the long-range vision for multi-sensory environments exploiting only mental activity. EEG-based BCI systems have a bad signal-to-noise ratio and display a drop of classification accuracy when more than 2 mental states have to be classified [3, 24, 26]. The ultimate idea behind is to use direct implants into the brain (for completely paralyzed patients) for computer control, as discussed recently by Nicolelis [27] and analyzed directly the activity of single neurons. In this case the signal-to-noise ration and more than 2 mental states can be classified with high accuracy.

Maybe the vision of the science fiction authors to use the brain as the ultimate interface will become reality sometime in the future.

Acknowledgements

The work was funded by the European PRESENCIA project (IST-2001-37927). Thanks to Vinoba Vinayagamoorthy and Marco Gillies for providing the street scene virtual environment and special thanks to Jörg Biedermann, Gernot Supp and Doris Zimmermann for their participation.

References:

[1] A. C. Clark, *3001, The Final Odyssey*. London: HarperCollinsPublisher, 1997.

[2] M. Slater, M. Usoh, and A. Steed, "Taking steps: the influence of a walking technique on presence in virtual reality," *ACM Trans. Comput.-Hum. Interact.*, vol. 2, pp. 201-219, 1995.

[3] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control," *Clin Neurophysiol*, vol. 113, pp. 767-91, 2002.

[4] G. Pfurtscheller and C. Neuper, "Motor imagery and direct brain-computer communication," *Proceedings of the IEEE*, vol. 89, pp. 1123-1134, 2001.

[5] R. Leeb, R. Scherer, C. Keinrath, C. Guger, and G. Pfurtscheller, "Exploring virtual environments with an EEG-based BCI through motor imagery," *Biomed Tech (Berl)*, vol. 50, pp. 86-91, 2005.

[6] J. D. Bayliss, "Use of the evoked potential P3 component for control in a virtual apartment," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on [see also IEEE Trans. on Rehabilitation Engineering]*, vol. 11, pp. 113-116, 2003.

[7] G. Pfurtscheller, C. Neuper, G. R. Muller, B. Obermaier, G. Krausz, A. Schlogl, R. Scherer, B. Graimann, C. Keinrath, D. Skliris, M. Wortz, G. Supp, and C. Schrank, "Graz-BCI: state of the art and clinical applications," *IEEE Trans Neural Syst Rehabil Eng*, vol. 11, pp. 177-80, 2003.

[8] H. H. Jasper, "The ten-twenty electrode system of the International Federation," *Electroencephalogr Clin Neurophysiol*, vol. 10, pp. 370-375, 1958.

[9] C. M. Bishop, *Neural Networks for Pattern Recognition*. Oxford, UK: Clarendon Press, 1995.

[10] C. Guger, A. Schlogl, C. Neuper, D. Walterspacher, T. Strein, and G. Pfurtscheller, "Rapid prototyping of an EEG-based brain-computer interface (BCI)," *IEEE Trans Neural Syst Rehabil Eng*, vol. 9, pp. 49-58, 2001.

[11] R. Scherer, A. Schlögl, G. R. Müller-Putz, and G. Pfurtscheller, "Inside the Graz BCI: rtsBCI," *Proc. 2nd International Brain-Computer Interface Workshop and Training Course, Graz, Austria, Biomed Tech (Berl)*, vol. 49 (Suppl.Vol.1), pp. 81-82, 2004.

[12] A. Schlögl, "The BIOSIG-Project," in *http://biosig.sf.net/*, 2003-2005.

[13] G. Pfurtscheller and F. H. Lopes da Silva, "Event-related EEG/MEG synchronization and desynchronization: basic principles," *Clin Neurophysiol*, vol. 110, pp. 1842-57, 1999.

[14] T. M. Vaughan, W. J. Heetderks, L. J. Trejo, W. Z. Rymer, M. Weinrich, M. M. Moore, A. Kubler, B. H. Dobkin, N. Birbaumer, E. Donchin, E. W. Wolpaw, and J. R. Wolpaw, "Brain-computer interface technology: a review of the Second International Meeting," *IEEE Trans Neural Syst Rehabil Eng*, vol. 11, pp. 94-109, 2003.

[15] VRjuggler, "Open Source Virtual Reality Tool," in *http://www.vrjuggler.org/*.

[16] R. Leeb and G. Pfurtscheller, "Walking through a Virtual City by Thought," *Proc. of the 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society - EMBC 2004, San Francisco, USA*, pp. 4503-6, 2004.

[17] C. Cruz-Neira, D. J. Sandin, and T. A. DeFanti, "Surround-screen projection-based virtual reality: the design and implementation of the CAVE," in *Proceedings of the 20th annual conference on Computer graphics and interactive techniques*: ACM Press, 1993, pp. 135-142.

[18] M. Slater, A. Steed, and Y. Chrysanthou, *Computer Graphics and Virtual Environments: From Realism to Real-Time*. Harlow, UK: Addison-Wesley, 2002.

[19] E. Frecon, G. Smith, A. Steed, M. Stenius, and O. Stahl, "An overview of the COVEN platform," *Presence-Teleoperators and Virtual Environments*, vol. 10, pp. 109-127, 2001.

[20] VRPN, "Virtual Reality Peripheral Network," in *http://www.cs.unc.edu/Research/vrpn/*.

[21] G. Krausz, R. Scherer, G. Korisek, and G. Pfurtscheller, "Critical decision-speed and information transfer in the

"Graz Brain-Computer Interface"," *Appl Psychophysiol Biofeedback*, vol. 28, pp. 233-40, 2003.

[22] M. Slater and M. Usoh, "Presence in Immersive Virtual Environments," in *Proc. IEEE Virtual Reality Annual International Symposium, Seattle, WA*, 1993, pp. 90-96.

[23] D. Friedman, R. Leeb, A. Antley, M. Garau, C. Guger, C. Keinrath, A. Steed, G. Pfurtscheller, and M. Slater, "Navigating Virtual Reality by Thought: First Steps," *Proc. 7th Annual Int. Workshop PRESENCE, Valencia*, pp. 160-167, 2004.

[24] R. Scherer, G. R. Muller, C. Neuper, B. Graimann, and G. Pfurtscheller, "An asynchronously controlled EEG-based virtual keyboard: improvement of the spelling rate," *IEEE Trans Biomed Eng*, vol. 51, pp. 979-84, 2004.

[25] G. Rizzolatti, L. Fogassi, and V. Gallese, "Neurophysiological mechanisms underlying the understanding and imitation of action," *Nat Rev Neurosci*, vol. 2, pp. 661-70, 2001.

[26] G. Pfurtscheller, C. Neuper, and N. Birbaumer, "Human Brain-Computer Interface," in *Motor cortex in voluntary movements: a distributed system for distributed functions. Series: Methods and New Frontiers in Neuroscience*, E. Vaadia and A. Riehle, Eds.: CRC Press, 2005, pp. 367-401.

[27] M. A. Nicolelis, "Actions from thoughts," *Nature*, vol. 409, pp. 403-7, 2001.

Reliving VE Day With Schemata Activation

Phil Turner[1], Susan Turner[1] and Dimitrios Tzovaras[2]
[1]Social Informatics Research Group, School of Computing, Napier University, Edinburgh
[2]Informatics and Telematics Institute, CERTH, Thermi-Thessaloniki, Greece

Abstract

This paper reports some preliminary work on the IntoMyWorld candidate Presence II project. One of the key components of IntoMyWorld is a mixed reality 'album' of important events. The album will contain entries which will allow people, for example, to re-immerse themselves in their own weddings or other significant events. Among our first tasks are (a) to understand the salient features and dimensions of events which must be captured and (b) identify the range of environmental cues required to trigger memories and re-immersion. Our psychological position has been draw upon contemporary thinking on autobiographical memory and add to this aspects of schema theory. Schema theory claims that memories are encoded and recalled in structured packets that have 'slots' which can store either default or individual values. While psychologists have studied schemata and have experimentally manipulated their formation and recall, we propose to activate them using technology. The IntoMyWorld system will make use of situated, interactive schemata to help trigger memories and with this a sense of being present once more at specific events and places. In this initial work we consider some of the issues in re-creating a significant event in the lives of our parents and grandparents, namely VE day – May 1945.

1. Introduction

IntoMyWorld is a candidate Presence II project which has been developed by a consortium of universities, research labs and industrial partners across Europe led by Centre for Research and Technology Hellas. As currently formulated, IntoMyWorld comprises three paradigmatic development streams, the first of which is *My Life Album*, an intelligent album that will selectively record, store and replay important events of our everyday life. The second is the *My Life Learning Notebook* which is an advanced notebook containing procedural and propositional knowledge acquired from practice and finally, *My Memory Aid* which will be of particular value to people experiencing memory problems and other forms of cognitive deficit. IntoMyWorld is an interactive system which aims to support recall, recognition and immersion in memories by supplying a cue-rich environment. The focus of this paper is a discussion of some of the psychological issues involved in constructing an entry in *My Life Album* and how this entry might then be used to immerse an individual in this memory. Among the major psychological research which is required to underpin the development of the IntoMyWorld concept includes:

- understanding 'everyday presence', that is, identifying the salient features of the environment – people, events, places – which are central to the experience of being present. Having identified these features, the technologists in IntoMyWorld will then be tasked to 'record' and 'replay' them.

- understanding the relationship between memory and presence. The IntoMyWorld technology cannot hope to faithfully replay all aspects of the occasion: instead the mixed reality system will present an evocative chiaroscuro of the event.

2. VE Day

This year being the sixtieth anniversary of the end of the Second World War, we have decided to consider what would be required to re-create the experience of VE day. As a first step in the development of the IntoMyWorld system we need to understand what people remember of a significant event such as this. While there are a number of psychological studies of remote memories and reminiscences (e.g. [1], [2], [3], [4], [5]), none have considered them from the perspective of providing a technological memory prosthesis.

VE day – Victory in Europe – was 8th May 1945. In Britain it was marked by street parties and wild celebration. We have chosen this event not only because of its significance but as there are abundant first-hand accounts of the day readily available on the Web. The results of an analysis of several sets of archived accounts are presented in section 6. There is also a significant historical archive of the event.

By a recreation of the experience of VE day we intend considering what would be required to re-immerse someone (who was there) in the sights and sounds of the event. (This is not to suggest that other people might not be able to engage in this sort of 'time travel' but this is outwith the scope of the current discussion.) The question then is, what are the environmental cues which might serve to trigger autobiographical memories of the event and together re-create the sense of being there. From this description it should be clear that we have not adopted the traditional approach to memory. For many psychologists, memory is a purely cognitive faculty but studied in the laboratory. For IntoMyWorld, memory is taken to be the everyday process of reconstruction which relies upon environmental, situated cues.

3. Memory and Presence

Memory must have a role in presence but there has been little or no empirical research conducted into this to date. A significant exception to this is Riva *et al.*'s (2004) three-layer model of presence in which memory plays an important role [6]. The three layers of the model are:

- *proto presence* – the embodied aspect of presence relating to the differentiation of the self from the world;
- *core presence* – a process of selective attention to perceptual stimuli, supporting the discrimination of external reality from the contents of one's consciousness, dreams or memories;
- *extended presence* - which serves to assess the relationship and significance of events in the world in the context of the memories and so forth which make up the autobiographical self.

It is evident that they see a role for memory in two of the three layers of presence. It is also worth quoting at length the scenario used to illustrate how the model works.

"To understand how these components are related we can use an example: the way our self experiences our first view of the Colosseum in Rome. We receive sensory signals from our eyes, ears, nose and sense of touch that are mapped by the proto self—the feeling of something happening. […] this leads to perceptual activity which is monitored by the core self and becomes the content of core consciousness […] Some milliseconds later, it adds dispositional records of that place (or similar places), records which typically include stored sensory, motor response and emotional data. If these records are also part of autobiographical memory — the organized record of the main aspects of our biographies we may consciously recognize the place because we studied it in architectural history; and we may have emotional ties because we associate the place with special memories […]. The result is a single conscious experience integrating perceptions, emotions and feeling. Once the event has ended, it is restored in dispositional space with new data about our most recent experience."

This 'thought experiment' clearly illustrates the intimacy of memory and presence. In reading this it is immediately evident that memory is a major substrate upon which presence 'resides'.

4. Our Epistemological Orientation

Despite the (above) discussion of the three-layer model of presence we do not subscribe to Riva *et al.*'s description of autobiographical memory as simply "the organized record of the main aspects of our biographies" (ibid: 408). Instead we have adopted an epistemological position similar to that of Clark and others who argued for a situated, interactionist account of complex cognitive and affective phenomena (e.g. [7] [8], [9], [10]).

Gero and Peng (2004: 3) define situated-ness as "where you are when you do what you do matters". They go on, "[It] states that an agent's knowledge depends on the conext in which it is situated. What can be cognized is also related to agent's experiences which are grounded from memory constructed though agent-environment interactions". Gero also regarded memory as a constructive process – constructed from the experiential responses to environmental cues, the activated memory and past experiences.

Clancey (1991: 91) writing from an artificial intelligence / cognitive science perspective, describes situated cognition as "the study of how representations are created and given meaning. An essential idea is that this process is perceptual and inherently dialectic. That is, the organization of mental processes producing coherent sequences of activity and the organization of representational forms … arise together". He continues in the same vein noting that Bartlett – a pioneer in memory research - observed that "mental organizations do not merely drive activity like stored programs, but are created in the course of the activity, always as new, living structures" [11]. The situated perspective should be seen as a continuum ranging from the very radical formulations which reject computation and representation (e.g. [12] [13]) to those which recognise that complex human behaviour cannot be described in a context-free manner ([14] [15]). However, situated cognition generally does recognise that such behaviour cannot simply be attributed to pre-existing internally in neural structures **or** features of the world *per se* but from an interaction between the two.

We argue for an interaction between internal representation and external, situated cues, consistent with the common experience that an environmental cue can evoke a memory which in turn can conjure an entire scene in which we can feel immersed, involved and present. Having made a case for a situated, interactionist account, we need to discuss the nature of the internal representation. this brings us to a review of the various aspects of memory.

5. A Proliferation Of Memories

There is a profusion of terms, metaphors, models and methodologies with respect to memory. After reviewing a number of candidate models we have adopted a *schematic* account of *autobiographical memory*. While we are presenting a psychological perspective on memory and presence, we must equally retain an engineering perspective too. Clearly, recalling personal memories, by definition, involves autobiographical memory and there is abundant evidence that such memories are recalled as stories and these stories have a schematic structure ([16] [17]). We begin, however, with a discussion of autobiographical memory.

5.1 Autobiographical and Episodic Memories

Tulving [18] was the first to propose two different forms of long term memory, namely episodic and semantic memory. Semantic memory hold general knowledge, facts about the world. This is not seen to be of personal significance. In contrast, episodic memory consists of a record of personal experiences. While Tulving initially considered episodic memory to be synonymous with autobiographic memory, his position was eventually modified to recognize that autobiographical memory is a special kind of episodic memory concerned with life events. Moving beyond these early studies, we have found the Conway's work on autobiographical memory, episodic memory, mental models and their inter-relationships both compelling and convincing [19] [20]. Conway has argued that episodic memories comprise highly detailed sensory perceptual knowledge of recent experiences. These memories are typically retained for intervals of minutes to hours. They might also be thought of as a sort of 'sample' of past experiences and this sample is sparse compared with the number formed. These episodic memories may then be integrated with the autobiographical memory 'knowledge base'. Conway conceives of autobiographical memory as an account of our memories from the perspective of personal goals. As he puts it "a central tenet of this account is that a fundamental function of human memory is to retain knowledge on the progress of personal goals, i.e. whether they have been achieved or not" (Conway, 2001: 1375). Conway's model of autobiographical memory has potentially some interesting consequences for presence research. He describes our "mental model of the current situation" which might correspond to our sense of presence as comprising representations of the current situation in the episodic buffer *and* patterns of activation over knowledge structures in semantic memory *and* activation of the goal hierarchy of the working self *and* the affective state. Unfortunately space precludes a more detailed discussion of this model.

5.3 Schematic memory

We noted above that people recall events from their autobiographical memory as stories and these stories have a schematic structure. Schema theory can be traced back to the seminal work of Bartlett [11]. Bartlett introduced the notion of schemata in order to explain how it is that when people remember stories. He found, for example, that we typically omit some details and introduce rationalizations, reconstructing the story so as to make sense in terms of their own knowledge and experience. According to Bartlett, the story is assimilated to pre-stored schemata based on previous experiences.

Schema theory argues that our knowledge of the world is stored in memory as schemata, each of which incorporates all the knowledge of a given type of object or event that we have acquired from experience. Schemata operate in a bottom-up direction to help us interpret the bottom-up flow of information from the world. New experiences are not just passively copied or recorded into memory. A mental representation is activity constructed by processes influenced by schemata.

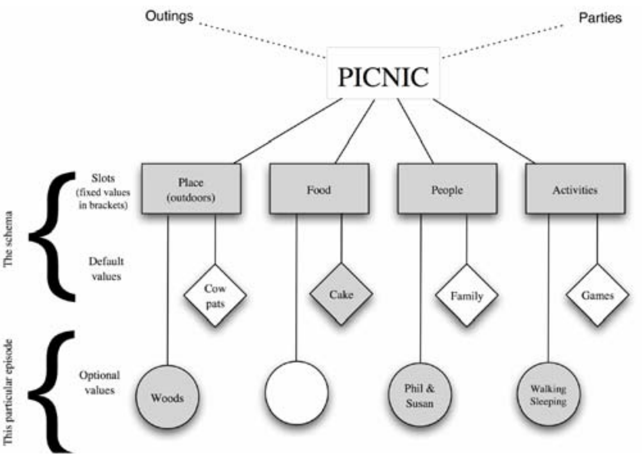


Figure 1: The picnic schema – after [21]

Current versions of schema theory have incorporated many of Bartlett's ideas, specifically the concept that what is encoded, stored and retrieved from memory is determined by existing schemata. Thus schemata drive the selection of what aspects of a new input will be stored and may modify the memory representation of a new experience so as to bring it into line with prior expectations and make it consistent with past experience. New experiences in turn can be stored as new schemata or modifications of old schemata, adding to our store of general knowledge. Structurally schemata are packets of information representing knowledge about objects, situations, events, or actions. Rumelhart and Norman [22] list five characteristics of schemata:

1. Schemata represents knowledge of all kinds from the simple to the more complex including *episodic memory and autobiographical memory*.
2. Schemata are linked together into related systems as can be seen in figure 3, the picnic schema is linked to 'outings' and 'parties'.
3. A schema has *slots* which may be filled with fixed, compulsory values or variable, optional, personal data.
4. Schema incorporate all kinds of different information we have accumulated.
5. Schemata operating at different levels may re-organize and interpret new inputs.

The interactionist, situated account of memory which we have proposed underpins the modified version of schemata which we now suggest as a psychologically-plausible means of *indicating* the contents for a particular type of (re)-experience. (It is tempting to suggest we could *specify* the contents, but it is much too early in the development of these concepts to make such a strong claim.) These schemata contain one set of slots whose default values are pre-identified and re-created using the IntoMyWorld technology. These values will draw on data from analysis of the contents of individuals' memories of the type of event concerned and historical accounts where available. A complementary set of slots have contents which are necessarily unique and personal to the individual: their contents cannot be pre-specified. The instantiation of these elements of the schemata, we propose, is achieved through the triggering of memories of one's emotions, activities, close companions and so forth by the IntoMyWorld experience. We argue that this is an interactive, situated process which will engender re-immersion in the event and a sense of being there again. However we do recognize that there is a good deal of conceptual work to be done to bring together Conway's work with schema theory.

As an illustration, section 6 describes how an analysis of individual memories of VE day might indicate the range of elements to be included as slots in an IntoMyWorld schema, the default values for some of these slots, and the sort of instance values which might be expected to be triggered when an individual re-experiences the events of this particular piece of the past.

5.4 Activating Schematic Memories

Central to our reasoning is the belief that the IntoMyWorld mixed-reality system can trigger or more correctly, activate networks of autobiographical schemata. In this section we discuss prior studies intended to specifically activate schematic memories. Unhappily the evidence is 'sharply contradictory' as noted by Rojahn and Pettigrew's meta-review of 60 independent studies with 165 comparative tests [23]. They found overall result shows a slight overall memory advantage for schema-*inconsistent* information. However they are quick to note that the effects are highly heterogeneous. They go on to remark that "schema-based processing is moderated by an array of variables" including "guessing and whether the measurements were of recall or recognition, length of exposure to inconsistent information, delay between presentation of the stimulus and the memory test, proportion of inconsistent items, order of schema-presentation, degree of inconsistency and importance of categories to subjects all had significant impacts on inconsistency resolution". To this should be added that while there is considerable interest in the fine psychological detail as to how schemata operate, for the current purposes we have settled for the psychological plausible. And there is evidence that schemata do apply to memories mediated by virtual environments. For example, Flannery and Walles [24] have reported a study in which explored how schemata operate in a well-known environment and to examine whether or not schemata operate differently in real versus virtual environments. They found that the virtual reality situation produced similar outcomes compared to the real world.

6. Remembering VE Day

The preliminary study below illustrates how schemata might be derived from a set of real world accounts of being present at a memorable event. There is of course a wide spectrum of events which might be re-created in IntoMyWorld, from the small-scale intimacy of a child's first words to the large public occasion. The event discussed here is largely of the latter type, namely the VE day celebrations of 7/8 May 1945. VE day immediately followed the Nazi surrender to Allied forces towards the end of the Second World War. In Britain Winston Churchill announced the end European hostilities on the evening of 7 May, declaring May 8 as a public holiday and day of celebration. Historical records and contemporary accounts document people celebrating in the streets on the evening of May 7 and throughout May 8, celebrations which culminated in London with the appearance of members of the Royal family and Churchill on the balcony of Buckingham palace. Outside London celebrations were naturally smaller in scale and varied in character.

6.1. Archive sources

A number of publicly available web archives are available which bring together personal, individual memories of VE day. Of these, the largest and most accessible collections identified were those collated by SAGA[1] magazine [25], the Museum of London [26], and the BBC [27]. The memory texts have been volunteered by individuals and range in length from a couple of sentences to 10 or so paragraphs. The wording of the call for contributions to the SAGA magazine is not available. The Museum of London specifically prompts its contributors "How did you find out that the war in Europe was over? What did you do? How did you feel? What did it mean? What do you think now, looking back? ". On the BBC site (dedicated to WW2 history in general), contributors are asked to be "post authentic stories based on their own, honest interpretations of the time." It is also suggested that they may wish to check facts against material elsewhere on the site, and to read other stories to gain inspiration. There do not, however, appear to be systematic differences in style or content between these three archives analysed, and the description following aggregates the texts from the three collections.

A caveat here: it is impossible to know from the information available on the web archives exactly how far, if at all, the texts have been edited. Certainly all examples are grammatical, correctly spelled and so forth and are coherent 'stories' or fragments of stories, and some of the BBC stories which appear on the main page of the archive in question have had their text "polished and cross-referenced". However, since the point of this study is to identify possible schemata in the content of VE day memories we do not consider the likelihood of some traces of an editorial hand to be a problem.

24 texts in total described the experience of VE day somewhere in Britain, whether from the perspective of a child, a civilian teenager or adult, or a member of the armed forces joining public celebrations. A further 11 texts – treated separately – are accounts of VE day from the perspective of members of the forces on active service overseas.

---

[1] SAGA is a British organization 'providing high-quality services for people aged 50 and over'.

6.2. Identifying Potential Schemata And ‘Slots’

An iterative process of categorization and re-categorization of the content of the accounts resulted in the identification of recurrent elements in the 24 ‘home’ texts which were agreed by both authors. Both first-person and third-person perspectives were evident: representative examples of the former being the identification of one’s own vantage point in a crowded public space or feeling thrilled to see the Royal Family, of the latter descriptions of crowds singing and dancing, or the sight of many bonfires.

Only 11 texts were generated by members of the services abroad. Given this small body of evidence these do not merit systematic comparison with the civilian data. The main differences, however, appear to be more reporting of one’s own activities, fewer mentions of large masses of people and less explicit retrospection.

Even a passing consideration of the main dataset, civilian celebrations, suggests that the four of main elements of any story are present – who, what, when, where. Why is usually implicit, but is the historical fact of the end of the European war. More usefully, however, we can begin to see what a set of ‘VE day celebration’ schemata might look like – as shown in table 1 below.

6.3 IntoMyWorld In Practice

We are now in a position to visualize IntoMyWorld in operation. We have used a data-rich historical event to illustrate some of the psychological challenges facing us but IntoMyWorld is expected to be used prospectively. We can imagine that it has been used to capture and record the salient features of the a significant event guided by a system of interlinked schemata. The mixed reality platform which is at the heart of IntoMyWorld is then able to create a chiaroscuro of the scene / event, and in doing so activates the interlinked, schematically organized autobiographical memories of the user. Then, technology and psychology work together to re-immersed the user in the event.

Mass public celebration

Crowds[2]	Extremely large crowds[3]
Public figures	Royal family and Churchill “shouting and cheering at the appearance of Winston Churchill and all the Royal Family” *Saga 4*
Central public space	In front of Buckingham palace “being in the crowd outside Buckingham Palace” *MoL 11*
Crowd behaviour	Singing & dancing, hugging and kissing, waving flags “Hundreds of people all waving flags were crowding in” *BBC 1*
Crowd emotions	Joyful and relieved. “The crowds were enormous and unbelievably joyful and happy” *MoL 9*
Physical sensations	Tired at the end of the day “reached my home with very sore feet” *MoL 9*

Being part of a historic event

What led up to the event	Wartime events as experienced by London population “We had just come through the horrors of ‘Doodlebugs’ & V2 rocket attacks” *MoL 3*
Who I was with	Friends and family “went by train from New Eltham to Charing Cross to join up with our fellow students.” *MoL 9*
Where I was	In central London
What I feel about it now	Grateful to others. “Looking back now, I really appreciate all the efforts of our forces and those at home.” *MoL 5*

Table 1: Possible VE-day Schemata

7. Discussion

In this paper we have reviewed the main strands of current autobiographical memory research and indicated the relationship of autobiographical memory to presence. We then presented a case for an interactive, situated treatment of memory based on schema theory and illustrated how a modified version of schemata might serve as a plausible way of embodying the aspects of memory theory which are most pertinent for applications aiming to recreate the personal past. The last section of the paper discussed how VE-day schemata might be defined and re-created, using the IntoMyWorld technology to provide environmental cues matching the default values of slots in the schemata, which together with the instantiation of an individual’s personal slots and values would trigger re-immersion in past memories.

We hope that such schemata are both psychologically plausible and sufficiently substantial to provide a starting point in guiding the design of the IntoMyWorld technologies and the contents of the experiences to be re-created. The success of the concept in this and any similar application contexts will demand close co-working between psychologists and technologists to create a conceptual framework and define associated terminology which is theoretically robust and practically useful. There are, of course, numerous questions raised but left unanswered by the argument and proposals above. To indicate just one question which is directly pertinent, what is the role of priming in successfully activating the contents of memory schemata in such contexts? There is little helpful applied material to guide us in this area – for example Nunez and Blake [28] who proposed a schema-based account of perception and presence, found no direct relationship between sense of presence and priming participants with literature relevant to the virtual environments experienced, so the effects are not as clear-cut as might be expected. Nonetheless, most documentaries and feature films

---

[2] Possible ‘slots’ (illustrative examples only)

[3] Possible default values for VE day from archived accounts with selected illustrative quotations. We are aware here that the provision of a set of slots with default values raises the unwelcome possibility of the manipulation of the contents of people’s memory. Clearly there are ethical issues here which demand very careful consideration.

featuring historical events prime their viewers with brief factual introductions, period music and other effects. How might these and other stimuli best be adapted to support the transition to a recreated past? What might be the role of tangible stimuli? Applications which seek to create a sense of presence in the past provide both stimulating technological challenges and a rich environment for the exploration and further development of the theory of memory and other related areas of psychology.

References

[1] Whitten, W.B. and Leonard, J.M. (1981) Directed search through autobiographical memory. *Memory and Cognition*, **9**, 566-579.

[2] Birren, J. E. and Deutchman, D. E. 1991. *Guiding autobiography groups for older adults: exploring the fabric of life. Baltimore*, Johns Hopkins University Press.

[3] Haight, B. K. 1991. Reminiscing: the state of the art as a basis for practice. *International Journal of Aging and Human Development*, **33**, 1-32.

[4] Hughston, G. and Merriam, S. 1982. Reminiscence: a nonformal technique for improving cognitive function in the aged. *Journal of Aging and Human Development*, **15**, 139-149.

[5] McCloskey, L. J. 1990. The silent heart sings. Generations, **14**, 63-65.

[6] Riva, G., Waterworth, J. A. and Waterworth E. L. (2004) The Layers Of Presence: A Bio-Cultural Approach To Understanding Presence In Natural And Mediated Environments, *CyberPyschology & Behavior*, **7(4)**, 405-419.

[7] Clark, A. (1995) Moving minds: re-thinking representations in the heat of situated action. In J. Tomerlin (Ed.) Philosophical Perspectives 9: AI, Connectionism and Philosophical Psychology.

[8] Clark, A. (1997) *Being There*. MIT Press, Cambridge: MA.

[9] Clancey, W. J. (1991) Situated cognition: Stepping out of representational flatland. *AI Communications—The European Journal on Artificial Intelligence*, **4(2/3)**, 109-112.

[10] Gero, J.S. and Peng, W. (2004) A Situated Agent-Based Design Assistant. In HS Lee and JW Choi (eds) *CAADRIA 2004*, Yonsei University Press, 145-157.

[11] Bartlett, FC (1932) *Remembering*. Cambridge University Press.

[12] Maturana, H. R. and Varela, F. (1980) Autopoiesis and Cognition: The Realisation Of The Living. Dordrecht: Reidel.

[13] Gibson, J. J. (1986) *The Ecological Approach to Visual Perception*. Lawrence Erlbaum Associates, Publishers. Hillsdale, New Jersey.

[14] Suchman, L. A. (1987) *Plans and situated action*. Cambridge: CUP.

[15] Bannon, L. (1991) From Human Factors to Human Actors. In J. Greenbaum and M. Kyng (Eds.) *Design at work: Cooperative Design of Computer Systems*. Hillsdale: Lawrence Erlbaum Associates, 25-44.

[16] Brown, N.R. (1990) Organization of public events in long-term memory. *JEP: General*, **19(3)**, 297-314.

[17] Reiser B.J., Black, J.B. and Abelson, R.P. (1985) Knowledge structures in the organization and retrieval of autobiographical memory. *Cognitive Psychology*, **17**, 89-137.

[18] Tulving, E. (1972) Episodic And Autobiographical Memory. In E. Tulving and W. Donaldson (Eds.) *Organization of Memory*. Academic Press.

[19] Conway, M.A. (2001) Sensory-perceptual episodic memory and its context: autobiographical memory. *Philosophical Transactions of the Royal Soc. London B.*, **356**, 1375-1384.

[20] Conway, M.A. (2003) Commentary: Cognitive-Affective mechanisms and processes in autobiographical memory, *Memory*, **11(2)**, 217-224.

[21] Cohen, G., Kiss, G. and LeVoi, M. (1993) *Memory – Current Issues*. Open University Press.

[22] Rumelhart and Norman (1983) Representation in Memory. In RC Atkinson, HJ Herrnstein, G Lindzey and RD Luce (eds) Handbook Of Experimental Psychology. Wiley and Son.

[23] Rojahn K. and Pettigrew, T.F. (1992) Memory for schema-relevant information: a meta-analytic resolution. *British Journal of Social Psychology*. **31(2)**, 81-109.

[24] Flannery K.A. and Walles, R. (2003) How Does Schema Theory Apply to Real versus Virtual Memories? CyberPsychology & Behavior, **6(2)**, 151-159

[25] www.saga.co.uk/magazine/article/B9866C43-E1C7-4646-A69D-0E6716B6CF89.asp?bhcp=1 (last accessed 05.05.05)

[26] www.museumoflondon.org.uk/frames.asp?http%3A//www.museumoflondon.org.uk/MOLsite/londonsvoices/veday/ (last accessed 05.05.05)

[27] www.bbc.co.uk/dna/ww2/C1198 (last accessed 05.05.05)

[28] Nunez, D. and Blake, E. (2003) Conceptual priming as a determinant of presence in virtual environments. Proceedings of the 2nd international conference on computer graphics, virtual reality, visualisation and interaction in Africa, 101-108

Session 2

21st September, 2005

Body, Space and Motion

15.45-16.15 *Is this My Hand I see before me? The Rubber Hand Illusion in Reality, Virtual Reality and Mixed Reality*
Wijnand IJsselsteijn, Yvonne de Kort and Antal Haans
Human-Technology Interaction Group, Department of Technology Management,
Eindhoven University of Technology, The Netherlands

16.15-16.30 *Influence of Auditory Cues on the Visually Induced Self-Motion Illusion (Circular Vection) in Virtual Reality*
Bernhard Riecke, Jörg Schulte-Pelkum, Franck Caniard, and Heinrich H.Bülthoff
Max Planck Institute for Biological Cybernetics, Tübingen, Germany

16.30-17.00 *Neural Processing of Spatial Information: What we know about place cells and what they can tell us about presence*
Jorge Brotons[1], Shane O'Mara[2] and Mavi Sanchez-Vives[1]
[1] Instituto de Neurociencias de Alicante, Universidad Miguel Hernandez-CSIC, Spain
[2] Institute of Neuroscience, Trinity College, Dublin, Ireland.

Is This *My* Hand I See Before Me?
The Rubber Hand Illusion in Reality, Virtual Reality, and Mixed Reality

W. A. IJsselsteijn, Y. A. W. de Kort, A. Haans
Eindhoven University of Technology, Human-Technology Interaction Group
{w.a.ijsselsteijn@tue.nl, y.a.w.d.kort@tm.tue.nl, a.haans@tm.tue.nl}

Abstract

This paper presents a first study in which a recently reported intermodal perceptual illusion known as the rubber hand illusion is experimentally investigated under mediated conditions. When one's own hand is placed out of view and a visible fake hand is repeatedly stroked and tapped in synchrony with the unseen hand, subjects report a strong sense in which the fake hand is experienced as part of their own body. In our experiment, we investigated this illusion under three conditions: (i) unmediated condition, replicating the original paradigm, (ii) virtual reality (VR) condition, where both the fake hand and its stimulation were projected on the table in front of the participant, and (iii) mixed reality (MR) condition, where the fake hand was projected, but its stimulation was unmediated. Dependent measures included self-report (open-ended and questionnaire-based) and drift, that is, the offset between the felt position of the hidden hand and its actual position. As expected, the unmediated condition produced the strongest illusion, as indicated both by self-report and drift towards the rubber hand. The VR condition produced a more convincing subjective illusion than the MR condition, although no difference in drift was found between the mediated conditions. Results are discussed in terms of perceptual mechanisms underlying the rubber hand illusion, and the illusion's relevance to understanding telepresence.

Keywords--- **Rubber hand illusion, multisensory integration, body image, virtual reality, mixed reality**

1. Introduction

One need not be a chamber to be haunted,
One need not be a house;
The brain has corridors surpassing
Material place

– Emily Dickinson

Human brains seem to support highly malleable body images. Although intuitively we expect our body image to be durable and permanent, evidence is mounting that suggests that our sense of bodily self-identification – the ability to distinguish what's contained within versus what's beyond our familiar biological shell – is a flexible, temporary construct and not a fixed property. Having a negotiable body image has clear survival value when considering the profound bodily changes that the brain has to accommodate during a lifetime of physical development and change. What is most surprising here, however, is the relative speed at which the brain appears to support a significantly altered body image after just a few minutes of the right kind of sensory stimulation.

A particularly interesting and relevant phenomenon in this respect is a recently reported intermodal perceptual illusion known as the rubber hand illusion [1,2,3]. When a person is watching a fake, rubber hand being stroked and tapped in precise synchrony with his or her own unseen hand, the person will, within a few minutes of stimulation, start experiencing the rubber hand as an actual part of his or her own body. In part, this illusion illustrates the importance of visual information in specifying limb location and constructing the body image (cf. [4]). For example, when seen and felt hand position are in conflict, as is the case when one wears a prism that displaces the entire visual field to one side, the visually displaced hand is usually felt where it is seen, a phenomenon know as immediate visual capture [5]. The visual adaptation of proprioceptive position that occurs during the rubber hand illusion is related, though not identical, to prism adaptation (see [6] for an overview). After prolonged exposure to prism-induced visual displacements, after-effects will occur including misreaching in the direction opposite to the previous visual displacement. Similar effects have been reported in adapting to tele-systems and virtual environments (see, e.g., [7]). However, a key distinguishing feature of the rubber hand illusion is that it emerges from closely correlated visual and tactile stimulation, resulting in a strong sense of body ownership of the fake hand. The correlation between visual, tactile and proprioceptive information can be thought of as self-specifying for bodily self-identification, as the brain has learned from a very early age onwards that it can only be the body, and no other object, that can register these specific intersensory correlations [8].

The extent to which non-biological artefacts, such as a rubber hand, can be incorporated as a phenomenal extension of the self has clear relevance to the area of telepresence [9]. Understanding the conditions under which such integration may or may not occur has implications for the design of virtual environments, teleoperation and mixed reality systems, and ways in which the body may be optimally represented in such mediated environments. More importantly, it enhances our fundamental understanding of the phenomenal experience of telepresence and the psychological and brain mechanisms involved in distinguishing self from non-self, and reality from mediation.

In this paper, we report on an experiment we performed to investigate the rubber hand illusion under mediated conditions. However, before describing the rationale of the experiment, we will first turn to the rubber hand illusion in more detail.

1.1. The Rubber Hand Illusion

Botvinick and Cohen [1] provided a first description of the rubber hand illusion. This crossmodal perceptual illusion occurred when participants' left hand was placed out of view and a life-size rubber facsimile of a human hand was placed in front of them. Subsequently, both the rubber hand and participants' left hand were gently stroked by two small paintbrushes, synchronizing timing as closely as possible. Subjects reported feeling a sense of ownership of the rubber hand, as if it was actually their own. In addition to self-report, Botvinick and Cohen also employed a measure of drift, where subjects were asked to close their eyes and align their right index finger with the index finger of their unseen left hand. Results showed a proprioceptive drift towards the rubber hand, with the magnitude of drift correlating significantly with the reported duration of the illusion.

Although Botvinick and Cohen interpret their results as an effect of visual information overriding the incongruent proprioceptive information, Armel and Ramachandran [2] contest this claim, demonstrating that the illusory sensation can also be elicited by merely stimulating the tabletop in front of a participant, which bears no visual resemblance to a hand (see also [10]). They argue that the illusion mainly arises "from the 'Bayesian logic' of all perception; the brain's remarkable ability to detect statistical correlations in sensory inputs in constructing useful perceptual representations of the world – including one's body."([2], p. 1500). Armel and Ramachandran (2003) further showed that when the physical integrity of the rubber hand was threatened (bending a finger backwards to seem painful), a clear skin conductance response was generated. The illusion could even be projected to anatomically impossible locations, with the rubber hand positioned at a distance. It is important to note, however, that although Armel and Ramachandran's study showed that the rubber hand illusion is relatively robust to manipulations of form or location (i.e., the illusion still occurs to an extent), the subjective intensity appears to be much lower under these circumstances, and in particular in the tabletop condition. This questions the authors' interpretation that the illusion is resistant to top-down knowledge from cognitive body representations and is solely governed by the brain's ability to extract statistical correlations when perceptions from different modalities co-occur with a high probability. Indeed, a series of experiments recently reported by Tsakiris and Haggard [3] support the contention that bottom-up visuotactile correlations are modulated by top-down influences originating from one's body representation in creating the rubber hand illusion. However, Tsakiris and Haggard's results are solely based on measuring drift, making direct comparisons between their results and those of Armel and Ramachandran difficult.

1.2. Rationale of the Current Experiment

The experiment reported in this paper was performed for three reasons. First of all, we wanted to introduce intermediate levels in form manipulation between the original rubber hand illusion as reported by Botvinick and Cohen [1] and the 'table illusion' as reported by Armel and Ramachandran [2]. Teasing apart and testing the various form factors that influence the vividness of the rubber hand illusion will allow us to better understand the contributing processes, in particular the role of the cognitive body representation, underlying the illusion. To this end, we chose to use a video-projection of a rubber hand (and its synchronous stimulation) onto the flat tabletop surface (we dubbed this the 'virtual reality' condition), thus reproducing the rubber hand form in terms of basic contour, size, texture and colour. The main perceptual difference was in terms of perceived rubber hand volume. By using a non-tracked, monoscopic projection, the stereoscopic and motion parallax cues to object shape were absent, allowing us to assess the impact that these cues have in activating our cognitive body scheme by comparing this condition with the unmediated condition, where these cues are available.

Secondly, Armel and Ramachandran [2] reported anecdotally that the table illusion was more vivid if subjects could see a common texture being synchronously manipulated – in their case a band-aid placed on both the subject's real hand and the table surface. To test this, we chose to project the rubber hand on the tabletop in front of the participant (as before), however with the touch stimulation being unmediated, that is, applied directly to the tabletop projection visible in front of the participant instead of to the rubber hand which was being recorded. Thus, this 'mixed reality' condition would allow us to check whether inconsistencies in perceived texture would diminish the vividness of the illusion.

Lastly, since the rubber hand illusion appears to be a cognitively impenetrable perceptual illusion, the level to which it can be reproduced under mediated conditions may provide us with an interesting indicator of the perceptual quality of a particular form of mediation, and thus a potential indicator of presence.

2. Method

2.1. Design

In this study, we aimed to compare the 'traditional' unmediated rubber hand condition (see Figure 1A) with 2 types of mediated conditions. First, in what we call the Virtual Reality condition (VR), both the rubber hand and the stimulation of the rubber hand (via a small painter's brush held by the experimenter) were projected on the table in front of the participant (see Figure 1B), thus providing a fully mediated equivalent of the original rubber hand experiment, as reported by Botvinick & Cohen [1], and employed in subsequent studies by various others. Second, in the Mixed Reality condition (MR), the rubber hand was again projected in front of the participant (as in the VR condition), yet this time the stimulation by the brush was physically applied to the projection of the rubber hand, rather than to the rubber hand itself (see Figure 1C).

As the existing literature points to significant variations between individuals in both the experienced nature and vividness of the rubber hand illusion, we decided to use a basic within-subjects design to control for this potential variation and increase our experiment's sensitivity. To compensate for potential order effects, the three conditions were presented in fully counterbalanced order.

2.2. Participants

Twenty-four participants, 15 male, 9 female, between 20 and 32 years of age, all with normal or corrected-to-normal vision, took part in this study. Twenty participants were right-handed, three were left-handed, and one had mixed handedness. Participants were either students or employees at the Eindhoven University of Technology in Eindhoven, The Netherlands. They were naïve to the hypothesis under test. Students were compensated with € 7 for their participation.

2.3. Setting and Apparatus

The experiment was conducted at the UseLab of the Human-Technology Interaction Group. The UseLab is a usability laboratory equipped with standard living room furniture as well as state-of-the-art observational technologies and tools. Figure 1 shows the setup that was used in this experiment for the three conditions (A: unmediated, B: virtual reality, and C: mixed reality). The fake "rubber" hand used in all conditions was highly realistic in terms of colour, skin texture, size and shape. It was originally developed by Otto Bock Benelux B.V. as a prosthetic left hand and kindly donated to the authors for research purposes.

A wooden separating screen was used to obscure the view the participants had onto their own left hand. Also, in the VR and MR conditions, the rubber hand was itself placed out of view, behind the separating screen. The rubber hand, or its projection, was placed in a natural position in relation to the participant's torso, slightly left in front of the participant. This would be a comfortable position if it were the participant's own hand (i.e., not an anatomically implausible location – cf. [2]). The distance between the participant's left hand, placed out of view, and the rubber hand (or its projection) was approximately 30 cm. Two small brushes were used to synchronously stroke congruent positions on both the rubber hand (or its projection) and the participant's unseen left hand.

A standard mini-DV camera, mounted on a tripod, was used to record the rubber hand and the stimulation in the VR condition, or only the rubber hand in the MR condition. The camera was mounted such that it had a top view of the recording area, on the left side of the separating screen, as depicted in Figure 1 (panels B & C). The camera's output was connected to an InFocus LP750 projector, which projected directly onto the tabletop surface in front of the participant. Care was taken that the rubber hand projection was of the same size as the rubber hand itself, and that its perspective was matching the participant's viewpoint.

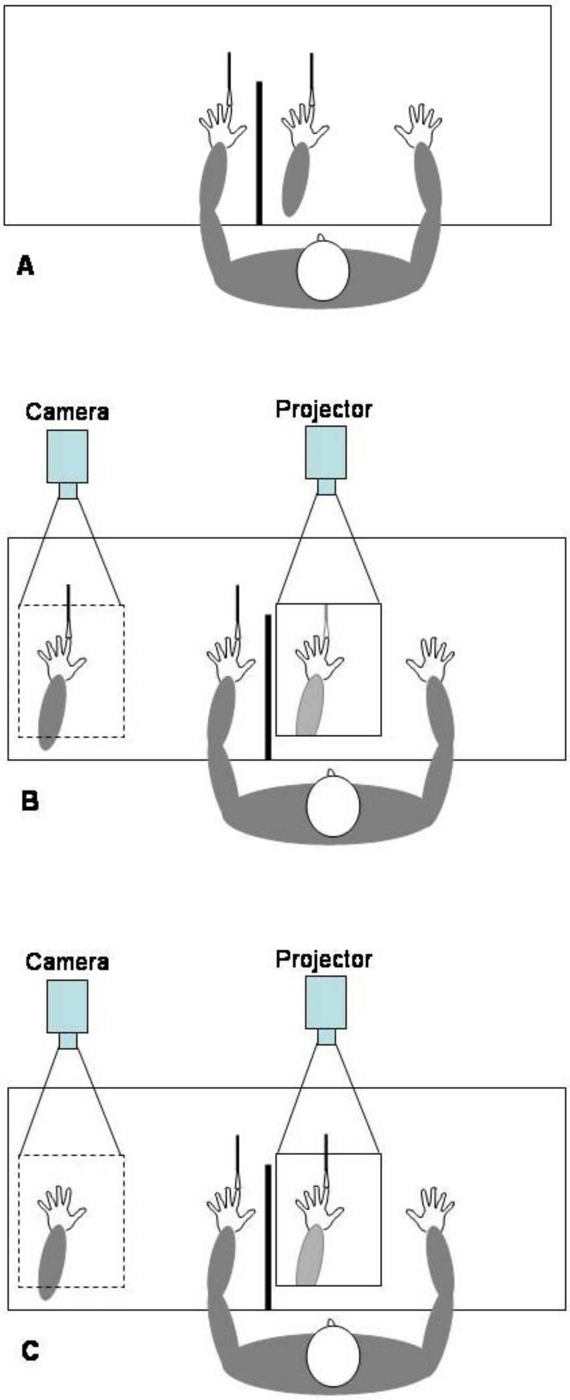


Figure 1. Overview of the three experimental conditions: A) *Unmediated condition*: the rubber hand and its stimulation are both physically present in the participant's field of view. B) *Virtual Reality condition*: both the rubber hand and its stimulation are presented as projections on the table surface in front of the participant. C) *Mixed Reality condition*: the rubber hand is projected in front of the participant and unmediated stimulation is applied to this projection.

2.4. Measurement

In the present experiment we employed self-report to directly assess participants' experiences, and measured drift as an objective corroborative measure of the rubber hand illusion. Self-report included a questionnaire as well as an open-ended, qualitative description of the experience.

2.4.1. Questionnaire The questionnaire was adopted from Botvinick and Cohen [1]. Their questionnaire consists of 9 statements describing specific perceptual effects associated with the rubber hand illusion, such as "I felt the rubber hand was my hand" or "It seemed as though the touch I felt was caused by the paintbrush touching the rubber hand". All items were translated into Dutch.

Three changes were made to this questionnaire. Firstly, the last item ("The rubber hand began to resemble my own (real) hand, in terms of shape, skin tone, freckles or some other visual feature") was divided into two separate items, one tapping resemblance between the rubber hand and the real hand in terms of shape, the other in terms of texture. Secondly, one item was added describing a sensation that a number of people reported during the pilot phase of the study: "It felt as if my hand was inside the rubber hand". Lastly, the 7-point response scale used by Botvinick and Cohen, running from '---' via '0' to '+++' was reformulated to run from 'not at all' to 'completely'. The resulting 11 items are reported in the caption of Figure 2 in the results section.

2.4.2. Drift Drift is a measure gauging a distortion of proprioception in participants that typically occurs after exposure to the rubber hand stimulation. With eyes closed and keeping their left hand in place on the table, participants were asked to indicate the location of their left hand by moving their right hand in a straight line below the table until they feel both hands are in alignment with each other. This task was performed before and after each condition. Drift was calculated by subtracting the displacement to the right (i.e., towards the rubber hand) pre-exposure from the displacement post-exposure (similar to the method used in [3]).

2.5. Procedure

As our study was aimed to elucidate to what extent the rubber hand illusion would occur under mediated conditions, we selected participants on the basis of a short pilot test in which it was established that they indeed were able to experience the rubber hand illusion. Of the 30 participants that were recruited, 6 were excluded from partaking in the study as they did not report any sign of the illusion. The pilot test and the main experimental study were at least one week apart.

On arrival at the UseLab, participants were seated behind a standard office table with a white tabletop surface, and were asked to place their left hand palm face down in a relaxed position on top of a marker behind the wooden partition. This setup ensured that participants were unable to view their real left hand and arm. Participants were instructed not to move their left hand during the experiment, and to focus their attention on the fake hand that was placed in a natural position in front of them.

The experiment was divided into three sessions, one for each condition. Conditions were completely counterbalanced, yielding 6 unique orders. In the unmediated and VR conditions, the experimenter synchronously stroked the fingers of the participant's invisible left hand and the rubber hand for approximately 7,5 minutes, using a small brush. In the MR condition, the experimenter stroked the *projection* of the rubber hand on the table surface in front of the participant instead of the rubber hand itself. After 7,5 minutes of synchronous stimulation in each condition, participants were asked to immediately close their eyes and indicate the felt position of their left hand, in order to establish a measure of drift. Subsequently, participants were asked to fill out the questionnaire. Finally, after each condition, participants were asked to recount *in their own words* what the experience had felt like to them, plus any other remarks they would like to make about the experiment itself. The total experiment took approximately 45 minutes to complete.

3. Results

The rubber hand illusion was measured with an 11 item questionnaire [1] and a drift measure. Furthermore, qualitative data were obtained from the participant's open-ended descriptions. Results from the questionnaire and drift measures will now be presented separately, followed by some illustrative quotes from the participants, recorded after each session.

3.1 Questionnaire

Scores on the questionnaire items for the three experimental conditions are reported in Figure 2. A clear picture emerges of the rubber hand illusion being strongest in the unmediated condition, followed by the VR and lastly the MR one. Similar to the findings by Botvinick and Cohen [1], the first three items showed greatest variance and effects of our manipulations. These were studied more rigorously employing repeated measures analyses of variance (REMANOVA).

The first item ('It seemed as if I were feeling the touch in the location where I saw the rubber hand touched') was analysed in a REMANOVA with Mediation as the independent factor. This factor was significant ($F(2,46) = 10.70$, $p<.001$, partial eta squared $=.32$)[1]. Subsequent contrast analyses revealed a significant difference between the unmediated condition and the two mediated conditions ($p<.001$), but not between the VR and MR condition.

---

[1] Partial eta squared is an estimate of the degree of association between the dependent and independent variables for the sample and can be interpreted as the proportion of variance in the dependent variable that is attributable to the effect of the independent variable. It is used as an indicator of effect size and its value varies between 0 and 1.

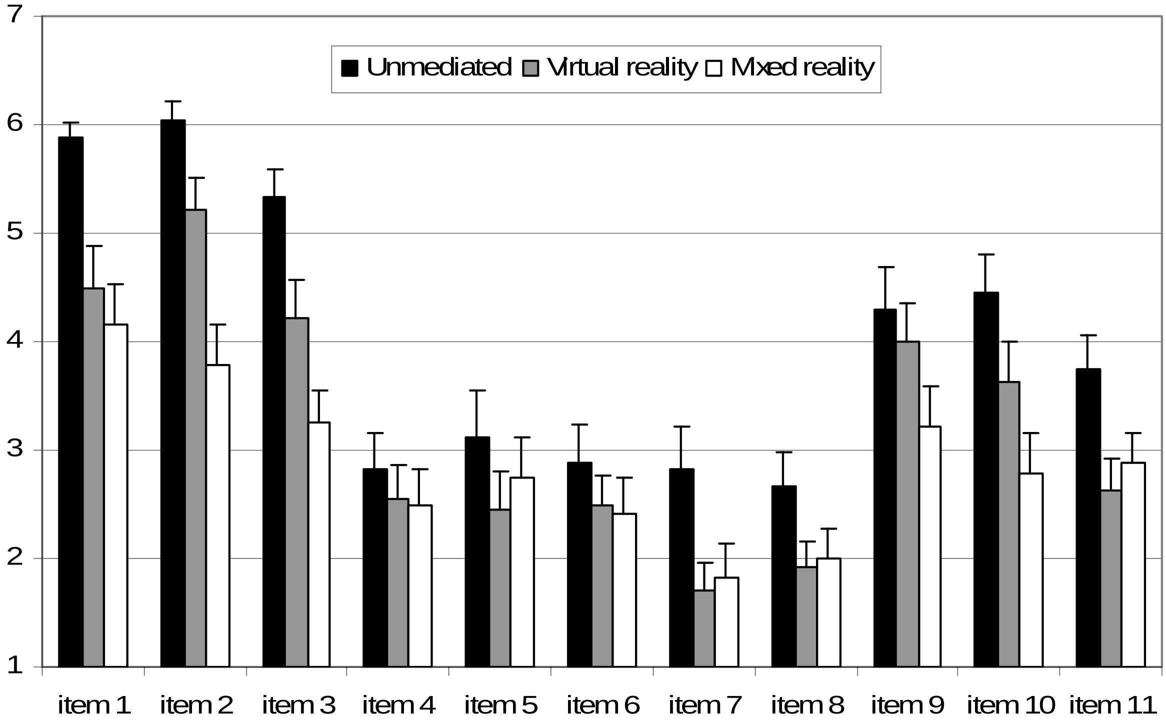


Figure 2: Questionnaire data, presenting means and standard errors of each item for the three experimental conditions. Item 1- It seemed as if I were feeling the touch in the location where I saw the rubber hand touched; item 2 – It seemed as though the touch I felt was caused by the paintbrush touching the rubber hand; item 3 - I felt as if the rubber hand were my hand; item 4 – It felt as if my hand were drifting towards the rubber hand; item 5 – It seemed as if I had more than 1 left hand or arm; item 6 – It seemed as if the touch I was feeling came from somewhere between my own hand and the rubber hand; item 7 – It felt as if my hand was turning rubbery; item 8 – It appeared as if the rubber hand were drifting towards my hand; item 9 – The rubber hand began to resemble my hand in form; item 10 – The rubber hand began to resemble my hand in texture; item 11 – It felt as if my hand was in the rubber hand.

Similar analyses with the second item (‘It seemed as though the touch I felt was caused by the paintbrush touching the rubber hand’) again revealed a significant effect of Mediation (F(2,46)=25.87, p<.001, partial eta squared = .53). This time all contrasts were significant (p<.001).

Analyses of the third item (‘I felt as if the rubber hand were my hand’) also resulted in a significant effect of Mediation (F(2,46)=15.98, p<.001, partial eta squared = .41) and all contrasts significant (p≤.01).

The remaining items showed similar patterns as those described earlier, although in general effects were smaller and not always significant. As a final check, we performed a REMANOVA with the mean score on the 11 items as the dependent variable, and Mediation, Gender, Handedness and Experimental Order as independent variables. Again, Mediation turned out significant (F(2,9)=8.39, p=.009, partial eta squared = .65), while no remaining significant main or interaction effects emerged.

3.2 Drift

Drift measurements for the three experimental conditions are summarized in Figure 3. Although less clear, the pattern resembles the one found in the questionnaire data: proprioceptive drift of the left hand towards the location of the rubber hand is strongest in the unmediated condition, and weaker in both mediated conditions. A REMANOVA with drift dependent and Mediation independent resulted in marginally significant effects (F(2,42)=2.64; p=.08). After discarding 1 outlier who had standardised scores over 1.96 in all conditions[2], differences became a bit more pronounced, resulting in a significant effect of Mediation (F(2,40)=3.71, p=0.03, partial eta squared=.16). Contrast analyses revealed significant differences only between the unmediated condition on the one hand and the mediated conditions on the other.

3.3 Open-ended descriptions

The open-ended description proved to be quite informative. In the unmediated condition there were many cases in which participants were using descriptions that signalled a sense of bodily ownership of the rubber hand. For instance:

“The feeling seems to build up the first few minutes and then, all of a sudden, the hand feels like my own. And after a while they start to look the same as well!”

“Soon you have the feeling the rubber hand is really your hand, you can really feel it being touched.”

---

[2] This criterion is known as Grubb’s test [11], and is similar to discarding data that differ more than two standard deviations from the mean.

Participants remarked that the illusion was particularly vivid when somewhat more force was applied by the experimenter, and the fingers of the rubber hand moved a little as a result. In both the VR and MR conditions there were several instances where participants also reported a strong sensation of ownership of the mediated rubber hand. In the VR condition, a number of participants also claimed that they felt as if the projection of the rubber hand was a projection of their own hand:

"*I had a feeling I was looking at a projection of my own hand.*"

"*It soon appeared as if the projection was my own hand, and my own hand was being touched.*"

Interestingly, in the MR condition some participants noted that the flat image appeared to obtain volume:

"*It felt as if the projection became three-dimensional, just like my own hand.*"

"*The illusion was not strong, but the image appeared to become 3D as time passed.*"

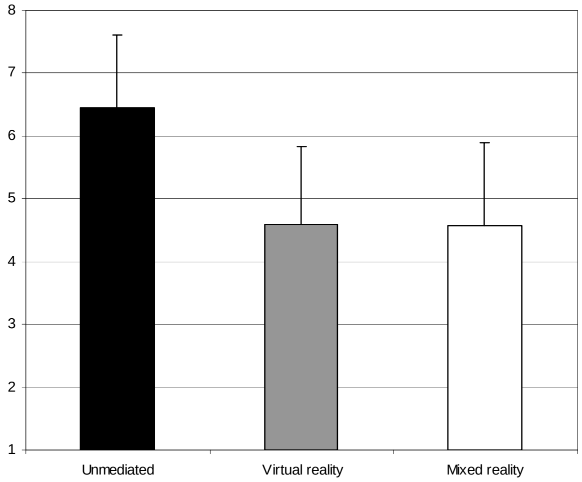


Figure 3: Mean drift in cm for the three experimental conditions. Error bars indicating standard error.

4. Discussion

Our questionnaire results in the unmediated condition clearly replicate the original Botvinick and Cohen results, although with somewhat lower variability in the data. This was to be expected as we selected participants on the basis of a pilot test that showed they were sensitive to the rubber hand illusion to some extent (only 6 of the 30 people tested did not reach this criterion). Nevertheless, the results of the unmediated condition illustrate that the rubber hand illusion can be reliably reproduced when similar procedures are being employed. The existence of the rubber hand illusion demonstrates that intermodal correlations between vision, touch, and proprioception can specify self-attribution of a non-self object [1]. That is, the rubber hand becomes part of the body image, thereby illustrating that the body image is a plastic, temporary construction that can be altered within a relatively short time-span.

The results of the self-report and drift measurements for the mediated conditions indicate that the rubber hand illusion still occurs, albeit to a significantly lesser degree than in the unmediated condition. This result partially contradicts Armel and Ramachandran's [2] claim that the rubber hand illusion is purely the result of Bayesian learning, whereby reliable correlations of visuotactile events are necessary and sufficient by themselves to constitute self-attribution. If this were true, no difference ought to be found between the VR and the unmediated condition, for instance. The fact that we did find a significant difference, however, points to the role of top-down mechanisms that specify requirements for a plausible and congruent (hand-shaped) visual object, if it is to be integrated within the body image. It should be noted however, that Armel and Ramachandran's own results also point to a potential role of top-down mechanisms, as both the subjective ratings and the electrodermal responses were significantly lower in the tabletop condition as compared to the rubber hand condition. Moreover, our results are in agreement with Tsakiris and Haggard [3] who also argue in favour of a combination of bottom-up and top-down processes in explaining the rubber hand illusion. Based on our results, we can argue that the top-down cognitive body representation needs to include a specification of the 3D shape of the hand, as this was the main difference between the projected (VR) and unmediated rubber hand conditions. In the near future, we will employ stereoscopic imaging to further investigate this issue

The VR condition provided participants with a more vivid illusion than the MR condition. This was also in line with our hypothesis, based on the assumption that in the MR condition, like in Armel and Ramachandran's table condition, there was an inconsistency in texture between the felt stimulation on one's skin, and the observed stimulation on the tabletop. This inconsistency was not present in the VR condition. However, after analysing the open-ended descriptions, an alternative explanation for the difference between the VR and MR results also needs to be considered. It appears that a significant number of participants had a quite convincing illusion that the remotely located rubber hand was their own, which was then subsequently being displayed in front of them. None of the participants mentioned this after the MR condition – this would not have made sense as the stimulation was happening on the table in front of them. In the MR condition then, the illusion appeared to suffer somewhat from the conflict between the real brush and the mediated hand. This points to the basic challenge of creating seamless perceptual fusion between the real and the virtual in mixed reality environments. Clearly, in our experiment, this was not yet the case, although for some participants, only in the MR condition, it appeared as though the 2D image became 3D. This illusion could be related to the perceptual system solving the "contradiction" of watching a flat hand being stroked by a 3D brush, and simultaneously feeling one's own unseen hand being stroked.

Overall, our experiment demonstrated that we can produce the rubber hand illusion using media, albeit somewhat less vivid than in the unmediated case. We have shown that form factors play a significant role in the occurrence and vividness of the rubber hand illusion, a fact that contradicts an exclusive adherence to Bayesian principles of statistical correlation. The fact that we can reproduce the rubber hand illusion under mediated conditions is promising for two reasons. First, to obtain a deeper understanding of the form, location, and temporal factors influencing the rubber hand illusion it is necessary to have complete and systematic control over the variables one may want to manipulate. Mediated environments provide such a level of control, combining ecological validity with the ability to systematically tweak relevant variables, and allow for precise replication of conditions [12]. Secondly, the extent to which the mediated rubber hand illusion occurs may in itself provide the research community with an interesting evaluation metric of the quality of the particular media environment under study. The fact that the vividness of the rubber hand illusion varied significantly across conditions in the experiment reported in this paper bodes well for the sensitivity of this measure.

In sum, the same sensorimotor and brain systems responsible for our sense of bodily boundaries are also remarkably adaptable to include non-biological artefacts within the perceptual-motor loop, provided reliable, real-time intersensory correlations can be established, and the artefact can be plausibly mapped onto the body image. When we interact with virtual or remote environments using intuitive interaction devices, isomorphic to our sensorimotor abilities, the real-time, reliable and persistent chain of user action and system feedback will effectively integrate the technology as a phenomenal extension of the self. This fluid integration of technology into the perceptual-motor loop eventually may blur the boundary between our ‘unmediated’ self and the ‘mediating’ technology.

Acknowledgements

The authors would like to thank Thomas Beffers, Ferry van Mil, Arjan Tekelenburg, and Bartjan Rijs for carrying out the experiment reported here as part of their Human-Technology Interaction Design Track. We also gratefully acknowledge Martin Boschman for assisting with the laboratory set-up, and Otto Bock Benelux B.V. in Son en Breugel, The Netherlands for donating the prosthetic left hand used in this experiment.

References

[1] M. Botvinick, J. Cohen. Rubber hands ‘feel’ touch that eyes see. *Nature, 391*, 756, 1998.

[2] K.C. Armel, V.S. Ramachandran. Projecting sensations to external objects: Evidence from skin conductance response. *Proceedings of the Royal Society of London B, 270*, 1499-1506, 2003.

[3] M. Tsakiris, P. Haggard. The rubber hand revisited: Visuotactile integration and self-attribution. *Journal of Experimental Psychology: Human Perception and Performance, 31*, 80-91, 2005.

[4] R.B. Welch. The effect of experienced limb identity upon adaptation to simulated displacement of the visual field. *Perception and Psychophysics, 12*, 453-456, 1972.

[5] R.B. Welch, D.H. Warren. Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin, 88*, 638-667, 1980.

[6] R.B. Welch. Adaptation of space perception. In: K.R. Boff, L. Kaufman, and J.P. Thomas (Eds.), *Handbook of Perception and Human Performance*, pp. 24.1-24.44. New York: Wiley, 1986.

[7] J. Groen, P.J. Werkhoven. Visuomotor adaptation to virtual hand position in interactive virtual environments. *Presence: Teleoperators and Virtual Environments, 7*, 429-446, 1998.

[8] M. Botvinick. Probing the neural basis of body ownership. *Science, 305*, 782-783, 2004.

[9] W.A. IJsselsteijn. Towards a neuropsychological basis of presence. *Annual Review of CyberTherapy and Telemedicine, 3*, 25-30, 2005.

[10] V.S. Ramachandran, W. Hirstein, D. Rogers-Ramachandran, D. Phantom limbs, body image, and neural plasticity. *Int. Brain. Res. Org. News, 26* (1), 10-11, 1998.

[11] F. Grubbs. Procedures for detecting outlying observations in samples. *Technometrics, 11*, 1–21, 1969.

[12] J. M. Loomis, J.J. Blascovich, A.C. Beall. Immersive virtual environment technology as a basic research tool in psychology. *Behaviour Research Methods, Instruments, & Computers, 31*, 557-564, 1999.

Influence of Auditory Cues on the visually-induced Self-Motion Illusion (Circular Vection) in Virtual Reality

Bernhard E. Riecke, Jörg Schulte-Pelkum, Franck Caniard, & Heinrich H.Bülthoff
Max Planck Institute for Biological Cybernetics, Tübingen, Germany
{bernhard.riecke@tuebingen.mpg.de, joerg.sp@tuebingen.mpg.de,
franck.caniard@tuebingen.mpg.de, heinrich.buelthoff@tuebingen.mpg.de}

Abstract

This study investigated whether the visually induced self-motion illusion ("circular vection") can be enhanced by adding a matching auditory cue (the sound of a fountain that is also visible in the visual stimulus). Twenty observers viewed rotating photorealistic pictures of a market place projected onto a curved projection screen (FOV: 54˚x45˚). Three conditions were randomized in a repeated measures within-subject design: No sound, mono sound, and spatialized sound using a generic head-related transfer function (HRTF). Adding mono sound increased convincingness ratings marginally, but did not affect any of the other measures of vection or presence. Spatializing the fountain sound, however, improved vection (convincingness and vection buildup time) and presence ratings significantly. Note that facilitation was found even though the visual stimulus was of high quality and realism, and known to be a powerful vection-inducing stimulus. Thus, HRTF-based auralization using headphones can be employed to improve visual VR simulations both in terms of self-motion perception and overall presence.

Keywords---**Vection, self-motion perception, spatial orientation, virtual reality, motion simulation, human factors, psychophysics, multi-modal cue integration, auditory cues, HRTF**.

1. Introduction

This paper addresses the visually induced self-motion illusion known as vection, and investigates whether additional matching auditory cues might be able to facilitate the illusion – if this were the case, it would have important implications for both our understanding of multi-modal self-motion perception and optimizing virtual reality applications that include simulated movements of the observer. Most people know the phenomenon of vection from real-world experience: When sitting in a train waiting to depart from the train station and watching a train on the neighboring track pulling out of the station, one can have the strong impression of moving oneself, even though it was in fact the train on the adjacent track that just started to move. A similar effect can be observed when sitting in the car waiting for the traffic light to turn green and when a close-by large truck slowly starts to move.

Such self-motion illusions can be reliably elicited in more controlled laboratory settings. Typically, vection has been investigated by seating participants in the center of a rotating optokinetic drum that is painted with simple geometrical patterns like black and white vertical stripes. When stationary observers are exposed to such a moving visual stimulus, they will at first correctly perceive motion of the visual stimulus (object motion). After a few seconds, however, this perception typically shifts toward oneself being moved and the moving visual stimulus slowing down and finally becoming earth-stationary. This self-motion illusion is referred to as circular vection, and the illusion has been studied extensively for more than a century [10, 21]. Excellent reviews on the phenomenon of vection are provided by [6, 15, 43]. More recently, the vection literature has also been revisited in the context of virtual reality (VR) and ego-motion simulation applications [13, 30]. So why is the phenomenon of illusory self-motion interesting in the context of VR?

Being able to move about one's environment and change one's viewpoint is a fundamental behavior of humans and most animals. Hence, being able to simulate convincing self-motions is a key necessity for interactive VR applications. There are a number different approaches to simulating ego-motion in VR, including motion platforms, free walking using head-mounted displays (HMDs), locomotion interfaces such as treadmills, or simply just presenting visual information about the self-motion. Each of these approaches offers distinct disadvantages: The drawback of using motion platforms is that they require a considerable technical and financial effort, and even then performance in VR is not necessarily comparable to corresponding real-world tasks like driving or flight simulations [1, 3, 23]. An often used alternative is to allow users to freely walk around while wearing a position-tracked head-mounted display. For most tasks, however, this requires a rather large walking area in which the observer's position is precisely tracked. This is, however, often infeasible or simply too costly. Using locomotion interfaces like treadmills or bicycles to allow for proprioceptive cues from physically walking or cycling etc. is often believed to be an optimal solution – there are, however, many open design and implementation issues that need to be carefully evaluated to come up with an optimal (and affordable) solution for a given task, especially if self-rotations are involved [14]. There has been only little research on the perception of ego-motion (vection) using treadmills, and informal observations suggest that participants hardly ever report compelling sensations of self-motion that is comparable to vection as experienced in optokinetic drums, even in the most advanced linear treadports. Durgin and Pelah state, for example, that “during treadmill locomotion, there is rarely any illusion that one is actually moving forward” [8]. Finally, when only visual information about the self-motion is provided, users hardly ever have a convincing sensation of self-motion, especially for the relatively small field of views that are common for off-the-shelf VR display devices.

In sum, despite tremendous progress in VR simulation technology, self-motion simulation in VR still poses a major challenge, and self-motion simulation is typically not as effective and convincing as corresponding real-world motions. This can lead to a number of problems including disorientation, reduced or misadapted task performance, general discomfort, and motion sickness (see, e.g., the discussion in [5, 29, 30]).

Nonetheless, it is known that moving visual stimuli *can* in certain situations be sufficient for triggering a compelling sensation of (illusory) self-motion, as is illustrated by the train illusion described above. This motivated us to investigate how far we can get without moving the observer at all, and how using VR technology might allow to optimize self-motion perception compared to the traditionally used optokinetic drums displaying abstract black and white patterns (instead of a natural scene as in the train illusion example).

Recent studies demonstrated that vection can indeed be reliably induced and investigated using VR setups that used video-projection setups [20, 13, 31, 32]. Lowther and Ware [20], Palmisano [25], and Riecke et al. [28] showed, for example, that the ability of VR to provide stereoscopic cues and to display naturalistic scenes instead of more abstract geometrical patterns can enhance vection reliably. Multi-modal contributions to vection have, however, received only little attention in the past. A noteworthy exception is the study by Wong and Frost [44], which showed that circular vection can be facilitated if participants receive an initial physical rotation (“jerk”) that accompanies the visual motion onset. One could imagine that the physical motion – even though it did not match the visual motion exactly – nevertheless provided a qualitatively correct motion signal, which might have reduced the visuo-vestibular cue conflict and thus facilitated vection. More recently, Schulte-Pelkum et al. [35] and Riecke et al. [31] showed that simply adding vibrations to the participant’s seat and floor plate during the visual motion can also enhance the self-motion sensation of the otherwise stationary participants. Post-experimental interviews revealed that the vibration were often associated with an actual motion of the VR setup (which never happened), thus making the simulation more believable.

Even though the auditory modality plays a rather important role in everyday life when moving about, there has been surprisingly little research on the relation between auditory cues and induced self-motion sensations.

This is all the more striking as auditorily induced circular vection and nystagmus have been reported as early as 1923 [7] and later been replicated several times [12, 17, 22]. Lackner demonstrated, for example, that an array of speakers simulating a rotating sound field can indeed induce vection in blindfolded participants [17]. Only recently has auditory vection received more interest, and a small number of studies were able to induce auditory vection in at least some of the participants, both for rotational and translational motions [16, 18, 32, 33, 38, 41, 39, 40]. While most researchers used artificial sounds (e.g., pink noise) [16, 17, 33], Larsson et al. [18] and Riecke et al. [32] hypothesized that the nature or interpretation of the sound source might also be able to affect auditory vection. In line with their hypothesis, they were able to demonstrate that sound sources that are typically associated with stationary objects (so-called “acoustic landmarks” like church bells) are more effective in triggering auditory circular vection than artificial sounds like pink noise or sounds that normally stem from moving objects (e.g., footsteps). These results strongly suggest the existence of higher cognitive or top-down contributions to vection, as the interpretation or meaning associated with a sound source affected the illusion. These results challenge the prevailing opinion that vection is mainly a bottom-up driven process. A more in-depth discussion of top-down and higher level influences on auditory as well as visual vection can be found in [32]. A similar benefit for using “acoustic landmarks” has recently been shown for translational vection [39]. Even non-spatialized sound was found to enhance vection if it resembled the sound of a vehicle engine [40].

Other factors that have been shown to facilitate auditory vection include the realism of the acoustic simulation and the number of sound sources [18, 32]. So far, though, there has been hardly any research on cross-modal contributions to auditory vection, and we are only aware of a study by Väljamäe et al. that showed that vibrations can enhance auditory vection [39], in line with experiments by Schulte-Pelkum et al. that showed a similar benefit of vibrations for visually-induced vection [35]. A comparable enhancement of auditory vection was observed when infrasound was added to the rotating sound sources (15Hz) [39].

Compared to visually induced vection, which is quite compelling and can even be indistinguishable from real motion [2], the auditory induced self-motion illusion is much weaker and less compelling. Furthermore, auditory vection occurs only in about 25-60% of the participants. Hence, even though auditory vection *can* occur, auditory cues alone are clearly insufficient to reliably induce a compelling self-motion sensation that could be used in applications. Therefore, the current study investigated whether *additional* spatial auditory cues can be utilized to *enhance* visually induced self-motion. Even though there is a large body of literature on visual vection, audio-visual interactions for vection have hardly if at all been investigated before.

Figure 1: *Top:* 360° roundshot photograph of the Tübingen market place, which was wrapped onto a cylinder to provide an undistorted view of the scene for the simulated viewpoint centered in the cylinder. *Bottom:* Participants were seated at a distance of about 1.8m from a curved projection screen (left) displaying a view of the market place (right).

Instead of using the classic black-and-white striped patterns as vection-inducing visual stimulus – which is not really suitable for VR applications – we opted here for using a naturalistic visual stimulus that has previously been shown to be quite powerful in inducing visual vection [28].

2 Hypotheses

Two main hypotheses on how adding auditory cues could potentially facilitate visual vection were investigated in the current study:

Hypothesis 1: Influence of adding non-spatialized auditory cues: First, one might imagine that there is a rather unspecific facilitation of vection by the auditory cues increasing the overall believability of the simulation and the resulting presence and involvement in the simulated scene, independent of the spatial content of the auditory cues. To address this issue, we compare a no-sound condition with a simple mono rendering of an auditory landmark in the scene (the sound of the fountain on the market place scene that was used as the visual stimulus).

Hypothesis 2: Influence of adding spatialized acoustic landmarks: Second, the spatial content of the auditory simulation could directly enhance vection by providing additional information about the spatial location of an acoustic landmark and hence the current orientation of the observer. This hypothesis was tested by comparing the above-mentioned mono-condition with a proper spatialized acoustic rendering of the correct location of the landmark using a generic head-related transfer function (HRTF). Furthermore, the simulation might appear more realistic in the spatialized condition, as the acoustic landmark should appear properly externalized and spatialized. This might also increase overall believability and presence in the simulated scene [11, 24, 38].

3 Methods

Twenty naive participants (eight male) took part in this experiment and were paid at standard rates[1]. All

---

[1] A subset of the experimental conditions with a smaller number of participants has previously been presented in an overview talk at the IEEE VR 2005 conference in Bonn [31].

participants had normal or corrected-to-normal vision and were able to locate the spatialized sound source without any problems.

3.1 Stimuli and Apparatus

Participants were comfortably seated at a distance of 1.8m from a curved projection screen (2m curvature radius) on which the rotating visual stimulus was displayed (see Fig. 1, bottom). The visual stimulus consisted of a photorealistic view of the Tübingen market place that was generated by wrapping a 360˚ roundshot (4096 × 1024 pixel) around a virtual cylinder (see Fig. 1, top). The simulated field of view (FOV) was set to 54˚×45˚and matched the physical FOV under which the projection screen was seen by the participants. Black curtains covered the side and top of the cabin surrounding the projection screen in order to increase immersion and block vision of the outside room. A force-feedback joystick (Microsoft force feedback 2) was mounted in front of the participants to collect the vection responses. Visual circular vection was induced by rotating the stimulus around the earth-vertical axis with alternating turning direction (left/right). Auditory cues were displayed using active noise-canceling headphones (Sennheiser HMEC 300) that participants wore throughout the experiment. Active noise cancellation was applied throughout the experiment to eliminate auditory cues from the surrounding room that could have interfered with the experiment.

In the spatialized auditory condition, a generic HRTF and a Lake DSP system (Huron engine) with multiscape rendering were used. Note that in the spatialized auditory condition, the fountain sound was always audible (as we have omni-directional hearing), even when the visual counterpart was outside of the current field of view. Participants perceived the spatialized fountain sound properly externalized and associated it readily with the visual counterpart as intended. None of the participants commented on any mismatch between the spatialized auditory cues and visual counterpart. In the mono sound condition, the sound was perceived “inside the head” as is to be expected for mono sound, and we are not aware that any participant experienced any ventriloquism effect in the sense that the moving visual stimulus might have created the illusion of a rotating sound.

3.2 Procedure and experimental design

Each participants performed 48 trials, consisting of a factorial combination of 3 auditory conditions (no sound, mono sound, HRTF-spatialized sound; these conditions were randomized within each session) × 2 turning directions (left/right; alternating) × 2 sessions × 4 repetitions of each condition. Participants were instructed to indicate the **onset of vection** by deflecting the joystick in the direction of perceived self-motion as soon as it was sensed. The amount of deflection indicated the **vection intensity**, and the time between vection onset and maximum vection (joystick deflection) reached indicated the **vection buildup time**. After each trial, participants indicated the **convincingness** of the perceived self-motion on a 0-100% rating scale (in steps of 10%) using a lever next to the joystick.

Participants started each trial by pressing a dedicated button on the joystick, which caused the static image to start rotating clockwise or counterclockwise (alternating, in order to reduce motion after-effects) around the earth-vertical axis with constant acceleration for 3s, followed by a constant velocity (30˚/s) phase. The maximum duration of constant velocity rotation was 46s, after which the stimulus decelerated at a constant rate for 3s. Stimulus motion stopped automatically once maximum joystick deflection (vection intensity) was sustained for 10s (otherwise it continued for 46s) to reduce the potential occurrence of motion sickness. Participants were asked to initiate each trial themselves to ensure that they could prepare for the next trial and paid attention to the stimulus[2].

Between trials, there was a pause of about 15 seconds to reduce potential motion aftereffects. In order to familiarize participants with the setup, a practice block containing 4 trials preceded the main experimental blocks. Furthermore, because none of the participants had experienced vection in the laboratory before, they were exposed, prior to beginning the practice block, to a vection stimulus for about 2 minutes or until they reported a strong sense of self-motion.

Overall between-subject differences in vection responses were removed using the following normalization procedure: Each data point per participant was divided by the ratio between the mean performance of that participant across all conditions and the mean of all participants across all conditions. In addition to the vection measures, **spatial presence** was assessed after the experiment using the Igroup Presence Questionnaire (IPQ) [34].

Participants were always instructed to watch the stimuli in a natural and relaxed manner, just as if looking out of the window of a moving vehicle. Furthermore, they were told to neither stare through the screen nor to fixate on any position on the screen (in order not to suppress the optokinetic reflex). Instead, they were instructed to concentrate on the central part of the projection screen.

4 Results

The vection data for the three sound conditions are summarized in Figure 2. The results of paired t-tests are indicated in the top inset of each plot. Adding mono sound increased the convincingness ratings slightly but insignificantly by about 10%. All other vection measures showed no difference between the no sound and mono sound condition.

---

[2] This procedure is not uncommon in psychophysical studies and implies that they might have been able to anticipate vection. We are, however, not aware of any study showing that this anticipation has any detrimental effect on the resulting data. If anything, we would rather expect that it might reduce the within-subject variability or random noise, as participants could start the next trial when they were ready for it and focusing on the stimulus to be presented.

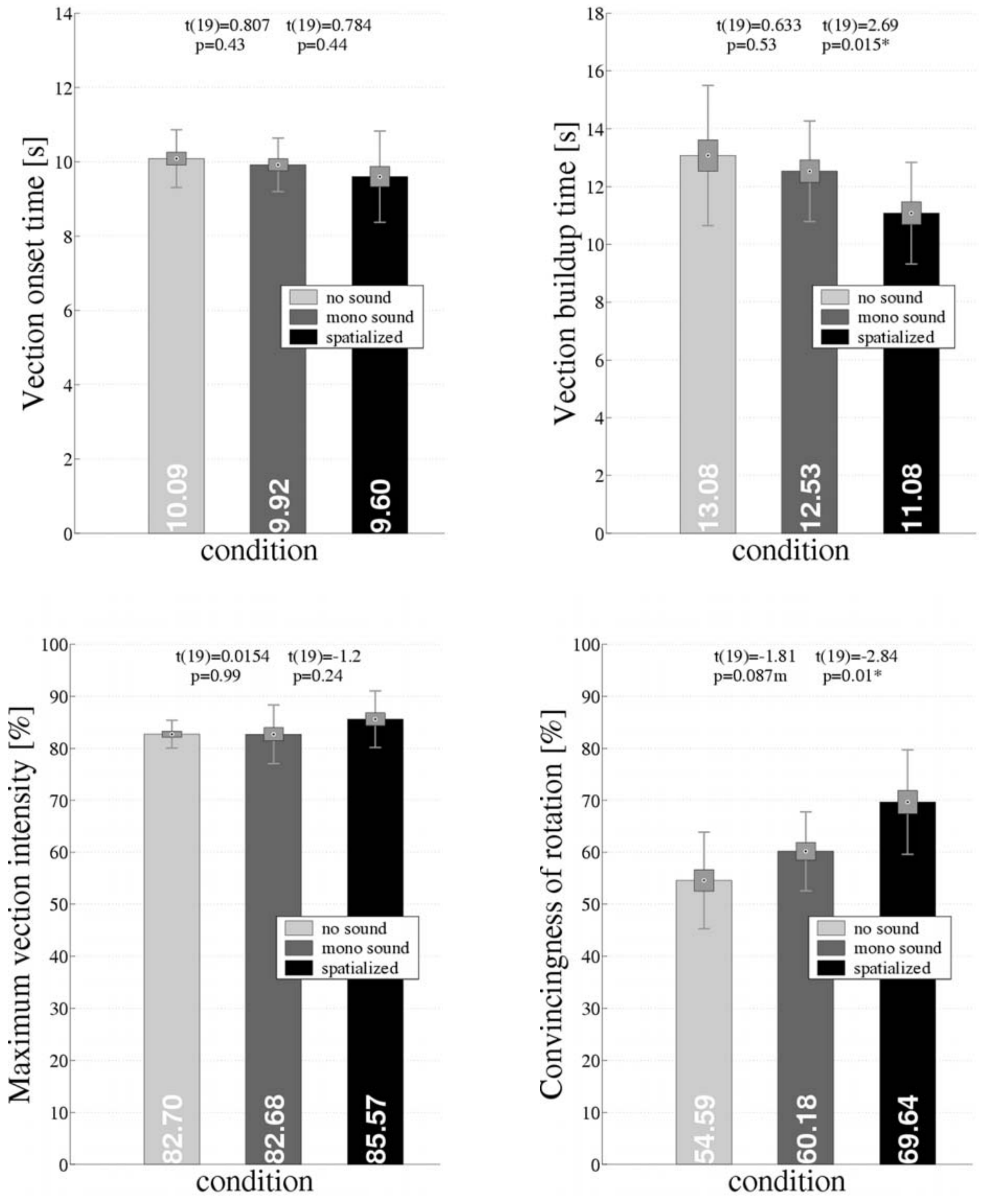


Figure 2: Mean of the four vection measures, averaged over the 20 participants. Boxes indicate one standard error of the mean, whiskers depict one standard deviation. The results of pairwise comparisons between the three sound conditions using paired t-tests are indicated in the top inset of each plot. An asterisk '*' indicates that the two conditions differ significantly from each other on a 5% level, an 'm' indicates that the difference is only marginally significant ($p < 0.1$). Note the small but consistent vection-facilitating effect of the proper spatialized auditory rendering of the fountain sound (right bars) as compared to simple mono display (middle bars). There were no significant differences between using mono sound and no sound at all.

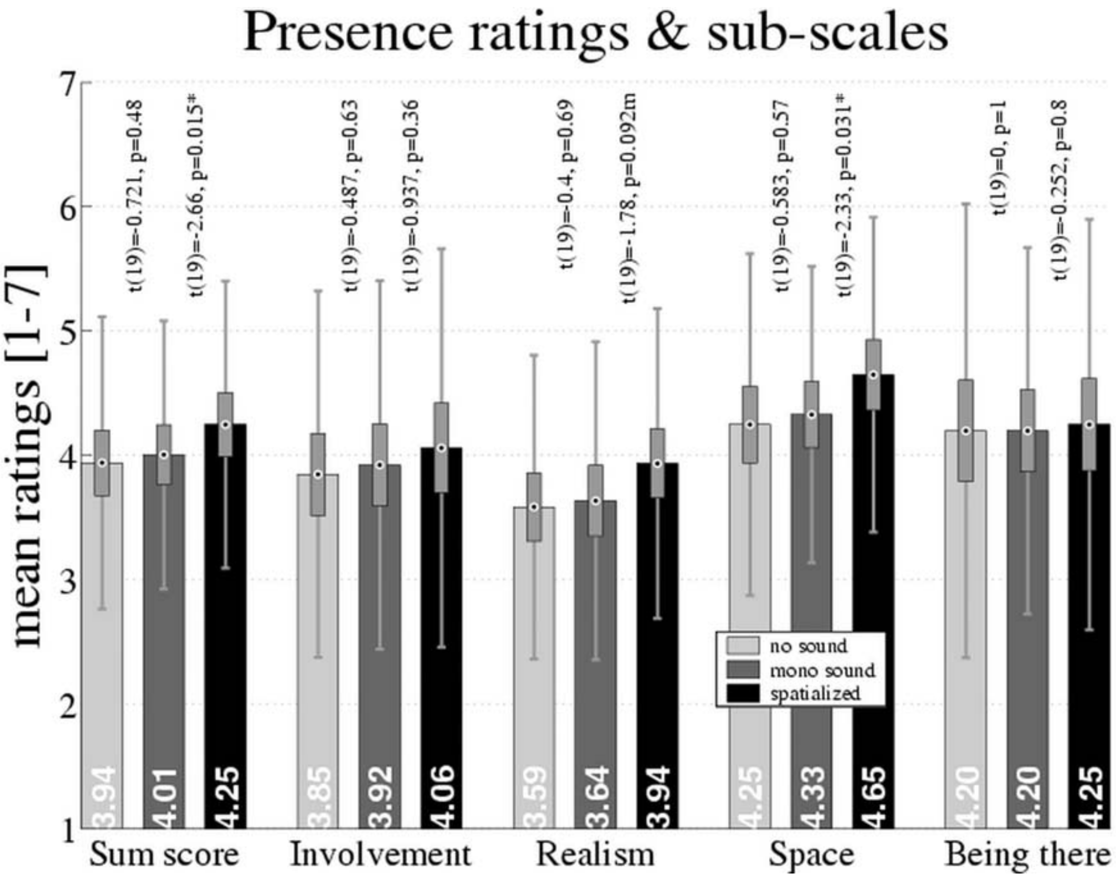


Figure 3: **Presence ratings for the three sound conditions. The sum score over all 14 items of the Igroup Presence Questionnaire (left three bars) were split up according to the four original sub-scales described by Schubert el al. [34]: "Involvement", "realism", "space", and "being there". Even though the effect size was quite small (<6%), the presence ratings were consistently higher for the spatialized sound condition.**

Comparing the mono condition with the spatialized sound condition demonstrates, however, a small but consistent vection-facilitating effect of the sound spatialization. The strongest effect was observed for the convincingness ratings (16% increase) and the vection buildup time (12% decrease). The other vection measures show only small and insignificant effects, albeit in the correct direction.

A similarly small, but consistent advantage for the spatialized sound can be observed for the presence ratings, which are summarized in Figure 3. This effect reached significance for the presence sum score and the "space" sub-scale. In addition, the "realism" sub-scale showed a marginally significant effect. The other presence sub-scales did not show any significant effects.

5 Discussion

Even though adding mono sound increased (insignificantly) the convincingness of the motion simulation by about 10%, neither the presence ratings nor any of the other vection measures were affected. That is, merely adding an audio cue that is associated with the fountain on the market place but not spatially aligned with it did not increase vection or presence significantly. This argues against an unspecific benefit of just adding audio cues. Only when the sound source was actually perceived to originate from the same location as its visual counterpart did we observe a significant facilitation of both vection and presence, which argues for a *specific* facilitation due to the spatialization of the sound source. This indicates that cross-modal consistency is indeed an important factor in improving VR simulations. This is all the more relevant as most existing VR simulations have rather poor audio quality, especially in terms of localizability of the sound sources (and externalization if headphone-based auralization is used).

As this study demonstrated, adding HRTF-based auralization using headphones can reliably be used to improve self-motion perception as well as presence in VR, even when the visual rendering is already of high quality and realism. This has many practical advantages, especially for applications where speaker arrays are unsuitable or where external noise must be excluded.

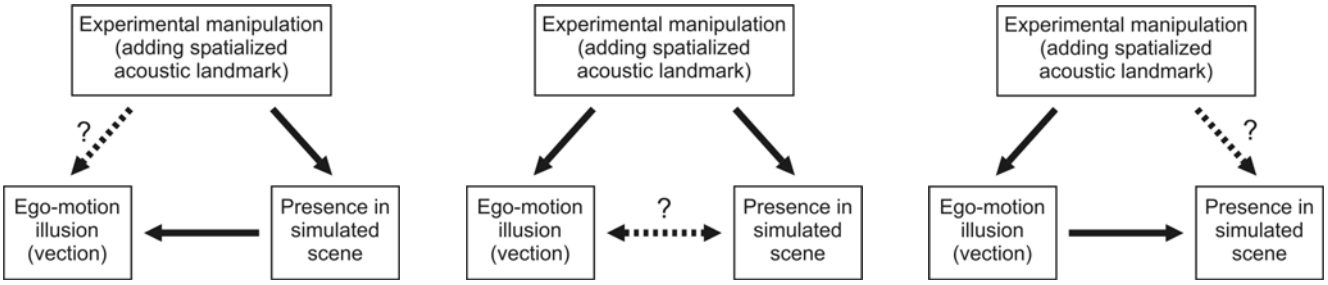


Figure 4: Schematic illustration of potential causal relations between adding the acoustic landmarks and the resulting facilitation of both vection and presence, as described in the text.

From the current data, it is, however, unclear whether there might also be a *causal* relationship or mediation between presence and vection, as is illustrated in Figure 4. On the one hand, it is conceivable that the observed increase in self-motion sensation might be mediated by the increase in presence (cf. Fig. 4, left).

A study by Riecke et al. [28] on visually induced circular vection suggests that an increase in presence might indeed be able to enhance vection: As an attempt to indirectly manipulate spatial presence without altering the physical stimulus properties too much, a photorealistic view onto a natural scene (just like in the current experiment) was compared to several globally inconsistent visual stimuli that were generated by scrambling image parts in a random manner. Thus, the stimulus could no longer be perceived as a globally consistent three-dimensional scene, which was expected to decrease spatial presence. The data showed both a decrease in presence and in vection for the globally inconsistent, scrambled stimuli. The authors suggest that higher-level factors like spatial presence in the simulated scene, global scene consistency, and/or consistent pictorial depth cues might have mediated the change in self-motion perception.

On the other hand, it is also feasible that an increase in the self-motion sensation might in some situations also be able to enhance overall presence and involvement (cf. Fig. 4, right), as suggested by Riecke et al. [27] and discussed in more detail in [30]. This seems sensible, as actual self-motions in the real world are typically accompanied by a corresponding sensation of self-motion. Hence, if self-motions simulated in VR are unable to evoke a natural percept of self-motion, the overall believability of the VR simulation and presence in the virtual environment in particular might also be affected.

In the long run, a deeper understanding of any potential causal relations between presence and the effectiveness of a simulation for a given task or goal (here: self-motion perception) would be rather helpful for optimizing VR simulations from a perceptual point of view. Further, carefully designed experiments are, however, required to tackle these issues.

In the debriefing after the experiment, participants rated the motion simulation as much more convincing when the spatialized sound was included. Nevertheless, the effect size of adding spatialized sound was rather small, both in terms of vection and rated presence. We propose two potential reasons here. First, it might reflect a ceiling effect, as the visually induced vection was already quite strong and showed relatively low onset latencies without the auditory cues. Second, auditory cues are known to be far less powerful in inducing vection than visual cues, which might explain the small effect size. Hence, we would expect a larger benefit of adding spatialized auditory cues if the auditory and visual vection inducing potential were equated in terms of their effect strength. On the one hand, the vection-inducing potential of the auditory cues could probably be increased by using more sound sources and rendering acoustic reflections and later reverberations in the simulated scene properly [18]. On the other hand, one could try to reduce the vection-inducing potential of the visual cues to a level comparable to the auditory cues by degrading the visual stimulus or by reducing the visual field of view. According to the latter, we would predict that the benefit of adding spatialized sound to VR simulations should be highest for low-cost simulators with poor image quality and/or a small field of view. Further experiments are currently being performed to investigate these hypotheses.

Apart from a specific vection-enhancing effect, adding spatialized auditory cues to VR simulations can have a number of further advantages, as is discussed in more detail in [19, 42, 37]: Adding auditory cues is known to increase presence in the simulated world, especially if spatialized auditory cues are used that are perceived as properly externalized and can be well localized, for example by using individualized HRTFs [11, 24, 38]. This is in agreement with the observed presence-facilitating effect of spatialized auditory cues in the current study. Furthermore, auditory cues provide the advantage of extending the perceivable virtual space beyond the limits of the visual field of view of the setup. This makes auditory cues perfectly suited for warning signals or for guiding attention. The omni-directional characteristics of human hearing enables us to get also a decent impression of the size and layout of a (real or simulated) scene without the need to turn our head and face the direction or object of interest

[26]. In general, whenever the corresponding situation in the real world would be accompanied with specific sounds, one would probably expect to hear those sounds in VR, too. This is of particular importance for achieving high perceptual realism in specific applications like driving and flight simulations, where adding appropriate engine sounds or environmental sounds is of crucial importance. One of the most frequent usages of audition is probably due to its clear potential to elicit emotional responses, a fact that is well-known and frequently employed by, for example, the movie industry. Last but not least, including auditory cues can also be particularly important for people who's preferred modality or cognitive style is auditory (and not visual or kinesthetic).

Hence, adding spatialized auditory cues to (predominately visual) VR simulations and ego-motion simulations in particular can have a number of advantages including an increase in the perceived self-motion. Relatively little research has been performed in this area, and additional studies are required to investigate these issues further. It is conceivable, however, that the requirements for visual rendering quality could be relaxed when appropriate simulation of the auditory modality (and potential other modalities) is provided [9]. As high quality auditory rendering can be achieved at relatively low cost, adding spatialized auditory cues might allow us in the future to increase simulation effectiveness while reducing the overall simulation effort, especially when the attention guiding potential of auditory cues is employed. Using a selective rendering approach, guiding attention has, for example, been shown to reduce computational costs of the visual rendering considerably [4, 36]. This is promising for the usage of auditory cues for optimizing VR simulations both on a computational and perceptual level.

Acknowledgements

This research was funded by the European Community (IST-2001-39223, FET Proactive Initiative, project "POEMS" (Perceptually Oriented Ego-Motion Simulation, see www.poems-project.info) and the Max Planck Society.

References

[1] E. R. Boer, A. R. Girshik, T. Yamamura, and N Kuge. Experiencing the same road twice: A driver-centred comparison between simulation and reality. In *Proceedings of the Driving Simulation Conference 2000*, Paris, France, 2000.

[2] T. Brandt, J. Dichgans, and R. Held. Optokinesis affects body posture and subjective visual vertical. *Pflugers Archiv-European Journal of Physiology*, 339:97–97, 1973.

[3] J. Burki-Cohen, T. H. Go, W. Y. Chung, J. Schroeder, S. Jacobs, and T. Longridge. Simulator fidelity requirements for airline pilot training and evaluation continued: An update on motion requirements research. In *Proceedings of the 12th International Symposium on Aviation Psychology, April 14-17*, pages 182–189, Dayton (OH), USA, 2003.

[4] K. Cater, A. Chalmers, and G. Ward. Detail to attention: Exploiting visual tasks for selective rendering. In *EGRW '03: Proceedings of the 14th Eurographics Workshop on Rendering*, pages 270–280, 2003.

[5] S. S. Chance, F. Gaunet, A. C. Beall, and J. M. Loomis. Locomotion mode affects the updating of objects encountered during travel: The contribution of vestibular and proprioceptive inputs to path integration. *Presence - Teleoperators and Virtual Environments*, 7(2):168–178, April 1998.

[6] J. Dichgans and T. Brandt. Visual-vestibular interaction: Effects on self-motion perception and postural control. In R. Held, H. W. Leibowitz, and H.-L. Teuber, editors, *Perception*, volume VIII of *Handbook of Sensory Physiology*, pages 756–804. Springer, 1978.

[7] R. Dodge. Thresholds of rotation. *Journal of Experimental Psychology.*, 6:107–137, 1923.

[8] F. H. Durgin and A. Pelah. Self-motion perception during locomotor recalibration: More than meets the eye. *Journal of Experimental Psychology: Human Perception and Performance*, in press.

[9] N. I. Durlach and A. S. Mavor, editors. *Virtual reality: Scientific and technological challenges*. National Academy Press, 1995.

[10] M. H. Fischer and A. E. Kornmüller. Optokinetisch ausgelöste Bewegungswahrnehmung und optokinetischer Nystagmus [Optokinetically induced motion perception and optokinetic nystagmus]. *Journal für Psychologie und Neurologie*, pages 273–308, 1930.

[11] C. Hendrix and W. Barfield. The sense of presence within auditory virtual environments. *Presence – Teleoperators and Virtual Environments*, 5(3):290–301, 1996.

[12] P. E. Hennebert. Audiokinetic nystagmus. *Journal of Auditory Research*, 1(1):84–87, 1960.

[13] L. J. Hettinger. Illusory self-motion in virtual environments. In Kay M. Stanney, editor, *Handbook of Virtual Environments*, chapter 23, pages 471–492. Lawrence Erlbaum, 2002.

[14] J. M. Hollerbach. Locomotion interfaces. In Kay M. Stanney, editor, *Handbook of Virtual Environments*, chapter 11, pages 239–254. Lawrence Erlbaum, 2002.

[15] I. P. Howard. The perception of posture, self motion, and the visual vertical. In K. R. Boff, L. Kaufman, and J. P. Thomas, editors, *Sensory processes and perception*, volume 1 of *Handbook of human perception and performance*, pages 18.1–18.62. Wiley, New York, 1986.

[16] B. Kapralos, D. Zikovitz, M. Jenkin, and L.R. Harris. Auditory cues in the perception of self-motion. In *Proceedings of the 116th AES convention, Berlin 2005*, Berlin, Germany, 2004.

[17] J. R. Lackner. Induction of illusory self-rotation and nystagmus by a rotating sound-field. *Aviation Space and Environmental Medicine*, 48(2):129–131, 1977.

[18] P. Larsson, D. Västfjäll, and M. Kleiner. Perception of self-motion and presence in auditory virtual environments. In *Proceedings of Seventh Annual Workshop Presence 2004*, pages 252–258, 2004. Available: www.kyb.mpg.de/publication.html?publ=2953.

[19] P. Larsson, A. Väljamäe, D. Västfjäll, and M. Kleiner. Auditory-induced presence in mediated environments and related technology. In *Handbook of Presence*. Lawrence Erlbaum, 2005. (submitted).

[20] K. Lowther and C. Ware. Vection with large screen 3d imagery. In *ACM CHI '96*, pages 233–234, 1996.

[21] E. Mach. *Grundlinien der Lehre von den Bewegungsempfindungen*. Leipzig, Germany: Engelmann, 1875.

[22] A. M. Marmekarelse and W. Bles. Circular vection and human posture ii: Does the auditory-system play a role. *Agressologie*, 18(6):329–333, 1977.

[23] M. Mulder, M. M. van Paassen, and E. R. Boer. Exploring the roles of information in the control of vehicular locomotion - from kinematics and dynamics to cybernetics. *Presence - Teleoperators and Virtual Environments*, 13:535–548, 2004.

[24] K. Ozawa, Y. Chujo, Y. Suzuki, and T. Sone. Psychological factors involved in auditory presence. *Acoustical Science and Technology*, 24:42–44, 2004.

[25] S. Palmisano. Consistent stereoscopic information increases the perceived speed of vection in depth. *Perception*, 31(4):463–480, 2002.

[26] J. Pope and A. Chalmers. Multi-sensory rendering: Combining graphics and acoustics. In *Proceedings of the 7th International Conference in Central Europe on Computer Graphics*, pages 233–242, Czech Republic, 1999.

[27] B. E. Riecke, J. Schulte-Pelkum, M. N. Avraamides, and H. H. Bülthoff. Enhancing the visually induced self-motion illusion (vection) under natural viewing conditions in virtual reality. In *Proceedings of Seventh Annual Workshop Presence 2004*, pages 125–132, 2004. Available: www.kyb.mpg.de/publication.html?publ=2864.

[28] B. E. Riecke, J. Schulte-Pelkum, M. N. Avraamides, M. von der Heyde, and H. H. Bülthoff. Influence of scene consistency and spatial presence on circular vection in virtual reality. In *ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization* (APGV), 2005. (accepted).

[29] B. E. Riecke, J. Schulte-Pelkum, and H. H. Bülthoff. Perceiving simulated ego-motions in virtual reality - comparing large screen displays with HMDs. In *SPIE - Invited paper on VALVE: Vision, Action, and Locomotion in Virtual (and Real) Environments*, San Jose, CA, USA, 2005. Available: www.kyb.mpg.de/publication.html?publ=3233.

[30] B. E. Riecke, J. Schulte-Pelkum, and Franck Caniard. Using the perceptually oriented approach to optimize spatial presence & ego-motion simulation. In *Handbook of Presence*. Lawrence Erlbaum, 2005. (submitted).

[31] B. E. Riecke, J. Schulte-Pelkum, Franck Caniard, and H. H. Bülthoff. Towards lean and elegant self-motion simulation in virtual reality. In *Proceedings of IEEE VR2005*, pages 131–138, Bonn, Germany, 2005. www.vr2005.org.

[32] B. E. Riecke, Daniel Västfjäll, Pontus Larsson, and J. Schulte-Pelkum. Top-down and multi-modal influences on self-motion perception in virtual reality. In *HCI international 2005 (accepted)*, Las Vegas, NV, USA, 2005. www.hci-international.org.

[33] S. Sakamoto, Y. Osada, Y. Suzuki, and J. Gyoba. The effects of linearly moving sound images on self-motion perception. *Acoustical Science and Technology*, 25:100–102, 2004.

[34] T. Schubert, F. Friedmann, and H. Regenbrecht. The experience of presence: Factor analytic insights. *Presence - Teleoperators and Virtual Environments*, 10(3):266–281, June 2001.

[35] J. Schulte-Pelkum, B. E. Riecke, and H. H. Bülthoff. Vibrational cues enhance believability of ego-motion simulation. In *International Multisensory Research Forum (IMRF)*, 2004. Available: www.kyb.mpg.de/publication.html?publ=2766.

[36] V. Sundstedt, K. Debattista, and A. Chalmers. Selective rendering using task-importance maps. In *ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization (APGV)* (p. 175), 2004.

[37] A. Väljamäe, A. Kohlrausch, S. van de Par, D. Västfjäll, P. Larsson, and Kleiner M. Audio-visual interaction and synergy effects: implications for cross-modal optimization of virtual and mixed reality applications. In *Handbook of Presence*. Lawrence Erlbaum, 2005. (submitted).

[38] A. Väljamäe, P. Larsson, D. Västfjäll, and Kleiner M. Auditory presence, individualized head- related transfer functions, and illusory ego-motion in virtual environments. In *Proceedings of Seventh AnnualWorkshop Presence 2004*, pages 141–147, Valencia, Spain, 2004. Available: www.kyb.mpg.de/publication.html?publ=2954.

[39] A. Väljamäe, P. Larsson, D. Västfjäll, and Kleiner M. Effects of vibratory stimulation on auditory induced self-motion. In *Proceedings of IMRF 2005*, Rovereto, Italy, 2005. Poster presented at IMRF 2005; Available: www.kyb.mpg.de/publication.html?publ=3379.

[40] A. Väljamäe, P. Larsson, D. Västfjäll, and Kleiner M. Sonic self-avatar and self-motion in virtual environments. In *Proceedings of the 8th Annual Workshop of Presence, London 2005*, London, England, 2005. (submitted).

[41] A. Väljamäe, P. Larsson, D. Västfjäll, and Kleiner M. Travelling without moving: Auditory scene cues for translational self-motion. In *Proceedings of ICAD 2005*, Limerick, Ireland, 2005.

[42] A. Väljamäe, D. Västfjäll, P. Larsson, and Kleiner M. Perceived sound in mediated environments. In *Handbook of Presence*. Lawrence Erlbaum, 2005. (submitted).

[43] R. Warren and A. H. Wertheim, editors. *Perception & Control of Self-Motion*. Erlbaum, New Jersey, London, 1990.

[44] S. C. P. Wong and B. J. Frost. The effect of visual-vestibular conflict on the latency of steady-state visually induced subjective rotation. *Perception & Psychophysics*, 30(3):228–236, 1981.

Neural processing of Spatial Information: What we know about Place cells and what they can tell us about Presence

Jorge Brotons [1,2], Shane O'Mara [2] and Maria V. Sanchez-Vives [1]
[1] Instituto de Neurociencias de Alicante, Universidad Miguel Hernández-CSIC, San Juan de Alicante, Spain.
[2] Institute of Neuroscience, Trinity College, Dublin, Ireland.
jbrotons@umh.es; smomara@tcd.ie; mavi.sanchez@umh.es

Abstract

Brain processing of spatial information is a very prolific area of research in neuroscience. Since the discovery of place cells (PCs)[1] researchers have tried to explain how these neurons integrate and process spatial and non-spatial information. Place cells (PCs) are pyramidal neurons located in the hippocampus and parahippocampal region which fire with higher frequency when the animal is in a discrete area of space. Recently, PCs have been found in the human brain. The processing of spatial information and the creation of cognitive maps of the space is the result of the integration of multisensory external and internal information and the brain's own activity. In this article we review some of the most relevant properties of PCs and how this knowledge can be extended to the understanding of human processing of spatial information and to the generation of spatial presence.

Keywords--- ***Hippocampus, Place cells, Subiculum, Spatial Processing, Spatial Presence****.*

1. Introduction.

Spatial navigation is a fundamental form of interaction with the environment. Animals and humans must move about in their environments in search for food, shelter or mate, actions which are basic for the survival of the individual and the species. The brain in different species has evolved in an effort to make individuals capable of navigating their environments in an efficient manner. The understanding of the brain mechanisms underlying the generation of internal maps of the external world, the storage (or memory) of these maps, and the use of them in the form of navigation strategies is the field of study of a large number of researchers in the neuroscience community. On the other hand, the study of navigation in real and virtual environments (VE) has been a broad field of study, including a diverse range of topics from model city design to the generation of VEs that successfully result in spatial presence and that are optimal for the transfer of spatial information between virtual and real worlds.

In this article we review data (including our own) on the neural basis of spatial navigation, mostly centered on hippocampal and parahippocampal neurons called "place cells" that are specialized in responding to spatial position. The functional properties of these neurons embody many aspects of human navigation that are well known from a behavioral point of view. It is our purpose to demonstrate that the understanding of the neuronal basis of spatial processing is relevant to the understanding and successful generation of spatial presence. Furthermore, we will suggest that the similar activation of brain structures during navigation in virtual *compared to* real worlds can be in itself an objective measurement of presence. In other words if place cells activation occurs in the same way in a virtual environment (VE) as it does in a physical environment then this is one level of evidence, a very important one, that presence is occurring within that VE.

In Section 2 we review general mechanisms and strategies of navigation and the underlying brain structures that control them. We go on to center our attention on the best known structure that codes for spatial information (Section 3), the hippocampus and parahippocampal region. Its anatomical structure is briefly described, as well as the properties of one of the most prominent electrophysiological signatures of this region, the "theta rhythm". This rhythm is important because it synchronises activity within the hippocampal formation and it affects the firing of place cells. For this reason it has been repeatedly implicated in integrative functions related to the navigation tasks –i.e. sensory-motor integration-, and therefore it is worth mentioning. Once the general framework for investigating place cells has been described, we go on to explain their specific functional properties, with an emphasis on the factors that determine their spatial firing fields (location, visual or other sensory cues, behavioral relevance of the area, etc) and the involvement of other areas of the brain in other relevant aspects of navigation, such as place significance or reward. These functional properties that are studied at the cellular level are supposed to support many of the well known features of navigation and their understanding results in the knowledge of the elements that could induce spatial presence. Based on that knowledge, in Section 4 we review relevant aspects of place cells and we suggest how this information could be useful to the understanding on how the brain processes spatial information in VR. To expand on how this could be relevant to presence research, we suggest some empirical experiments and predictions based on observations made in place cells.

2. Spatial navigation in animals and humans.

Species varying from migratory birds to humans need to utilize different information to generate knowledge of environments to navigate successfully. O'Keefe and Dostrovsky [1] suggested that the hippocampus was the central brain structure implicated in spatial navigation and the neuronal substrate in which a "cognitive map" of the external environment is created. A "cognitive map" is an internal representation of an environment that allows subjects to choose the best way to get to an objective by making calculations based on the relations between different environmental landmarks. Other strategies could be used by humans and animals in an effort to navigate such as egocentric navigation (see below), and these route or 'taxon'-based strategies depend on non-hippocampal brain systems.

Birds with hippocampal lesions can navigate during migration using a compass strategy, following a fixed direction, but they get lost in their local area because they are not capable of generating a cognitive map of the area [2]. Classic studies of migratory birds shed light on the strategies of these expert navigators to make use of different types of available information to orient themselves throughout long distances in their migratory flights or in their short trips in search for food. [3] showed that if naïve migratory birds in their first flight were captured and transported in a perpendicular direction to that which they were directed they would miss the final destiny by the amount of kilometers they were transported. These birds were flying towards a fixed goal using a compass strategy [4]. On the other hand if the same procedure was implemented in experienced birds, these would correct the distance they were transported, reaching successfully the final goal. Experienced birds use a more elaborate approach to navigation involving knowledge of the environment. Therefore, cognitive mapping would depend on experience and learning, ruling out the possibility of instinctive knowledge of migratory routes. While using a compass strategy birds can use three different sources of information, the sun, geomagnetism and the stars [5]. Experiments which have manipulated the internal (circadian) clock of birds have demonstrated that they use the sun to orientate themselves with respect to their internal clock [6]. Animals use also geomagnetism to orientate themselves and by applying magnets in the head of the animals, they can be redirected towards a specific direction if the skies are overcast. More recently, some studies have demonstrated that pigeons, while flying to their nests, can also use highways and their exits as cues using compass adjustment during the middle part of the fly and a cognitive map when approaching the loft area [7].

No evidence has been found in the human brain of magnetic sensors contributing to spatial orientation. However, there are recent advances in the understanding of the cellular networks underlying human navigation by means of single neuron recordings in implanted patients [8] and fMRI studies [9] in virtual environments.

2.1. Allocentric and egocentric navigation

Two basic navigational strategies allow animals and humans to navigate successfully:

1) *Allocentric navigation* enables humans and animals to generate a internal representational system based on the global coordinates of the environment. Thus, a topographical representation of the environment is generated by using multiple relevant landmarks of their surroundings. These external cues are used to establish a complex representation which would include the distance between them and to the subject´s own relative position. This facilitates a precise navigation to specific goals even if those are not visible. For example it has been proved that rats are able of swimming to a hidden platform which allows them to escape from a pool. To achieve this kind of successful navigation rats had to use multiple environmental cues available in the experimental room which allow them to generate a map-like representation of the environment [10]. Subsequently, studies were performed on experimental groups that were trained to navigate in the watermaze searching for the hidden platform using external cues, while having different degrees of damage in the parahippocampal region [11]. Those studies revealed that rats had a strong degree of impairment to find the hidden platform but not in the visible platform version of the task. The results of these two experiments would suggest that these different structures of the parahipocampal complex and the hippocampus are necessary for the allocentric navigation strategy [12] supporting the original suggestion [1]. Similar evidence has been found in humans, whose hippocampus and parahipocampal region appeared activated in fMRI studies in which subjects navigated in a virtual environment [13, 14]. The activation of these structures followed a different pattern depending on the type of navigation, wayfinding or route following [9].

2) The *egocentric navigation* implies using other available information such as internal cues, motor input, vestibular and directional information. All these sources of information allow the subject to calculate its present and its future position by summing all different movements and turns, also called *path integration*. The hippocampus [15] and other areas such as the parietal cortex seem to be involved [15-17]. The parietal cortex and other structures are involved in path integration while humans navigate in a virtual reality environment using a route strategy [18].

The egocentric strategy would be the dominant while navigating in situations in which there is no allocentric information available, for example while navigating in the darkness.

3. Neuronal substrates of spatial navigation.

So far we have briefly described some of the neuronal bases of spatial navigation and the two basic strategies that are used during navigation. To better understand how the brain integrates spatial information it is necessary to briefly describe the anatomy of the hippocampus, its physiology, and the functionality of hippocampal place cells.

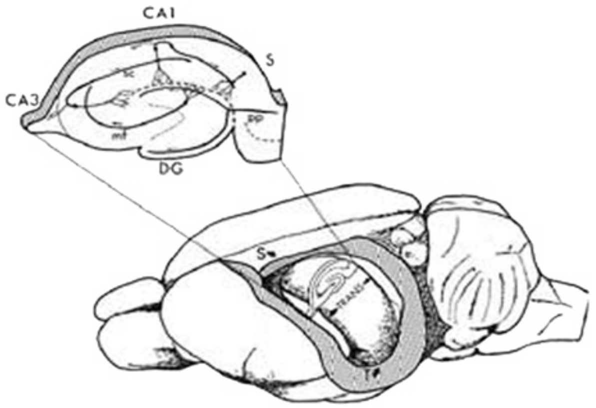


Figure 1. Anatomy of the rat hippocampus

3.1. The anatomy of the hippocampus.

A distinction between the hippocampus proper and the hippocampal region must be made. The hippocampus proper consists of two interlocked cell layers with the shape of a C consisting of the dentate gyrus and the cornus ammonis comprising areas CA1 and CA3, the two main subfields. The parahippocampal region comprises the entorhinal cortex, the periallocortical area of the perirhinal area, the subicular complex, presubiculum, parasubiculum and subiculum [19].

3.2. Electrophysiology of the hippocampus.

It has been suggested that the major electrophysiological activity involved in sensory and motor integration is hippocampal theta rhythm [for a review see 20]. Theta activity is characterized by a regular sinusoidal activity between 4-8 Hz. Its changes in amplitude and frequency are directly related to sensory inputs reflecting changes in any sensory pathway and also changes in motor behaviour [20-22]. The fact that hippocampal PCs firing is related to the theta rhythm [23] strengthens the idea that this sensory and motor integration process conveys at least some of the essential information required for spatial navigation. Theta rhythm has been detected in humans while navigating a virtual maze [24], being related to the difficulty of the maze [25]. This rhythm appears to be dissociated from other components of the task, being associated with navigation [26]. Nevertheless, some authors find association of theta rhythm just with the motor act of exploring, but find no correlations between any theta characteristics and the cognitive demand of the tasks [27].

3.3. Place cells

O´Keefe and Dostrovsky [1] recorded single neurons in the hippocampus from chronically implanted rats foraging freely for food in a small arena. They described a group of cells whose firing increased whenever the animal was in a discrete location of the environment and this location was called the “firing field” (FF) of that particular neuron.

The firing of these neurons seemed to be independent of other variables such as view, direction or speed of movement; location or position was the best predictor of their firing [28]. Subsequent research has supported the original finding and PCs were seen as the first objective measurable neuronal basis of an advanced or higher-order cognitive process. The study of PCs has generated a broad body of investigation and research but initially they were only recorded in rodents, and proved difficult to detect in primates [29, 30] until recently [31]. It was questionable if this same mechanism would be also present in the human brain. Lately, recordings from subcortical implanted electrodes in epileptic patients revealed that cells in the human hippocampus fire strongly in specific locations while the subject navigated a virtual environment [8], thus proving the existence of PCs in humans. Furthermore, it is evident that human hippocampal formation is strongly activated virtual navigation and exploration using brain imaging [9, 32].

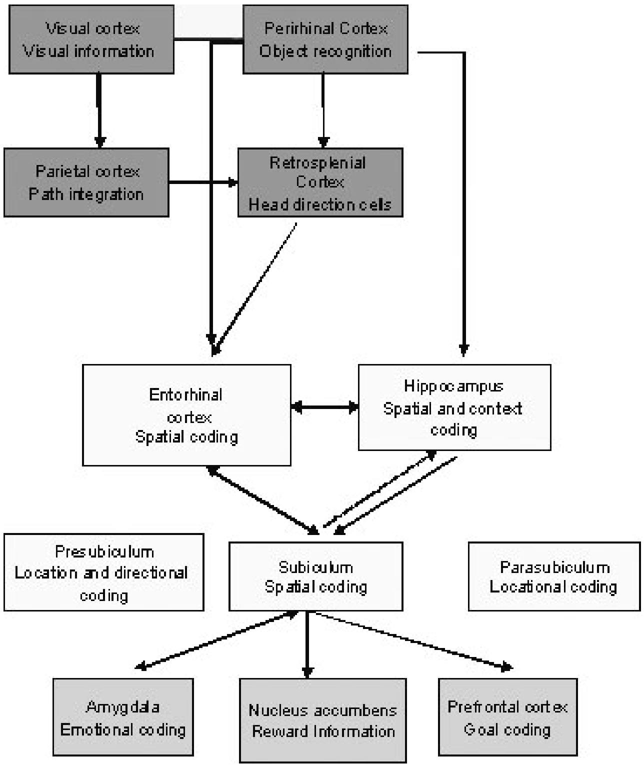


Figure 2. Simplifed anatomical afferent and efferent connections between the hippocampus and brain areas that are relevant to spatial processing, including some of their attributed functions.

Standard methods for studying the spatial selectivity of hippocampal formation neurons require freely-moving rats to traverse mazes or open-fields (sometimes foraging for food); neuronal activity is recorded and correlated with the rats' moment-to-moment position, from which colour-coded contour maps are generated (representing normalized/averaged spike firing density at all points occupied by the rat; see Figs. 3,4). Different parameters have been studied to better understand how PCs code spatial information among which we can highlight stability, directionality of PCs firing, sensory information and cue control of PCs firing.

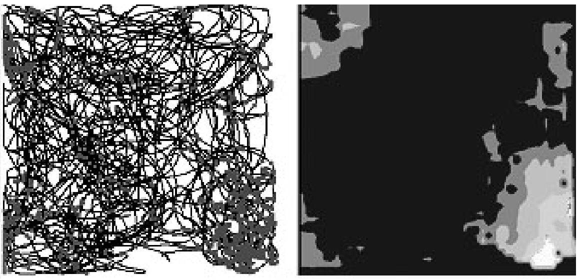


Figure 3. Recording from PCs. Left, animal tracking signal and spike firing. Right, firing density according to position or firing field of a particular neuron, located in the lower right corner.

3.3.1. Place cell stability. Stability of place cells, or the opposite, plasticity of place cells are relevant to the understanding of remapping of space when we enter a virtual world. How stable are the maps coded by PCs? PCs tend to fire in a stable manner if no spatial or other manipulation is implemented in the environment. Thompson and Best [33] reported a neuron whose firing field was stable over 153 days of recording, using the same recording arena. Hill [34] suggested that PCs firing fields (FF) are generated as quickly as the animal explores the environment. Subsequent studies have demonstrated that PCs learn to code salient cues in the environment. Thus, hippocampal PCs can generate a progressive differential representation of two different arenas [35]. Although firing patterns were similar in both arenas at the beginning this spatial representation, they diverged after repetitive exposure. This new representation was stable one month later for each of the environments. Therefore, although FFs can be stable for long periods of time they also reflect spatial and neuronal plasticity. Indeed, blockade of NMDA receptors (involved in synaptic plasticity) impaired PCs firing stability in new environments [36]. However, other authors [37, 38] postulated that PCs firing depends not only on a learning process but it is relatively hard-wired in the hippocampus during brain development, a view that is challenged by data showing the great plasticity of place cells under appropriate circumstances.

3.3.2. Directionality of place cells. Although it is clear that PCs fire in relation to the animals' location [39] it was not clear if PCs also coded for the direction of the movement. It has been reported that the firing frequency of PCs was higher when the animal was running in an inward direction in a radial arm maze [40]. Later research suggested that PCs directional firing was related to the physical characteristics of the maze and to the task´s demands. Thus, directional firing of PCs was higher in the radial arm maze and also in an open field arena whenever the animal had to move in a linear track to retrieve a reward [41]. Taube et al [42, 43] described a type of cell whose firing coded for head direction (HDC) firing only whenever the animal head it is oriented to a specific direction. These type of cells are found in different structures of the parahipocampal complex as well as in other subcortical structures [44]. The firing of these neurons conveys information about where the animal's head is pointing. They seem to use environmental cues to calibrate their directional firing and they depend on vestibular input without which their firing disappears. A group of cells were found in the presubiculum with firing codes for location and direction [45].This type of cells could synthesize spatial information and direction information being the bridge between both systems.

3.3.3. Place cells and goal navigation. An efficient navigational system must be able to integrate the significance of a place in the cognitive map for efficient spatial navigation. It is not enough to know where you are but to know where you want to go [29]. O'Keefe and Nadel (1978) suggested in their model that PCs do not code for goals or hedonics aspects of navigation. On the other hand some authors have suggested that place cells have to do with the meaning of a place [46]. Speakman and O'Keefe [47] found that goal location changes did not affect the location of FFs in a radial arm maze, although prefrontal lesions do impair performance on this goal navigation task [48] suggesting that goal-related information might be located in prefrontal cortex.. The fact that FFs of place cells are stable while the forage for pellets of food thrown to random locations in the arena would suggest that they are not coding for the goal aspect of navigation [29]. However, when animals were trained to escape a watermaze using in a hidden platform, a strong concentration of PCs near the escape platform was found, suggesting that areas of space of behavioural significance could be over represented by the hippocampus [49].

Subiculum and nucleus accumbens' cells firing predicted reward administration and also coded for spatial location [50]. Similarly, PC firing changes due to task demands and that those changes correlated with efficient performance [51]. Gemmell and O'Mara [48] have suggested that the prefrontal cortex might be the central structure involved in goal coding during navigation. Likewise [52] found cells in the prefrontal cortex of the rat which they suggested coded for goal and it has been proved that prefrontal cells are able of differentiate between high and low frequency rewarded arms in the radial arm maze [53].

The amygdala is another structure involved in place preference learning [49, 54]. In humans, while hippocampal PCs code for location, neurons in the parahippocampal region as well as throughout the frontal and temporal lobes were found to respond to the subject's navigational goals and to conjunctions of place, goal and view [8]. We could summarize that hippocampal place cells are susceptible to changes in navigational tasks adapting the to new demands in relation to reward location changes. Also, hippocampal place cells could over-represent relevant areas of space. These changes could be due to integration of place significance in other areas of the brain such as the nucleus accumbens, amygdala and prefrontal cortex, areas which are all strongly interconnected with the hippocampus [55].

3.3.4. What does affect the firing of PCs? Environmental cues help animals and humans to make navigational decisions, to locate themselves and to calculate different trajectories to reach relevant goals [29]. How does multisensory information affect position coding? [56] introduced different manipulations of the recording arena to study the different effects on PCs. The recordings were carried out in a cylindrical arena with a cue card attached at the wall acted as a distal cue. Rotation of this visual cue produced a rotation of the FFs keeping the same angular relation as in the original configuration and removal of this cue card produced FF to rotate to unpredictable positions. However, manipulations of the cue size did not affect FFs. Placing a small barrier over the location of a previously recorded FFs was enough to make the FFs disappear. Doubling the size of the area and walls height produced that some cells expanded their FFs in relation to the new size although most cells generated new FFs, producing what has been called remapping [57, 58]. In the same way if the arena shape was changed from a cylinder to square, cells also remapped.

The removal of existing cues has different effects depending on the proximity of the cues [59]. It was found that removal of a cue proximal to the FF reduced the size of the FF, while removal of a distal cue would produce an enlargement of FF size. In [60] a visual cue was manipulated either when the animal was present or before he was placed in the recording arena. PCs did not rotate their FF if the cue was moved in their presence but if the cue was rotated while away then FF would also rotated. Rats learned to rely on egocentric information when the visual cue was not reliable. [61] placed objects centrally in the arena. They found that this configuration did not exert any control on the FF. On the contrary if these objects were placed against the walls of the arena then they were able of exerting control on FF. [62] rotated in different directions proximal and distal cues producing that some of the cells rotated with the distal cues and other with the proximal cues. Recordings in animals deprived of visual and auditory information revealed that the PCs of these animals were stable in despite of the lack of sensory input [63] [64-66]. It is then clear that a mechanism other than allocentric information is being used by the animals. PCs would be using some sort of path integration or egocentric information to keep their firing stable [38].

PCs could strongly respond to features of the environment such as barriers [35, 67, 68]. The recorded PCs while animals foraged for food in an open field in which a high barrier was located. "Barrier cells" would fire around this and their FF would move with the barrier if this was moved. The barrier would exert similar control if the animal was located in second new environment.

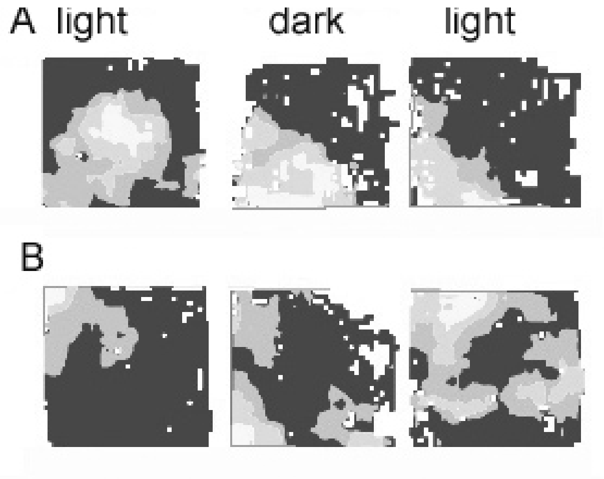


Figure 4. Firing fields of differing subicular place cells under different lighting conditions. Subicular firing fields are typically large. Firing fields (FFs) are optimally seen with a colour scale. A. Neuron with FF that remaps with a change in light condition and remains in the next location. B. Neuron whose firing field remaps from the top left corner to the left low corner in the dark and remaps again when back to light.

Previous research found that the geometry of the arena exert a quite strong control on PCs firing [67, 69]. Sharp [70] hypothesized that hippocampal place cells not only code spatial information but contextual spatial information. Place cells would be then modulated by geometric and non geometric changes in the environment. This would explain that subtle changes in context might generate extreme changes in the establishing firing field of a place cell and that geometric changes sometimes would not affect FFs that strongly. This hypothesis predicts that if place cells represent a unique spatial context then all place cells should remap under the different manipulations the experimenter could develop. There was, however, a high heterogeneity in the remapping of all different cells, leading to the conclusion that contextual information does not affect place cells in a whole block but in a fragmented way.

Similarly the fact that hippocampal PCs display different maps in different environments [71] could be seen as evidence that the hippocampus is coding spatial and non spatial aspects of the environmental context. On the contrary, subicular and entorhinal cortical PCs tend to represent different environments in similar ways [70]. However, in our laboratory we recorded place cells in a square arena of 50 cm x 50 cm and a 60 cm height wall. The animals were first trained to forage for food in the light and in the dark. We found that PCs in the subiculum do indeed show a large heterogeneity regarding their stability under different light conditions, as illustrated in Figure 4. In it we show two different types of subiculum PCs whose FFs remapped and remained (A) and remapped and returned to the original location (B) when studied in a light-dark-light protocol. Those neurons that remap under different light conditions would integrate visual information in their spatial coding. We also observed a third population with FFs that did not change with the light.

It is clear that multiple factors are being coded by the hippocampus and the parahippocampal region. There is a clear influence of visual information on PCs firing, influence that is not enough to disrupt firing of PCs under multiple circumstances. The fact that some PCs can keep their FF in the darkness or after being blind is strong evidence that the animals are using other information to keep their representation. Also, 3D objects are able to produce an effect on PC firing if these are located distal from the centre of the arena. PCs adapt in different ways when the size of the arena is manipulated but this adaptation seems to be different in different brain areas. It has been well described that the hippocampus would code more than spatial information while other areas of the region would be less sensitive to these aspects. Therefore it is of great interest to investigate how different areas of the parahippocampal region and the hippocampus code different aspects of the “where” experience as well other elements of the context. Recent research has probed that rats can learn to navigate in a VR environment [72] and this opens a new door to use VR as a valuable tool in the quest for the understanding of spatial processing.

4. Place cells and Presence research.

Presence research and research on spatial processing are strongly interrelated. On one side, the understanding of the factors that most influence our sense of location in space and that induce the creation of internal cognitive maps of the space can be exploited to induce presence. Reciprocally, the use of virtual environments is one of the fundamental tools to comprehend spatial processing.

We have reviewed in this article the neuroscience literature devoted to spatial coding, concentrating mostly on hippocampal and parahippocampal place cells which comprise the best defined neuronal populations that participate in an internal representation of the external world.

What can we learn from how spatial information is processed in the brain that can be useful in the field of presence research? We follow the operational definition of presence that it is successful substitution of real by virtual sensory data, where success is indicated by participants acting and responding to virtual sensory data in a VE as if it were real world sensory data, and where response is multi-level [73].

From that point of view, and since PCs code for particular locations in the space, we propose that if the firing of PCs during virtual navigation corresponds to the firing of these PCs in the equivalent real space, this would provide one component of a measure of presence based on brain activity. It has been shown that indeed PCs in humans respond to particular locations within VEs [8]. However, a systematic use of this tool to measure presence is so far unattainable since it is only rarely, in pre-surgical brain patients with deep implanted electrodes, that such kind of single unit recordings can be obtained in humans. Otherwise, it would be appealing to test if presence correlates with the appropriate firing of place cells in VEs under a variety of experimental conditions (differences in visual realism, frame rate, etc), or to measure to what extent the pattern of PCs activation was transferable from a real to a virtual representation of the same space and vice versa. Although the difficulties to carry out these experiments are obvious, in theory they could provide a tool to better understand brain processing of spatial information both in real and VE. This theoretical consideration will still be valid if we consider that other methods of measuring brain activity such as brain imaging (fMRI) have already been used to detect the activation of neural structures during virtual navigation [9]. The limitations in this case are determined by the spatial resolution of the techniques (no single PCs can be detected). Another limitation is that the subject must navigate while remains motionless, since fMRI cannot be performed so far in moving around subjects. Therefore, it does not provide the means to compare human brain activity under real and virtual navigation. However, with the fast transformation that brain recording techniques have experienced in the last decades, it is reasonable to think that all these limitations will only lessen over time.

So far, as we have presented in this review, most of the studies on the neural mechanisms underlying spatial navigation in real environments have been studied in animal models. Recently, the first really effective VE for rats has been described [72]. In it, a group of animals were trained to navigate to specific locations in order to obtain a series of rewards. A second group of rats were trained in the equivalent real environment without finding any sort of behavioral difference between them both [72]. We could take this result as an evidence of spatial presence in the VE. The obvious next step that has not been yet taken is to record from PCs in these animals in the equivalent real and VE and to try to correlate the stability of the PCs firing fields with the successful transfer of information between both experimental conditions. Accoding to our hypothesis and operational definition of presence, the similar firing of PCs in both environments would underlie a similar processing of the spatial information and would reveal presence in the VE. The fact that hippocampal cells are very sensitive to spatial contextual changes could be used to measure how different a VE is perceived in relationship to its correspondent real environment. It also provides the means to experiment on the impact that different streams of sensory information have on the brain processing of space, exploiting the possibility of disrupting sensory modalities in VEs that always appear together in real environments. Thus, in a Ves, visual, vestibular, somatosensory, auditory or propioceptive information could be dissociated, providing an excellent tool for the evaluation of their individual role on spatial processing.

It is a fact that has been described by different authors, that VEs are useful for acquiring spatial knowledge [74], although these findings are not exempt from controversy [75]: differences in the fidelity of the environments or the training methods can yield different behavioral results. We know that when cells “learn” to fire in order to code for a new space, this pattern of firing can be maintained for at least a month [35]. This variable transfer of spatial knowledge between virtual and real environments [75] could be due to the efficiency of the VE to generate a stable, cognitive map of space, that remains functional when the subject is moved to operate in the equivalent environment in the real world. The success of this transfer could therefore reflect the activation of the same network of PCs both in the virtual and the real environments. For this reason, the transfer success could be taken as a surrogate of the stability of the map coded in the PCs and, furthermore, as a measure of spatial presence during virtual navigation.

At the same time that spatial mapping in place cells can be very stable, PCs are plastic and one observation that reveals this plasticity is the fact that areas of the space that are relevant from a behavioral point of view, have been reported to have larger representation in the hippocampal map [50]. This means that if a particular area of the space goes on to increase its relevance for the subject, the number of neurons that code for that particular area of space increases. Based on this observation it seems reasonable to predict that those VEs with higher behavioral significance for the subject are going to induce higher spatial presence. Or, what is the same, that a relatively crude VE could induce high spatial presence if what is represented is behaviorally relevant for the subject.

Conclusions

Place cells in the hippocampus and parahippocampal formation create an internal cognitive map of the external space that integrates information about location, multisensory inputs and internal information (propioceptive, vestibular, etc). Chronic recordings of PCs in animal experiments and eventually in humans have yielded valuable information about the functional properties of these neurons that we have reviewed in this study. We believe that this information is relevant for presence research since these neurons constitute the roots of spatial presence, without understimating the involvement of other areas of the brain (parietal, frontal cortex) in the process.

In this paper we suggest that if place cells activation operates in the same way in a VE as it does in its equivalent physical environment then this is one level of evidence that presence is occurring within that VE. We propose that this similar activation of PCs in virtual and real spaces should have its behavioral correlation in a succesful transfer of spatial information across both environments.

Acknowledgements

This work has been sponsored by PRESENCIA (European Union FET IST-2001-37927) and MCYT (BFI2002-03643) to MVSV.

References

[1] O'Keefe, J. & Dostrovsky, J. The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Res* 34, 171-5. 1971

[2] Bingman, V. P. & Yates, G. Hippocampal lesions impair navigational learning in experienced homing pigeons. *Behav Neurosci* 106, 229-32. 1992

[3] Watson, J. & Lashley, K. Homing and related activities of birds. *Carnegie Institute Publication* 211, 1-104. 1915

[4] Griffin, D. Bird navigation. in *Recent studies oin avian biology*, ed. ed., A. W. (University of Illinois Press, Urbana, Illinois, pp. 154-197. 1955

[5] Keeton, W. T. The mystery of pigeon homing. *Sci Am* 231, 96-7. 1974

[6] Bingman, V. P. & Jones, T. J. Sun compass-based spatial learning impaired in homing pigeons with hippocampal lesions. *J Neurosci* 14, 6687-94. 1994

[7] Lipp, H. P., Vyssotski, A. L., Wolfer, D. P., Renaudineau, S., Savini, M., Troster, G. & Dell'Omo, G. Pigeon homing along highways and exits. *Curr Biol* 14, 1239-49. 2004

[8] Ekstrom, A. D., Kahana, M. J., Caplan, J. B., Fields, T. A., Isham, E. A., Newman, E. L. & Fried, I. Cellular networks underlying human spatial navigation. *Nature* 425, 184-8. 2003

[9] Hartley, T., Maguire, E. A., Spiers, H. J. & Burgess, N. The well-worn route and the path less traveled: distinct neural bases of route following and wayfinding in humans. *Neuron* 37, 877-88. 2003

[10] Morris, R. G., Garrud, P., Rawlins, J. N. & O'Keefe, J. Place navigation impaired in rats with hippocampal lesions. *Nature* 297, 681-3. 1982

[11] Schenk, F. & Morris, R. G. Dissociation between components of spatial memory in rats after recovery from the effects of retrohippocampal lesions. *Exp Brain Res* 58, 11-28. 1985

[12] Morris, R. G., Schenk, F., Tweedie, F. & Jarrard, L. E. Ibotenate Lesions of Hippocampus and/or Subiculum: Dissociating Components of Allocentric Spatial Learning. *Eur J Neurosci* 2, 1016-1028. 1990

[13] Aguirre, G. K., Detre, J. A., Alsop, D. C. & D'Esposito, M. The parahippocampus subserves topographical learning in man. *Cereb Cortex* 6, 823-9. 1996

[14] Jordan, K., Schadow, J., Wuestenberg, T., Heinze, H. J. & Jancke, L. Different cortical activations for subjects using allocentric or egocentric strategies in a virtual navigation task. *Neuroreport* 15, 135-40. 2004

[15] Whishaw, I. Q., Hines, D. J. & Wallace, D. G. Dead reckoning (path integration) requires the hippocampal formation: evidence from spontaneous exploration and spatial learning tasks in light (allothetic) and dark (idiothetic) tests. *Behav Brain Res* 127, 49-69. 2001

[16] Save, E., Paz-Villagran, V., Alexinsky, T. & Poucet, B. Functional interaction between the associative parietal

cortex and hippocampal place cell firing in the rat. *Eur J Neurosci* 21, 522-30. 2005

[17] Commins, S., Gemmell, C., Anderson, M., Gigg, J. & O'Mara, S. M. Disorientation combined with bilateral parietal cortex lesions causes path integration deficits in the water maze. *Behav Brain Res* 104, 197-200. 1999

[18] Wolbers, T., Weiller, C. & Buchel, C. Neural foundations of emerging route knowledge in complex spatial environments. *Brain Res Cogn Brain Res* 21, 401-11. 2004

[19] Witter, M. P., Groenewegen, H. J., Lopes da Silva, F. H. & Lohman, A. H. Functional organization of the extrinsic and intrinsic circuitry of the parahippocampal region. *Prog Neurobiol* 33, 161-253. 1989

[20] Bland, B. H. & Oddie, S. D. Theta band oscillation and synchrony in the hippocampal formation and associated structures: the case for its role in sensorimotor integration. *Behav Brain Res* 127, 119-36. 2001

[21] Whishaw, I. Q. & Vanderwolf, C. H. Hippocampal EEG and behavior: changes in amplitude and frequency of RSA (theta rhythm) associated with spontaneous and learned movement patterns in rats and cats. *Behav Biol* 8, 461-84. 1973

[22] Kramis, R., Vanderwolf, C. H. & Bland, B. H. Two types of hippocampal rhythmical slow activity in both the rabbit and the rat: relations to behavior and effects of atropine, diethyl ether, urethane, and pentobarbital. *Exp Neurol* 49, 58-85. 1975

[23] O'Keefe, J. & Recce, M. L. Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus* 3, 317-30. 1993

[24] Bischof, W. F. & Boulanger, P. Spatial navigation in virtual reality environments: an EEG analysis. *Cyberpsychol Behav* 6, 487-95. 2003

[25] Kahana, M. J., Sekuler, R., Caplan, J. B., Kirschen, M. & Madsen, J. R. Human theta oscillations exhibit task dependence during virtual maze navigation. *Nature* 399, 781-4. 1999

[26] de Araujo, D. B., Baffa, O. & Wakai, R. T. Theta oscillations and human navigation: a magnetoencephalography study. *J Cogn Neurosci* 14, 70-8. 2002

[27] Kelemen, E., Moron, I. & Fenton, A. A. Is the hippocampal theta rhythm related to cognition in a non-locomotor place recognition task? *Hippocampus* 15, 472-9. 2005

[28] Poucet, B., Lenck-Santini, P. P., Paz-Villagran, V. & Save, E. Place cells, neocortex and spatial navigation: a short review. *J Physiol Paris* 97, 537-46. 2003

[29] O'Mara, S. M. Spatially selective firing properties of hippocampal formation neurons in rodents and primates. *Prog Neurobiol* 45, 253-74. 1995

[30] O'Mara, S. M., Rolls, E. T., Berthoz, A. & Kesner, R. P. Neurons responding to whole-body motion in the primate hippocampus. *J Neurosci* 14, 6511-23. 1994

[31] Ludvig, N., Tang, H. M., Gohil, B. C. & Botero, J. M. Detecting location-specific neuronal firing rate increases in the hippocampus of freely-moving monkeys. *Brain Res* 1014, 97-109. 2004

[32] Roche, R., Mangaoang, M., Commins, S. & O'Mara, S. Hippocampal Contributions To Neurocognitive Mapping In Humans: A New Model. *Hippocampus* in press. 2005

[33] Thompson, L. T. & Best, P. J. Long-term stability of the place-field activity of single units recorded from the dorsal hippocampus of freely behaving rats. *Brain Res* 509, 299-308. 1990

[34] Hill, A. J. First occurrence of hippocampal spatial firing in a new environment. *Exp Neurol* 62, 282-97. 1978

[35] Lever, C., Wills, T., Cacucci, F., Burgess, N. & O'Keefe, J. Long-term plasticity in hippocampal place-cell representation of environmental geometry. *Nature* 416, 90-4. 2002

[36] Kentros, C., Hargreaves, E., Hawkins, R. D., Kandel, E. R., Shapiro, M. & Muller, R. V. Abolition of long-term stability of new hippocampal place cell maps by NMDA receptor blockade. *Science* 280, 2121-6. 1998

[37] McNaughton, B. L., Barnes, C. A., Meltzer, J. & Sutherland, R. J. Hippocampal granule cells are necessary for normal spatial learning but not for spatially-selective pyramidal cell discharge. *Exp Brain Res* 76, 485-96. 1989

[38] McNaughton, B. L., Barnes, C. A., Gerrard, J. L., Gothard, K., Jung, M. W., Knierim, J. J., Kudrimoti, H., Qin, Y., Skaggs, W. E., Suster, M. & Weaver, K. L. Deciphering the hippocampal polyglot: the hippocampus as a path integration system. *J Exp Biol* 199 (Pt 1), 173-85. 1996

[39] McNaughton, B. L., Barnes, C. A. & O'Keefe, J. The contributions of position, direction, and velocity to single unit activity in the hippocampus of freely-moving rats. *Exp Brain Res* 52, 41-9. 1983

[40] Markus, E. J., Qin, Y. L., Leonard, B., Skaggs, W. E., McNaughton, B. L. & Barnes, C. A. Interactions between location and task affect the spatial and directional firing of hippocampal neurons. *J Neurosci* 15, 7079-94. 1995

[41] Muller, R. U., Bostock, E., Taube, J. S. & Kubie, J. L. On the directional firing properties of hippocampal place cells. *J Neurosci* 14, 7235-51. 1994

[42] Taube, J. S., Muller, R. U. & Ranck, J. B., Jr. Head-direction cells recorded from the postsubiculum in freely moving rats. II. Effects of environmental manipulations. *J Neurosci* 10, 436-47. 1990

[43] Taube, J. S. Head direction cells recorded in the anterior thalamic nuclei of freely moving rats. *J Neurosci* 15, 70-86. 1995

[44] Goodridge, J. P., Dudchenko, P. A., Worboys, K. A., Golob, E. J. & Taube, J. S. Cue control and head direction cells. *Behav Neurosci* 112, 749-61. 1998

[45] Cacucci, F., Lever, C., Wills, T. J., Burgess, N. & O'Keefe, J. Theta-modulated place-by-direction cells in the hippocampal formation in the rat. *J Neurosci* 24, 8265-77. 2004

[46] Breese, C. R., Hampson, R. E. & Deadwyler, S. A. Hippocampal place cells: stereotypy and plasticity. *J Neurosci* 9, 1097-111. 1989

[47] Speakman, A. & O'Keefe, J. Hippocampal Complex Spike Cells do not Change Their Place Fields if the Goal is Moved Within a Cue Controlled Environment. *Eur J Neurosci* 2, 544-555. 1990

[48] Gemmell, C. & O'Mara, S. M. Medial prefrontal cortex lesions cause deficits in a variable-goal location task but not in object exploration. *Behav Neurosci* 113, 465-74. 1999

[49] Hollup, S. A., Molden, S., Donnett, J. G., Moser, M. B. & Moser, E. I. Accumulation of hippocampal place fields at the goal location in an annular watermaze task. *J Neurosci* 21, 1635-44. 2001

[50] Martin, P. D. & Ono, T. Effects of reward anticipation, reward presentation, and spatial parameters on the firing of single neurons recorded in the subiculum and nucleus accumbens of freely moving rats. *Behav Brain Res* 116, 23-38. 2000

[51] Kobayashi, T., Nishijo, H., Fukuda, M., Bures, J. & Ono, T. Task-dependent representations in rat hippocampal place neurons. *J Neurophysiol* 78, 597-613. 1997

[52] Hok, V., Save, E., Lenck-Santini, P. P. & Poucet, B. Coding for spatial goals in the prelimbic/infralimbic area of the rat frontal cortex. *Proc Natl Acad Sci U S A* 102, 4602-7. 2005

[53] Pratt, W. E. & Mizumori, S. J. Neurons in rat medial prefrontal cortex show anticipatory rate changes to predictable differential rewards in a spatial memory task. *Behav Brain Res* 123, 165-83. 2001

[54] White, N. M. & McDonald, R. J. Acquisition of a spatial conditioned place preference is impaired by amygdala lesions and improved by fornix lesions. *Behav Brain Res* 55, 269-81. 1993

[55] O'Mara, S. M., Commins, S., Anderson, M. & Gigg, J. The subiculum: a review of form, physiology and function. *Prog Neurobiol* 64, 129-55. 2001

[56] Muller, R. U. & Kubie, J. L. The effects of changes in the environment on the spatial firing of hippocampal complex-spike cells. *J Neurosci* 7, 1951-68. 1987

[57] Sharp, P. E. Subicular place cells expand or contract their spatial firing pattern to fit the size of the environment in an open field but not in the presence of barriers: comparison with hippocampal place cells. *Behav Neurosci* 113, 643-62. 1999

[58] Sharp, P. E. Subicular cells generate similar spatial firing patterns in two geometrically and visually distinctive environments: comparison with hippocampal place cells. *Behav Brain Res* 85, 71-92. 1997

[59] Hetherington, P. A. & Shapiro, M. L. Hippocampal place fields are altered by the removal of single visual cues in a distance-dependent manner. *Behav Neurosci* 111, 20-34. 1997

[60] Jeffery, K. J. & O'Keefe, J. M. Learned interaction of visual and idiothetic cues in the control of place field orientation. *Exp Brain Res* 127, 151-61. 1999

[61] Cressant, A., Muller, R. U. & Poucet, B. Failure of centrally placed objects to control the firing fields of hippocampal place cells. *J Neurosci* 17, 2531-42. 1997

[62] Shapiro, M. L., Tanila, H. & Eichenbaum, H. Cues that hippocampal place cells encode: dynamic and hierarchical representation of local and distal stimuli. *Hippocampus* 7, 624-42. 1997

[63] Hill, A. J. & Best, P. J. Effects of deafness and blindness on the spatial correlates of hippocampal unit activity in the rat. *Exp Neurol* 74, 204-17. 1981

[64] Quirk, G. J., Muller, R. U. & Kubie, J. L. The firing of hippocampal place cells in the dark depends on the rat's recent experience. *J Neurosci* 10, 2008-17. 1990

[65] Save, E., Cressant, A., Thinus-Blanc, C. & Poucet, B. Spatial firing of hippocampal place cells in blind rats. *J Neurosci* 18, 1818-26. 1998

[66] Paz-Villagran, V., Lenck-Santini, P. P., Save, E. & Poucet, B. Properties of place cell firing after damage to the visual cortex. *Eur J Neurosci* 16, 771-6. 2002

[67] O'Keefe, J. & Burgess, N. Geometric determinants of the place fields of hippocampal neurons. *Nature* 381, 425-8. 1996

[68] Rivard, B., Li, Y., Lenck-Santini, P. P., Poucet, B. & Muller, R. U. Representation of objects in space by two classes of hippocampal pyramidal cells. *J Gen Physiol* 124, 9-25. 2004

[69] Anderson, M. I. & Jeffery, K. J. Heterogeneous modulation of place cell firing by changes in context. *J Neurosci* 23, 8827-35. 2003

[70] Sharp, P. E. Complimentary roles for hippocampal versus subicular/entorhinal place cells in coding place, context, and events. *Hippocampus* 9, 432-43. 1999

[71] Deadwyler, S. A. & Hampson, R. E. Differential but complementary mnemonic functions of the hippocampus and subiculum. *Neuron* 42, 465-76. 2004

[72] Holscher, C., Schnee, A., Dahmen, H., Setia, L. & Mallot, H. A. Rats are able to navigate in virtual environments. *J Exp Biol* 208, 561-9. 2005

[73] Sanchez-Vives, M. V. & Slater, M. From presence to consciousness through virtual reality. *Nat Rev Neurosci* 6, 332-9. 2005

[74] Bliss, J., Tidwell, P. & Guest, M. The effectiviness of virtual reality for adminestering spatial navigation training to firefighters. *Presence-Teleoperators and Virtual Environments* 6, 73-86. 1997

[75] Darken, R. & Banker, W. in *Proceedings of the IEEE Virtual Reality Annual International Symposium*. ed. Press, C. S., Los Alamitos, CA). 1998

Session 3

21st September, 2005

Interaction with Avatars

17.30-18.00 *The Effect of Behavioral Realism and Form Realism of Real-Time Avatar Faces on Verbal Disclosure, Nonverbal Disclosure, Emotion Recognition, and Copresence in Dyadic Interaction*
Jeremy N. Bailenson[1], Nick Yee[1], Dan Merget[2] and Ralph Schroeder[3]
[1] Department of Communication, Stanford University, USA
[2] Department of Computer Science, Stanford University, USA
[3] Oxford Internet Institute, University of Oxford, UK

18.00-18.30 *BASIC: A Believable, Adaptable, Socially Intelligent Character for Social Presence*
Daniela Romano
Department of Computer Science, University of Sheffield, UK

18.30-18.45 *Virtual encounters. Creating social presence in net-based collaborations*
Sabine Rüggenberg, Gary Bente and Nicole C. Krämer
Department of Psychology, University of Cologne, Germany

18.45-19.00 *Non-verbal Communication for Correlational Characters*
Marco Gillies and Mel Slater
Department of Computer Science, University College London, UK

The Effect of Behavioral Realism and Form Realism of Real-Time Avatar Faces on Verbal Disclosure, Nonverbal Disclosure, Emotion Recognition, and Copresence in Dyadic Interaction

Jeremy N. Bailenson[1], Nick Yee[1], Dan Merget[2] and Ralph Schroeder[3]
[1] Department of Communication, Stanford University, USA
[2] Department of Computer Science, Stanford University, USA
[3] Oxford Internet Institute, University of Oxford, UK

Abstract

The realism of avatars in terms of behavioral and form is critical to the development of collaborative virtual environments. In the study we utilized state of the art, real-time face tracking technology to track and render facial expressions unobtrusively in a desktop CVE. Participants in dyads interacted with each other via either a videoconference (high behavioral realism and high form realism), voice only (low behavioral realism and low form realism), or an "emotibox" that rendered the dimensions of facial expressions abstractly in terms of color, shape, and orientation on a rectangular polygon (high behavior realism and low form realism). Verbal and non-verbal self-disclosure were lowest in the videoconference condition while self-reported copresence and success of transmission and identification of emotions were lowest in the emotibox condition. Previous work demonstrates that avatar realism increases copresence while decreasing self-disclosure. We discuss the possibility of a hybrid realism solution that maintains high copresence without lowering self-disclosure, and the benefits of such an avatar on applications such as distance learning and therapy.

1. Avatars

1.1. What is an avatar?

The study of virtual humans—from conceptual, design, and empirical perspectives—has progressed greatly over the past fifteen years. Traditionally, the field of research has delineated between *embodied agents* which are digital models driven by computer algorithms and *avatars* which are digital models driven by real-time humans. In terms of empirical behavioral research examining how people interact with virtual humans in social interaction, a majority of this work has utilized embodied agents (as opposed to avatars—see Bailenson & Blascovich [3] for a discussion of this disparity). One reason for this bias is because it is only over the past few years that readily available commercial technology has actually allowed people to make avatars that can look like and behave - via real-time tracking - like the user. In other words, up until now, producing real-time avatars that captured the user's visual features and subtle movements has been quite difficult to accomplish in a social science laboratory. Consequently, understanding the implications of the visual and behavioral veridicality of an avatar on the quality of interaction and on copresence is an important question that has received very little empirical attention. Schroeder [22] provides a review of the existing empirical work on avatars.

Avatars can be defined as digital models of people that either look or behave like the users they represent. In traditional immersive virtual environments, an avatar is the model that is rendered on the fly to reflect the user's behavior. However, the definition of an avatar certainly has blurry boundaries. For example, the definition including "looking like a user" would allow for a digital photograph of a person stored on a hard drive to be considered an avatar. Some would object that this archived image is not an avatar since it has no potential for behavior or for social interaction. On the other hand, some would include the photograph in the definition, arguing that people utilize static (i.e., non-animated) avatars with internet chat and emails. While people discuss the concept of avatars quite often in the literature on virtual humans and virtual environments, a standard definition of avatars has not emerged readily. But since avatars are playing an increasingly central role in virtual environments and other electronic media, it is important to investigate the suitability of different types of avatars for representing the user.

Figure 1 provides a preliminary attempt to provide a framework for considering representations of humans that is not limited just to digital avatars. The Y-axis denotes behavioral similarity—how much the behaviors of the representation correspond to the behaviors of a given person. The X axis indicates form similarity, how much the representation statically resembles features of a given person. On the left side are representations that correspond to a given person's form or behavior in real-time. On the right are representations that correspond to a person's form or behavior asynchronously. For example, a puppet is a representation of a person that has high behavioral similarity (the movements of the puppet are very closely tied to the person controlling it) but low form similarity (the puppet need not look at all like the person controlling it). Furthermore, the puppet's behaviors are expressed in real-time. On the other hand, an impressionist (i.e., someone who can very closely reproduce or mimic the behaviors of a person who is not physically present) has high behavioral similarity and low form similarity in that the impressionist need not look like the person being mimicked. Unlike the puppet, however, the impressionist is a non-real-time representation—the person being mimicked need not be present, aware of the impressionist's existence, or even still alive for that matter.

As Figure 1 demonstrates, there are lots of different types of representations of people utilized today. The shaded oval denotes the space in which we typically discuss avatars—digital representations of humans that are utilized in immersive virtual environments. Blascovich and colleagues [7] provide a theoretical framework to determine the interplay of behavioral and form realism for the avatars which fall into this shaded region.

1.2. Avatars and Copresence

A key reason why avatar form and behavior are so important is that they elicit an experience of being with another person; or copresence (also referred to as social presence). There are many definitions of copresence in the literature. Heeter defined copresence as the extent to which other beings, both living and synthetic, exist in a virtual world and appear to react to human interactants [15] . Slater and colleagues, in contrast, define copresence as the sense of being and acting with others in a virtual place [24] . Lee defines copresence as experiencing artificial social actors (agents) via objects that manifest humanness or actual social actors (avatars) connected via technology [18] . Finally, Blascovich and his colleagues have defined copresence as the extent to which individuals treat embodied agents as if they were other real human beings [7] .

Biocca, Harms and Burgoon [8] review the various definitions and measures of copresence. They discuss different media, including those in which the 'other' with whom one experiences presence can be an agent or other media-generated human-like appearance, and they include a broad range of phenomena within copresence partly so that they can compare different media (for example, para-social interaction with a film character). They also review several measures that have been proposed for copresence, including self-report, behavioural and psycho-physiological measures, but point out that little consensus has been reached on this issue. Their proposal to specify an extensive set of criteria and scope conditions for copresence is quite broad, including items such as "read[ing] minds" in both people and things' ([8] : 474). However, they also describe copresence as a more tightly defined subset of a larger phenomenon whereby people need to have a sensory experience of sharing the same space with someone else. This limits copresence to face-to-face experiences or experiences in which two (human) users *both* share the space and the sensory experience of each other (this also corresponds to Schroeder's strict definition of copresence [23]).

It is clear that different measures of copresence have drawbacks: self-report measures are subjective, but any objective (behavioural, cognitive, or psycho-physiological) measures will also be problematic since they will not directly reveal what people feel or how they interpret the

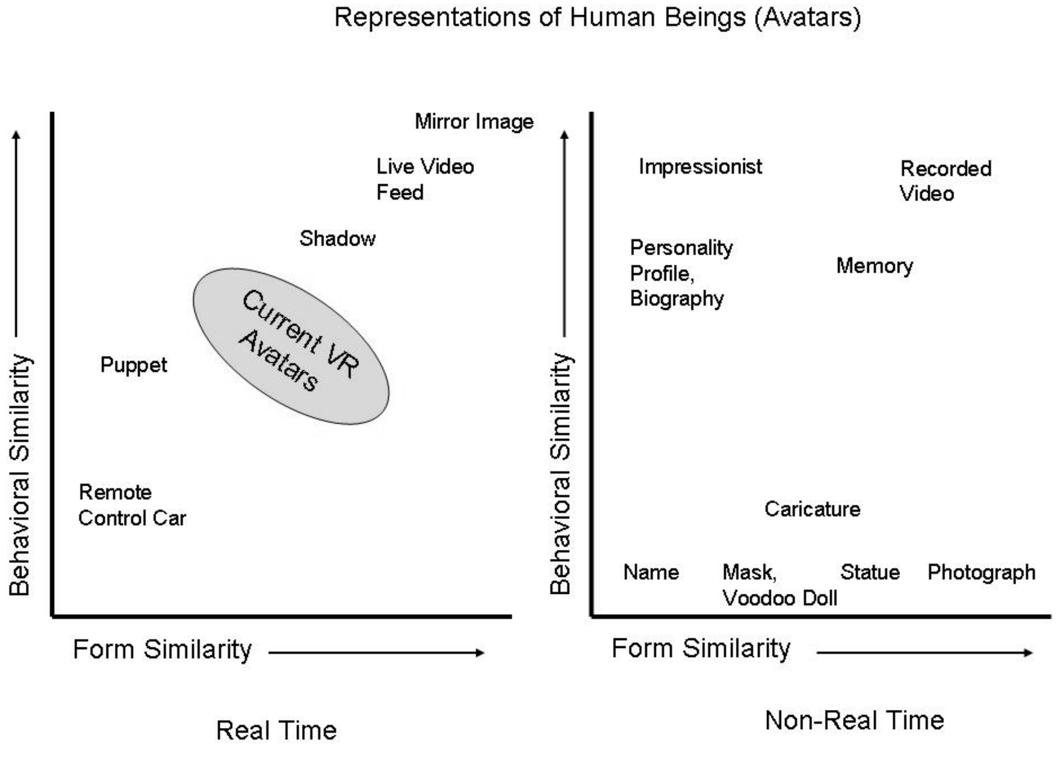


Figure 1: A framework for classifying representations of humans in physical and digital space.

presence of another. Obviously a combination of methods will provide the most well-rounded understanding and/or explanation of this phenomenon. Indeed, a recent empirical study by Bailenson, Swinth, Hoyt, Persky, Dimov, & Blascovich [4] directly compared the subjective, behavioural, and cognitive measures of copresence. Their results confirmed the hypotheses of Biocca et al.—by providing affective, behavioural and cognitive measures of copresence, they demonstrated that subjective reports alone were insufficient to highlight copresence differences between various types of agents. On the other hand linking the specific behavioural and cognitive results (which discovered differences not detected by self report measures) to the latent construct of copresence proved challenging.

A number of studies have explored avatar realism experimentally. Bailenson, Beall, & Blascovich [1] demonstrated that higher behavioural realism in terms of mutual gaze produced higher levels of copresence and produced changes in task performance. Garau [14] investigated photorealistic and behaviourally realistic avatars and showed that behavioural realism is more important than form realism in several different scenarios. Moreover, Bente [6] has shown that even avatars with very minimal levels of behavioural realism elicit responses from others.

There are also studies that have examined the interaction between avatars in ‘naturalistic’ settings. Becker and Mark [5] compared how social conventions are followed in three different online VEs: text-only, text-and-graphics, and voice plus talking head. They found, based on participant observation, that certain conventions from face-to-face interaction are followed in all three settings, but that certain of them are followed more in the more ‘realistic’ setting (i.e., interpersonal distance is kept more in the shared VE with voice). It has also been investigated what preferences people have for different avatar appearances. Cheng, Farnham and Stone [10] found that people in a text-and-graphics shared VE (V-chat) preferred representations of themselves that were neither too true-to-life to their own appearance nor too abstract. These studies demonstrate that people’s habits and preferences will shape avatar appearance.

A related topic is the extent to which avatars are developed sufficiently enough to allow the transmission of ‘social cues’ of face-to-face communication, which includes all the information about one another (pitch of the voice, non-verbal gestures, etc.—see Whittaker [29] for a review). Walther [27] has argued against the widely held view that interaction with avatars lacks ‘social richness’ or ‘media richness’. He has shown that it is not necessarily the case that less rich media prevent people from getting to know each other; it may just take more time. In fact, he argues that they may get to know each other better in certain respects in less rich media; he calls these ‘hyperpersonal’ relationships that are created among avatars and other representations in computer mediated communication in which people form extremely deep social ties with each other.

The literature on self-disclosure suggests that copresence mediates the effect of visual and behavioural realism on self-disclosure. For example, a meta-analysis of studies on self-disclosure in face-to-face interviews as compared with computer-administered interviews found that self-disclosure was higher in computer-administered interviews than in face-to-face interactions [28] . This suggests that less realistic avatars would elicit more self-disclosure from users. In a study where participants interacted with either a text-based or face-based agent, it was found that participants revealed more information about themselves when interacting with the text-based agent [25] . Previous researchers have also implemented and discussed self disclosure as a proxy for measures of copresence [20] .

1.3. Facial Expressions and Facial Tracking of Avatars

Research on transmitting as well as receiving facial expressions has received much attention from social scientists for the past fifty years. Some researchers argue that the face is a portal to the one’s internal mental state (Ekman & Friesen [12] , Izard [16]). These scholars argue that when an emotion occurs, a series of biological events follow that produce changes in a person—one of those manifestations is movement in facial muscles. Moreover, these changes in facial expressions are also correlated with other physiological changes such as heart rate changes or heightened blood pressure [11] .

The use of facial expressions to form attributions concerning others certainly changes during mediated communication. Telephone conversations clearly function quite well without any visual cues about another’s face. As Whittaker [29] points out in a review of the literature examining visual cues in mediated communication, adding visual features is not always beneficial, and can sometimes be counterproductive. Specifically, Whittaker’s survey of findings demonstrates that showing another person’s face during interaction tends to be more effective when the goal of the interaction is social than when it is purely task oriented. However, a large part of the problems with previously studied visual mediated communication systems have been due to bandwidth delay in videoconferences or from the stark conditions offered by other virtual solutions [17] . However, as virtual reality systems and other visually mediated communications systems improve the accuracy of visual representations will become closer to that seen in face-to-face interaction. Consequently, facial expressions seen during human-computer interaction will be more diagnostic of actual facial movements.

There has recently been a great surge of work to develop automatic algorithms to identify emotional states from a video image of facial movements. Early work developed a system of facial action coding system in which coders manually identified anchor points on the face in static images [12] . Similarly, computer scientists have developed vision algorithms that automatically find similar anchor points with varying amounts of success (see Essa & Pentland [13] for an early example). As computer vision algorithms and perceptual interfaces become more elegant (see Turk & Kölsch [26] for a review), it is becoming possible to measure the emotional state of people in real-time, based on algorithms that automatically detect facial anchor points without using markers on the face and then and categorize those points into emotions that have been previously identified using some type of learning algorithm. These systems sometimes attempt to recognize specific emotions [19] or alternatively attempt to gauge binary states such as general affect [21] .

2. Study Overview

In the current study we empirically test two of the dimensions of avatars depicted in Figure 1—behavioural and form realism. We varied the extent to which an avatar's face resembled and gestured similarly to the users' faces. Dyads interacted via a desktop virtual display, and we tracked in real-time 22 anchor points on their faces as well as position of the faces and orientation of the faces. We are interested in how people behaved towards one another's avatars and whether or not they revealed more about themselves (in terms of how much information they revealed verbally as well as how much information they revealed through facial gestures) when they encountered avatars that were less realistic in form and behaviour. Furthermore, we measured the ability of subjects to transmit and receive specific emotional expressions at various levels of behavioural and form realism as both a cognitive measure of copresence as well as a test of our face-tracking system's effectiveness.

We had three conditions: 1) voice only, 2) videoconference, and 3) the emotibox—a polygon that changed shape, colour and orientation in response to the user's head position and facial expressions. Figure 2 shows screenshots of these three conditions.

The emotibox is reminiscent of the 'blockie' avatars of the avatars that were used in some of the earliest research on CVEs [9] . Here, we implement this type of avatar because it is a manner to represent high behavioural realism (via facial emotion) with low form realism. By high behavioural realism, we simply mean that the avatar behaves in ways that are contingent upon the behaviours of a human. In other words, our definition of behavioural realism in the current study requires a) a high number of behaviours to be tracked, and b) a high number of behaviours rendered on the avatar that are contingent upon those tracked behaviours. In some ways, this definition is counterintuitive, because the behaviours do not look like the actual behaviours of the user since they are abstracted. The hypothesis in the current study was that demonstrating behavioural contingency (though not behavioural similarity) was the best compromise between high behavioural realism and low form realism. Because it is not possible to have facial movements reflected realistically on an avatar without facial features, the emotibox maintained the best balance between high behavioural realism and low form realism.

If one of the main difficulties of shared VEs and other computer-mediated communication is going to be the live capture of people's facial appearance and expressions, then the amount of realism required for non-verbal facial communication becomes an important question. To our knowledge this is one of the first empirical studies of copresence that utilizes avatars capable of rendering real-time emotional expressions via face-tracking. By examining the unique contribution of facial expressions as an independent variable, as well as using the amount of emotions conveyed as a dependent variable, we can potentially examine a unique level of avatar realism.

3. Method

3.1. Participants

Thirty undergraduate students (12 men and 18 women) were paid ten dollars each for their participation in the study. The gender makeup of dyads was 3 male-male pairs, 6 mixed pairs, and 6 female pairs.

3.2. Design

There were three conditions in the study: 1) voice only, 2) videoconference, and 3) emotibox. In all 3 conditions, participants were seated in front of a computer terminal equipped with a Logitech QuickCam Messenger digital camera mounted on top of the monitor. A conferencing application (Microsoft Netmeeting) was used in all three

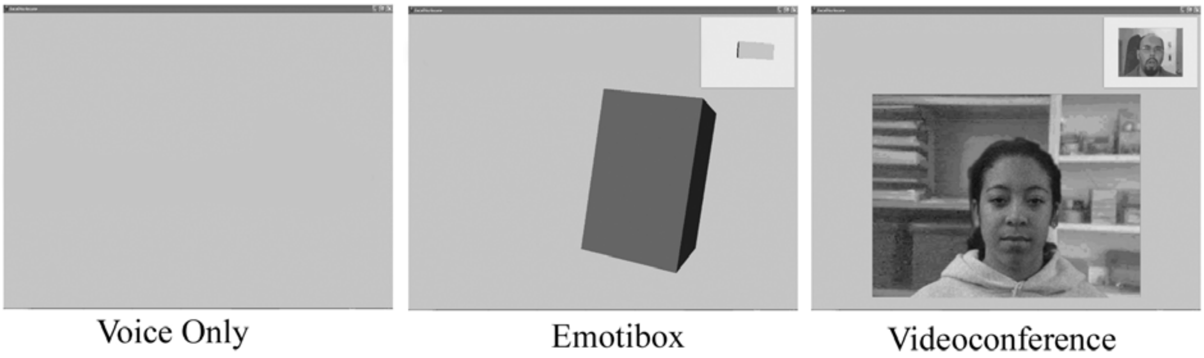


Figure 2: A subject's eye-view of the three conditions. In the right two panels, the center of the screen shows the avatar of the other interactant while the top right corner shows the subject's own avatar

conditions for voice.

In the *videoconference* condition, the conferencing application allowed participants to see each other via the digital cameras. The video feed was a gray-scale image with 256 levels, updated at 20 frames per second. The video was acquired at a resolution of 320x240, and then magnified to 760x570 so that it would fill most of a 1024x768 screen. While a videoconference image may not be traditionally categorized as an avatar, given that we were using digital video it does fit the definition discussed earlier on in this work. More importantly, for our purposes in this experiment, a videoconference worked most effectively as a high realism control condition.

In the *emotibox condition*, the Nevenvision Facial Feature Tracker, a real-time face-tracking solution, was integrated into Vizard 2.5, a platform for developing virtual environments, to capture key locations of the face. These anchor points, depicted in Figure 3, included 8 points around the contour of the mouth (three on each lip, and one at each corner), three points on each eye (including the pupil), two points on each eyebrow, and four points around the nose. The points were measured in a two-dimensional head-centred coordinate system normalized to the apparent size of the head on the screen; the coordinates were not affected by rigid head movements, and scaled well to different heads. The face-tracking software also tracked the pitch, yaw and roll of the face, the aspect ratio of the mouth and each eye, the coordinates of the face in the webcam image, and the scale of the face (which is inversely proportional to the distance from the face to the webcam). Our real-time face-tracking solution required no training, face-markers, or calibration for individual faces.

Figure 3: The 22 anchor points automatically tracked without using facial markers by the Nevenvision Facial Feature Tracker at 30 Hz.

The emotibox was based on the YUV colour scheme and had 11 degrees of freedom: 1) the eye aspect ratio controlled the Y-value (i.e., black-white spectrum) of the cube. In laboratory pilot studies, the aspect ratio of one eye was found to vary roughly between 0.10-0.35, so the aspect ratio of each eye was added together, truncated to the range 0.20-0.70, and linearly transformed to a Y-value from 0.5-1.0. Thus, the wider the person's eyes, the brighter the cube. A minimum Y of 0.5 kept the cube bright enough so the U and V could be seen. 2) The distance between the corners of the mouth and the eyes controlled the U-value (i.e., the blue-yellow spectrum) of the cube. The total distance was truncated to the range 85-100, and linearly transformed to a U of -0.4 to +0.4. Thus, the more a person smiled, the more yellow the cube became. And the more a person frowned, the more blue the cube became. These colours were chosen after extensive pre-testing indicated the most separability in terms of mapping discrete mental states. 3) The distance of the eyebrows from the pupils controlled the V-value (i.e., red-cyan spectrum) of the cube. Two different scales were used, since we found that relaxing the eyebrows brought them very close to their lowest extreme (at least according to our tracking software). Distances from 27-35 mapped to a V-value of 0.0 to +0.6, but distances from only 27 to 25 mapped to v-values of 0 to -0.6. The more you raised your eyebrows, the more cyan the cube would become. The 4) width and 5) height of the emotibox followed the width and height of the mouth: each dimension varied from 50% to 150% of the basic cube as the mouth width and height varied from 15-35 and 28-42, respectively. Finally, the emotibox followed the 6) pitch, 7) yaw, 8) roll, 9) x-coordinate, 10) y-coordinate, and 11) z-coordinate of the head.

The emotibox was also updated 20 times per second, even though the face-tracking software acquired images at 30 Hz. When the confidence of the face-tracking software fell below 40%, the data was discarded and the software was told to re-acquire the face from scratch. The other subject saw a frozen emotibox until this process was done. In the voice only condition, the sound system allowed participants to hear each other's voice.

In the *voice only* condition, subjects saw a blank screen and communicated through the audio software.

3.3. Materials

To generate two sets of questions (one for each interactant in the dyad) of a comparable degree of intimacy for the verbal self-disclosure task, 30 questions were pretested for their degree of intimacy. To pretest the materials, 15 undergraduates from a separate population from the experimental pool rated each of the questions on a 5-point, fully-labeled, construct-specific scale, ranging from "Not Personal At All" to "Extremely Personal". Six pairs of questions were chosen such that the questions in each pair did not differ significantly from each other in a t-test. In addition, we added a general self-disclosure question at the end of both sets - "Tell me a little more about yourself". These two sets of questions used in the main experiment are listed in the Appendix.

3.4. Procedure

Pairs of participants arrived at the laboratory for each session. Most participants did see one another in vivo before the experiment began. After signing informed consent, they were seated in front of the computer terminals in different rooms. Each pair of participants was assigned to the same condition using a predetermined randomization scheme. The study began with the verbal self-disclosure task. For all three conditions, the question sets were displayed textually on the monitor one at a time and alternated between the two participants. Participants were instructed to ask the other participant the question that was displayed (via text on the monitor) by speaking into a headset microphone. The participant that answered the question advanced to the next question by pressing the space bar when he or she was finished speaking. We randomized which participant would ask the first question. The audio from all interactions was recorded.

The second task was an emoting task. Participants were given a list of seven emotions, one at a time in random order - disgusted, angry, sad, joyful, afraid, interested, and surprised. For each emotion, participants were asked to convey that emotion to the other participant for 10 seconds. The video-feed and emotibox subjects conveyed the emotion via facial expression, while the voice-only subjects used nonverbal sounds (i.e., no words allowed) to express themselves. While this condition is somewhat unnatural, this was the best way for us to not allow for the use of language or grammar to clue the specific emotion. After each emotion, the other participant would be asked which emotion was conveyed, and how sure they were of their answer. One participant would be instructed to emote through all seven emotions and then the other participant would be instructed to do the same. The last task was filling out the copresence questionnaire. Participants saw one question on the screen at a time in a random order and responded using the keyboard.

4. Measures and Hypotheses

4.1. Verbal Self-Disclosure

Two coders blind to experimental condition listened to the audio recordings of all interactants and rated each one’s friendliness, honesty and how revealing their responses were on 5-point, fully-labeled, construct-specific scales. Thus, each participant had six ratings, three from each coder. The composite scale composed of these six items had a reliability of .85. We hypothesized that self-disclosure would be lowest in the videoconference condition and highest in the voice only condition, and that there would be more disclosure in front of the emotibox than the videoconference.

4.2. Non-Verbal Self-Disclosure

Previous research discussed above has indicated that people disclose more verbal information in a text interface than in an avatar-based interface. We were interested in testing for this effect in terms of non-verbal behaviors. We therefore predicted that participants in the voice only condition would disclose more non-verbal information than in the videoconference and emotibox conditions. The face tracking software was used to find 22 points on the face that varied with expression (see Figure 2), but were not affected by the position and/or orientation of the head as a whole. The standard deviation of each point (both x and y coordinates) measured how much activity occurred at that point, and the average of all 44 standard deviations served as a measure of how expressive the face was during the experiment. This metric is deliberately naïve, and some points, such as the corners of the mouth, were up to 6 times as mobile as others, and thus contributed more heavily to the face movement metric. Nonetheless, we used the simplest, least biased way of combining the measurements into a single score.[1] In future work, we plan on developing more elegant combinations of the facial feature points.

4.3. Copresence Ratings

Participants completed a 4-item copresence scale depicted in the Appendix, which was modeled after the scale developed by Biocca, Harms, & Burgoon [8]. The reliability of the composite scale was .62. We hypothesized that copresence would be highest in the videoconference condition and lowest in the voice only condition.

4.4. Emotion Detection

Participants were scored a 1 if they guessed the emotion correctly, a 0 if they were incorrect. The composite scale composed of the mean of the seven detection scores had a reliability of .62.

5. Results

5.1. Verbal Self Disclosure

We ran a between-subjects ANOVA with condition (voice only, emotibox, and videoconference) and subject gender as independent factors and self disclosure score as a dependent variable. There was a significant effect of condition, $F(2,24) = 5.80$, $p<.001$, partial Eta Squared = .33. As Figure 4 demonstrates, there was more disclosure in the voice only and the emotibox conditions than the videoconference conditions. The effect of participant gender was not significant, $F(1,24) = .02$, $p<.90$, partial Eta Squared = .00, and the interaction was not significant, $F(2,24) = 1.29$, $p<.29$, partial Eta Squared = .10.

---

[1] Participants were encouraged to always keep their heads in front of the camera, but we did not want to force artificial constraints into the interaction such as a chin-rest. Consequently, in the voice-only condition (in which subjects had no visual cue indicating their face was out of the camera tracking range), some participants kept their face out of the range of the tracking space for more than fifty percent of the time. When eliminating these subjects from the sample, the statistical significance of the results did not change at all. Consequently we leave all subjects in the analyses for the sake of simplicity.

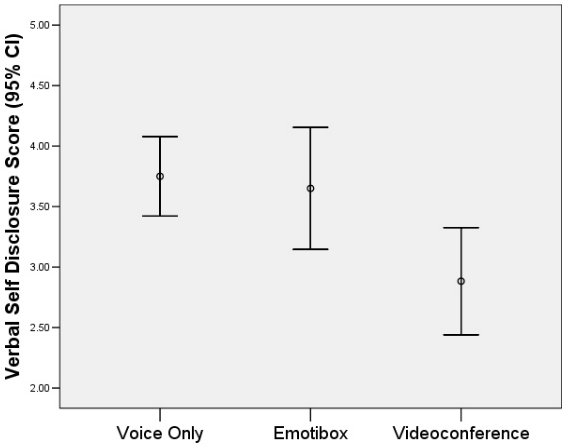


Figure 4: Verbal self disclosure scores by condition.

5.2. Nonverbal Disclosure

We ran a between-subjects ANOVA with condition (voice only, emotibox, and videoconference) and subject gender as independent factors and nonverbal disclosure score as a dependent variable. There was a significant effect of condition, *F*(2,24) = 6.45, *p*<.01, partial Eta Squared = .35. As Figure 5 demonstrates, there was more disclosure in the voice only condition than the emotibox or the videoconference conditions. The effect of gender was not significant, *F*(1,24) = .19, *p*<.67, partial Eta Squared = .01, and the interaction was not significant, *F*(2,24) = .65, *p*<.53, partial Eta Squared = .05.

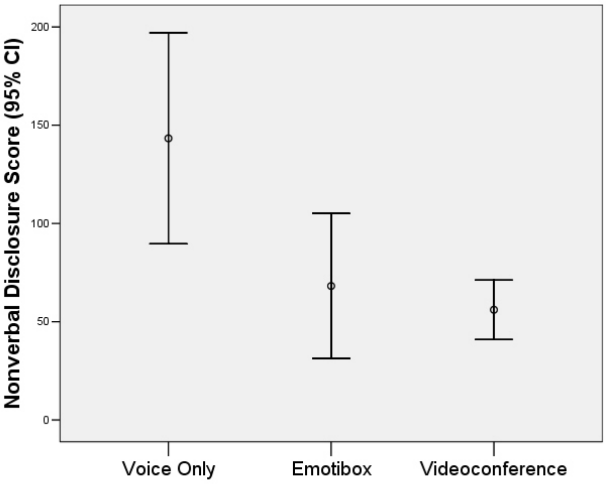


Figure 5: Average nonverbal disclosure score by condition. The scale of the Y-axis is normalized to the size of the head within the screen image and does not map onto a standard metric such as centimeters.

5.3. Copresence Ratings

We ran a between-subjects ANOVA with condition (voice only, emotibox, and videoconference) and subject gender as independent factors and self-report copresence score as a dependent variable. There was a significant effect of condition, *F*(2,24) = 3.55, *p*<.05, partial Eta Squared = .23. As Figure 6 demonstrates, there was less copresence in the emotibox condition than the voice only condition. The effect of gender was marginally significant, *F*(1,24) = 3.24, *p*<.08, partial Eta Squared = .12, and the interaction was not significant, *F*(2,24) = 1.36, *p*<.28, partial Eta Squared = .10.

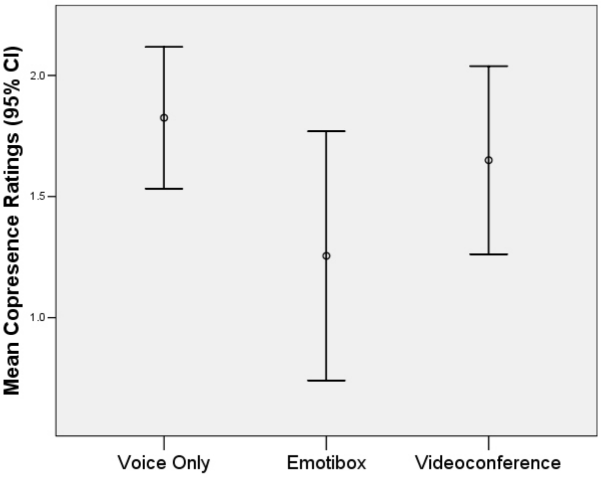


Figure 6: Mean copresence ratings by condition.

5.4. Emotion Detection

We ran a between-subjects ANOVA with condition (voice only, emotibox, and videoconference) and subject gender as independent factors and emotion detection score as a dependent variable. There was a significant effect of condition, *F*(2,24) = 18.05, *p*<.001, partial Eta Squared = .60. As Figure 7 demonstrates, there was worse performance in the emotibox condition than the voice only or the videoconference conditions. The effect of gender was not significant, *F*(1,24) = .12, *p*<.73, partial Eta Squared = .01, and the interaction was not significant, *F*(2,24) = .18, *p*<.83, partial Eta Squared = .02. In all three conditions, subjects were significantly above chance (depicted by the dotted line in Figure 7) at emotion detection.

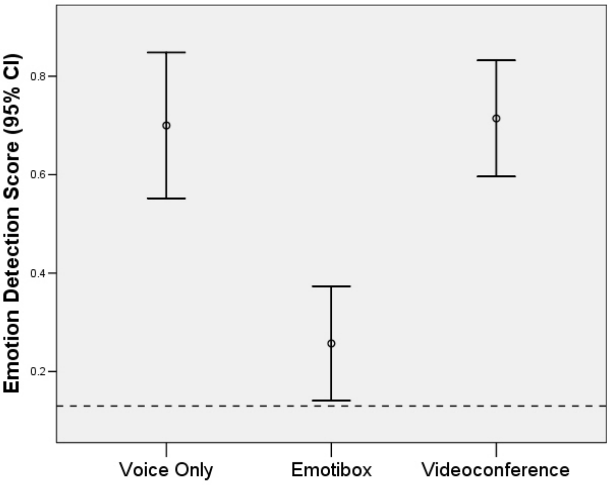


Figure 7: Mean percent correct on emotion detection task by condition. The dotted line indicates chance performance.

6. Discussion

6.1. Summary of Results

In this study, we compared the behavioral similarity and form similarity of avatar faces during real-time dyadic interaction. Our results demonstrated that, both verbally and nonverbally, people disclosed more information to avatars that were low in realism. In terms of verbal disclosure, subjects were perceived as less revealing, honest, and friendly in a videoconference then they were when interacting with either a text-only display or an avatar high in behavioral similarity but low in form similarity (the emotibox). In terms of nonverbal disclosure, subjects utilized more facial gestures and movements in a voice only interaction than in an interaction with either high behavioral realism (the emotibox) or high behavior and form realism (the videoconference). In other words, people emote more freely when their avatar does not express those emotions.

Overall, the emotibox proved to be a less effective interface than either of the two other alternatives in terms of copresence ratings and effectiveness in transmitting emotions. Nonetheless, without any training at all, on average subjects were above chance when attempting to identify the seven emotions with the emotibox, and on certain emotions were much higher than chance (e.g., 42% correct with “joyful”), which is encouraging considering that these emotions were expressed in a completely abstract fashion. With more elegant algorithms it should be quite possible to make more effective avatars that are high in behavioral similarity and low in form similarity.

6.2. Implications, Limitations and Future Directions

Earlier we discussed the defining characteristics of an avatar, and argued that a representation needs to have either high behavioral or form similarity in order to be utilized as an effective avatar in an interaction. In the current study, the emotibox was designed to elicit high behavioral similarity with low form similarity. However, by abstracting emotional expressions (as opposed to rendering the movements on a face-like digital model) we may have fallen short of our goal of producing high behavioral similarity. Participants may have been distracted by the foreign parameters of the box. In future work we plan on developing algorithms that are more stable (the same patterns emerge more readily across participants) and more intuitive (the mapping of color, shape, and orientation of the box is naturally tied to what we see on actual facial expressions).

Developing avatars that have high behavioral similarity and low form similarity is a worthy goal. The current study demonstrates that people are willing to disclose more personal information with an emotibox than with the avatar which is more realistic in form used in a videoconference. Unfortunately, the current instantiation of the emotibox elicited low copresence according to self report ratings and emotion recognition performance. If we can improve the quality of emotional transmittance of the emotibox, we can then create avatars in which people feel more comfortable using than ones highly realistic in form. Such avatars may be extremely useful for introverted students talking in front of a class in a distance learning scenario, patients interacting with a virtual therapist, and many other applications in which people interact with avatars in highly self-relevant and personal situations.

The current study is one of the first to use facial expressiveness as a dependent variable of copresence. Measuring people’s nonverbal facial disclosure can be an extremely powerful tool to uncover the elusively latent construct of copresence. Indeed, the finding that people utilize more facial expressions when the other interactants cannot see their avatars is quite counterintuitive, as one might predict more facial expressions to be used when another person can actually see those facial expressions. This counterintuitive finding supports the notion raised in the introduction that facial expressions are direct correlates of emotions, as opposed to a social tool that can be turned on and off strategically. Future work examining people interacting via avatars and embodied agents should build upon this methodology.

For example, research should explore the interplay between avatar realism and context. Even if the emotibox elicited low copresence and emotion recognition, this may not be important for some tasks or settings - and may in fact be an advantage. For certain object-focused tasks in CVEs, for example, participants may be completely focused on the task and not focus on each other’s faces. In this case an emotibox-type avatar could transmit only certain basic emotions that are designed to support the task (e.g., raising eyebrows translated into cyan cube color could transmit ‘I am concentrating’) which the collaborator could glance at occasionally without losing his or her concentration. Another type of avatar face might be developed for particular types of interpersonal interactions. The emotibox might, for example, transmit or signal only certain personal states, such as a smile translated into a yellow cube to signal ‘I am happy to continue our conversation’.

In turn, exploring different types of contexts will allow us to converge upon an optimal avatar design. In the current work, the emotibox avatar is at the most basic end of the continuum of form realism of avatar representations in CVEs. But it will be possible to ‘ramp up’ avatar realism by degrees. Further towards the realistic end of the continuum, there could, for example, be a cube with a human-like appearance (such as a cartoon face, not necessarily resembling the real user) and this could be given a more subtle range of emotions that are conveyable (for example, colors on the cheeks to convey degrees of shyness).

The current work also suggests new direction for measurement criterion in CVEs. Although presence and copresence are largely regarded as the ‘holy grail’ of virtual environments research, as CVE (and other new media) use increases, avatars will require different levels of self-disclosure and expressiveness, with the traditional notion of copresence weighed only as an additional factor in the mix. Findings such as those presented in the current paper will provide a useful tool for gauging the kinds of representations required for different forms of mediated communication, as well as providing insights into the nuances of face-to-face behavior that may be easier to measure and manipulate within CVE environments.

Furthermore with face-tracking and other technologies, users will be able to use self presentation as a mechanism to transform their avatar’s expressiveness. The possibilities for different forms of *transformed social interaction*– wearing different faces with capabilities for self-disclosure and emotional expressivity which can be changed ‘on the fly’ – offers potential for a number of training and related areas (see Bailenson & Beall [1] for other examples).

One of the most useful implications of the design of the emotibox is the idea of creating a framework within the notion of behavioral realism. Currently, behavioral realism is rarely discussed in a series of sub-dimensions. The emotibox raises issues in this regard. One dimension of behavioral similarity is the idea of *contingency*, the idea that for every performed behavior by the user, that behavior is tracked and then rendered on an avatar. Another one is *veridicality*, how much rendered behaviors resemble in terms of the actual animation. In other words, the emotibox from the current study was high in contingency but low in veridicality. A third type of realism is *correlation realism*. If not all behaviors of the human can be tracked, are there any behaviors that should be rendered? In other words, if it is not possible to track pupil dilation, but we know that pupil dilation correlates quite highly with heart-rate (which we can track), should we use probabilistic rendering of pupil dilation based on heart data? This is extremely important, given that tracking of human behaviors in real-time is currently quite difficult.

These areas of research and development will overlap, and there will be requirements for a variety of *degrees* of form and behavior realism in emerging media. Thus it is possible to envisage a range of avatar faces that could be combined in a pick-and-mix fashion to suit different types of interaction in CVEs, depending on the requirements for expressiveness and the task.

6.3. Conclusion

It is clear that avatar realism is critical to the future of collaborative virtual environment development. Highly realistic avatars with real-time facial form and tracking require more resources – both computationally and in terms of person-hours required to implement them. Moreover, the issue of the realism of digital human representations is a key question for a range of new media other than immersive virtual environments, such as videoconferencing, mobile telephony, online gaming, instant messaging and any other media that includes online representations of users. Understanding the relationship between form and behavioural realism is critical to begin examining the use of these new forms of media.

Appendix

Verbal Disclosure Question Set A:

1. Where do you live on campus?
2. Where did you grow up?
3. What do your parents do?
4. What has been the most stressful event of the last six months for you?
5. Of all the people you know, whose death would bring you the most sadness?
6. What's the longest relationship you've ever been in?
7. Tell me a little more about yourself.

Verbal Disclosure Question Set B:

1. What are you majoring in?
2. Do you have any siblings?
3. What's the scariest thing that's ever happened to you?
4. Do you think you're ready for a long-term romantic relationship? Why do you feel that way?
5. Which part of your body are you most uncomfortable with?
6. How much money do your parents make?
7. Tell me a little more about yourself.

Copresence scale:

1. How easily distracted were you during the interaction?
2. How easy was it for you to tell how your partner felt?
3. How responsive was your partner?
4. How often were your partner's behaviors clearly a reaction to your own behaviors?

Acknowledgements

The authors would like to thank Megan Miller for managing this project and Robby Ratan and Paola Worsley for assistance in data entry and analysis.

References

[1] Bailenson, J. and Beall, A. Transformed Social Interaction: Exploring the Digital Plasticity of Avatars. In R. Schroeder and A.-S. Axelsson (Eds.). *Avatars at Work and Play: Collaboration and Interaction in Shared Virtual Environments.* London: Springer, 2005.

[2] Bailenson, J.N., Beall. A.C., & Blascovich, J. (2002). Mutual gaze and task performance in shared virtual environments. *Journal of Visualization and Computer Animation, 13*, 1-8.

[3] Bailenson, J.N., & Blascovich, J. (2004) Avatars. *Encyclopedia of Human-Computer Interaction,* Berkshire Publishing Group, 64-68.

[4] Bailenson, J. N., Swinth, K. R., Hoyt, C. L., Persky, S., Dimov, A., & Blascovich, J. (in press). The independent and interactive effects of embodied agent appearance and behavior on self-report, cognitive, and behavioral markers of copresence in immersive virtual environments. *PRESENCE: Teleoperators and Virtual Environments*.

[5] Becker, B. and Mark, G. (2002), Social Conventions in Computer-mediated Communication: A Comparison of Three Online Shared Virtual Environments, in R. Schroeder (ed.), *The Social Life of Avatars*: *Presence and Interaction in Shared Virtual Environments*. London: Springer, 19-39.

[6] Bente (2004). Measuring Behavioral Correlates of Social Presence in Virtual Encounters. Paper presented in the *54th Annual Conference of the International Communication Association*, New Orleans, LA

[7] Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Hoyt, C. L., & Bailenson, J. N. (2002). Immersive virtual environment technology as a methodological tool for social psychology. *Psychological Inquiry, 13*, 146-149.

[8] Biocca, F.; Harms, C. and Burgoon, J.K. (2003). Toward a More Robust Theory and Measure of Social Presence: Review and Suggested Criteria', *Presence: Journal of Teleoperators and Virtual Environments*, vol.12, no.5, pp. 456-80.

[9] Bowers, J., Pycock, J. & O'Brien, J. (1996). Talk and embodiment in collaborative virtual environments. *Conference on Human Factors in Computing Systems (CHI'96)*, 58-65.

[10] Cheng, L., Farnham, S. and Stone, L. (2002), Lessons Learned: Building and Deploying Shared Virtual Environments, in R. Schroeder (ed.), *The Social Life of Avatars: Presence and Interaction in Shared Virtual Environments*. London: Springer, 90-111.

[11] Ekman P, Friesen WV (1976): Measuring facial movement. *Journal of Environmental Psychology and Nonverbal Behavior* 1:56-75.

[12] Ekman P, Friesen WV (1978): *Facial action coding system: A technique for the measurement of facial movement*. Palo Alto, CA: Consulting Psychologists Press.

[13] Essa, I. & A. Pentland. ``A Vision System for Observing and Extracting Facial Action Parameters'', In *Proceedings of IEEE Computer Vision Pattern Recognition Conference 1994*, Seattle, WA. June 1994.

[14] Garau, M. (2003). *The Impact of Avatar Fidelity on Social Interaction in Virtual Environments*. Ph.D. thesis, Department of Computer Science, University College London.

[15] Heeter, C. (1992). Being there: The subjective experience of presence. *PRESENCE: Teleoperators and Virtual Environments, 1(2)*, 262-271.

[16] Izard CE. 1971. *The Face of Emotion*. New York: Appleton Century Crofts.

[17] Lanier, J. (2001). Virtually there. *Scientific American,* April, 66-75.

[18] Lee, K., M. (2004). Presence, Explicated. *Communication Theory, 14*, 27-50.

[19] Michel, P. and El Kaliouby, R. (2003). Real time facial expression recognition in video using support vector machines. *Proc. 5th International conference on multimodal interfaces (ICMI)*, November 2003.

[20] Moon, Y. (200). Intimate Exchanges: Using Computers to Elicit Self-Disclosure from Consumers. *Journal of Consumer Research*, *26*, 323-339

[21] Picard, R. W. & S. Bryant Daily (2005), "Evaluating affective interactions: Alternatives to asking what users feel," *CHI Workshop on Evaluating Affective Interfaces: Innovative Approaches*, Portland Oregon, April 2005.

[22] Schroeder, R. (2002). Social Interaction in Virtual Environments: Key Issues, Common Themes, and a Framework for Research, in R. Schroeder (ed.), *the Social Life of Avatars: Presence and Interaction in Shared Virtual Environments*. London: Springer, pp.1-18.

[23] Schroeder, R. (2002). Copresence and Interaction in Virtual Environments: An Overview of the Range of Issues. In *Presence 2002: Fifth International Workshop. Porto*, Portugal, Oct.9-11, 274-295.

[24] Slater, M., Sadagic, A., Usoh, M., Schroeder, R. (2000), Small Group Behaviour in a Virtual and Real Environment: A Comparative Study. *Presence: Journal of Teleoperators and Virtual Environments, 9(1)*: 37-51.

[25] Sproull, L., Subramani, M., Kiesler, S., Walker, J., Waters, K. (1996). When the interface is a face. *Human-Computer Interaction, 11*, 97-124.

[26] Turk, M. and Kölsch, M., (2004), "Perceptual Interfaces," G. Medioni and S.B. Kang (eds.), *Emerging Topics in Computer Vision*, Prentice Hall.

[27] Walther, J. B. (1996). Computer-mediated communication: Impersonal, interpersonal and hyperpersonal interaction. *Communication Research, 23*, 3-43.

[28] Weisband, S. & Kiesler, S. (1996). Self disclosure on computer forms: Meta-analysis and implications. Presented at p*roceedings of the SIGCHI conference on human factors in computing systems*.

[29] Whittaker, S. (2002). Theories and Methods in Mediated Communication. In Graesser, A., Gernsbacher, M., and Goldman, S. (Ed.) *the Handbook of Discourse Processes*, 243-286, Erlbaum, NJ.

BASIC: A Believable Adaptable Socially Intelligent Character for Social Presence

Daniela M. Romano, Gary Sheppard, James Hall, Adam Miller, Zhinan Ma
D.Romano@dcs.shef.ac.uk, ac04jh@sheffield.ac.uk, ac04gs@sheffield.ac.uk;
ac04am@sheffield.ac.uk, ac04zm@sheffield.ac.uk
Computer Science, University of Sheffield
Regents Court, 211 Portobello Rd, Sheffield S1 4DP
United Kingdom

Abstract

Drawing inspiration from social science and psychology, a computational model of a personality model for a Believable, Adaptable, Socially Intelligent Character (BASIC) has been designed, implemented and tested to drive chimpanzees in a multi-agents scenario. The BASIC model can be customized to create different personalities that are able to trigger empathic responses in human spectators, otherwise known as social presence.

A multi-room event driven scenario, where queues propagate the social interactions amongst the characters, demonstrates the social interaction capabilities of the model embodied within the graphic visual representations. The system is efficient and can run on any mid-spec PC with over ten personalities being fully simulated.

The novelty of the approach lays in the factors combined in the personality model including mood, relationship-based memories, impulse based decision making, and gestural alteration through emotion. These are all steps towards the creation of virtual characters, able to elicit social presence.

Keywords--- **Social presence, multi-agents, synthetic personalities, empathy.**

1. Introduction

The creation of believable artificial characters has been the goal of many researchers in cross-disciplinary fields. Virtual characters are used with the aim of increasing the usability of the human-computer interaction in different ways, in particular in virtual environments to enhance the user experience and trigger social presence.

Biocca [1] defines physical presence, presence in virtual environments, self presence and social presence. Physical presence is the default sense of "being there". It is the basic state of consciousness that people attribute to the source of sensation from the physical environment. The sense of presence in virtual environments instead is a like a daydream in an imaginary environment [1][2]. Social presence has its roots in face-to-face interaction, and social interaction. Social presence is the sense of presence that is felt in mediated communication, where a user feels that a form, behavior, or sensory experience indicates the presence of another intelligence [1]. In this case, the amount of social presence is the degree to which a user feels access to the intelligence, intentions, and sensory impressions of another [1]. Biocca stresses that rather than seeing social presence as an insufficient replication of face-to-face communication, it should be seen as a simulation of another intelligence. Such simulation runs in the body and mind of the perceiver, and models the internal experience of some other moving, expressive body [1].

Furthermore [3] notes that social presence can be linked to a larger social context that includes motivation and social interaction. Social presence is the human ability to project oneself socially and effectively into a community [4].

In this paper we define achieving social presence as achieving *the illusion in the mind of the perceiver, that another intelligence exists in the environment. Such illusion is fostered by believable, to the perceiver internal experience, behavioral expressions of the character and its ability to engage in social interactions that trigger empathy in the human user.*

We present here a model and a realization of socially intelligent characters, able to adapt to the environment, with a tested ability to trigger emphatic reactions in the mind of the user.

Unlike other believable agents [5][6], the characters presented here express themselves only through their actions, gestures and facial expressions.

2. Background

Research on the link between realistic virtual characters and virtual presence has undergone some interesting advancement in recent years [7], but these are mostly based on characters as set actors in a preset scene.

2.1 Virtual personalities

The most common representations for modeling virtual humans in psychology are the OCC model of *Emotional State* [8] and *the Five-Factor Model of Personality* [9], both of which lend themselves to computer modeling [10][11][12]. Further work has proved that the full OCC model is not required for believable simulation of characters, and that a Restricted-OCC model, using only 10 of the original 22 emotions, can be reliably used [13]. Whilst not as accurate as the full model, the reduction in computational cost has proven enough to warrant the slight reduction in realism [10]. Kshirsagar [10]. also introduces *Mood* as a personality factor, suggesting that it is required to model slower changes of the current emotion; a concept that we found to be highly viable. Later works [12][14] also introduce memory to the personal state representation, indicating that current reactions and emotions are based on those that have passed in similar conditions. This improves the modeling of memory through the introduction of *Relationships*.

In addition to the emotional build-up of a character, it is also important to consider *Social Knowledge* and it's effects on the actions of an individual. Cervone [15] identifies several key areas of Social Cognitive Theory. They have been used in the personality simulation of our system. In order to co-ordinate social interactions, it is imperative for the agents to have some level of perception of their environment. [16] Rudomin et al. approaches this using a behavior map, but is somewhat limited in its domain. A more substantial event-driven model is proposed in [17] where events are fired into the environment for agents to react. The latter is the perception concept adopted here, as it is a well-proven approach as shown by similar techniques in modern day video games.

2.2 Character visualization

An important factor to consider in the visualization of interactive characters is the Uncanny Valley phenomenon [18], Figure 2.1, where slight inaccuracies in synthetic humans make the observer uncomfortable, reducing believability below what a less realistic representation would induce. An easy way to avoid all probability of being caught in this is to take a cartoon approach in the visual representation. Several systems have used this approach with good affect [14] [19] with minimal loss of user-empathy for the characters.

Considering the need for detailed control of the character a simple Forward Kinematical, skeleton based model is used.

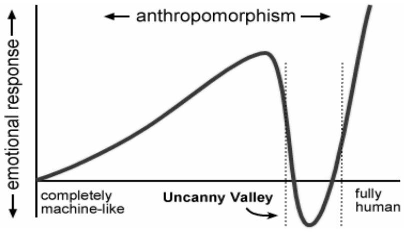


Figure 2.1 – The uncanny valley phenomenon

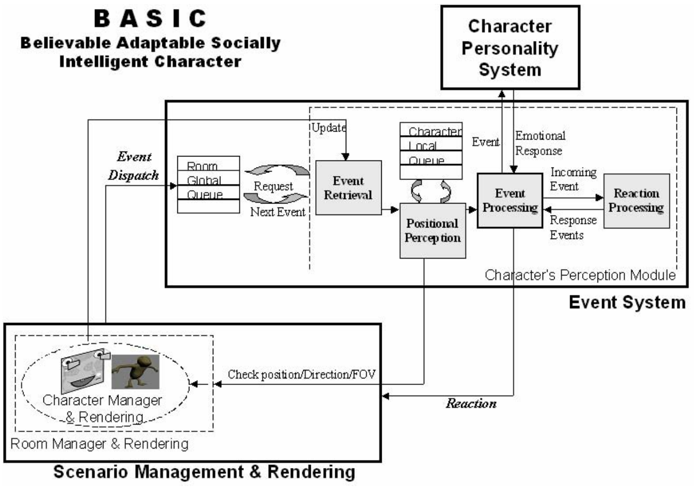
There are many approaches for facial animation from pseudo-bones to point-based manipulation [20]. This first version of the system has no natural language abilities, concentrating on the personality and behavioral simulations, consequently only basic expressions are required. For this purpose, a simplified muscle-based approach (loosely based on [21]) is adopted.

3. System overview

Here is a brief overview of the components that underpin the BASIC system, to aid the reader's understanding in subsequent sections, see figure 3.1.

The system is divided into three core components: Personality, Events System, Scenario Management & Rendering Engine. The primary focus of the research was on the level of believability created by virtual characters by using a dynamic model founded upon emotion, personality, mood, interpersonal relationships and impulse-based decisions. However, a model of this nature cannot inherently display its success without the other components.

Figure 3.1 – BASIC System Overview



The events system drives the interactions between the characters, enabling the passing of messages that inform one character of another's actions. These emotionally coded messages are placed in *queues* and processed by the character's *perception module* of they are directed to and those in the same room. The *event processing module* sends the interpreted event to update the internal status of the personality module receiving an emotionally coded response that is processed by the *reaction processing* sub-modules and consequently generates a reaction that drives the way the character responds and the scene is updated. The behavioral capabilities facilitated by the event system (in italics in figure 3.1) are scripted and external to the source code, to provide a more powerful data-driven simulation.

The personality system of each character receives events as stimuli to its internal state, processing each event it receives with regard to the emotional feelings, mood, personality type, social cognitive factors, and memory of previous interactions (relationship) with the other agents. At each event/interaction, the model generate an emotionally coded response of a certain intensity that the perception module uses to judge what response (if any) should occur.

The scenario management & rendering is required to visually demonstrate the results of our research. A house party, described in section 7, with multiple rooms, each having a specific scripted emotional attribute, has been created for this purpose. The Object-oriented Graphics Rendering Engine (OGRE)[1] has been used for scene management and rendering, populating the environment with bespoken cartoon-type human-like chimpanzee characters, to represent the model without implying any real-world restrictions that a human form would impose.

4. Personality model

The overriding methodology behind the implementation of personality is the dogma discussed in [22] were three factors: *environment, people and behavior* are constantly influencing each other. Behavior is not simply the result of the environment and the person, just as the environment is not simply the result of the person and behavior. The reciprocal nature of the theory has been captured in the model that is described in the following sections. A simple, trait-only, introspective approach to behavior, which does not take into account outside influences, may be able to provide a personality across certain contexts, but in different situations, behavior must vary. Therefore the aim is to produce a character that behaves in a consistent, yet variable manner, across a variety of social contexts learning from its experience of the world, as suggested by [23][24] this is a fundamental capability for autonomous animated agents.

Through the implementation of these theories in a simulated environment that is rich enough to depict whether or not the personality model is successful, we aim to show and evaluate the way that social factors have an impact upon a character's behavior, and ultimately their personality.

4.1 Representations

There are five data structures used in this system based upon three theories. These are represented in Figure 4.1, with the incoming event and the reaction. Three data structures are inspired by the OCC emotional model [8][13] and used to represent emotions, memory of relationships, and mood. The OCC is also used to code the emotional weighting of an incoming event and outgoing response. The second theory used is the five-factor model of personality traits [9] driving the personality, the third; the social cognitive factors [15] driving the behavior in social situations.

The OCC model of emotion is represented by five variables representing one of the emotional opposite pairs (fear/hope, joy/distress, hate/love, anger/pride, and relief/disappointment), ranging between -1 and 1. In addition to the normal range, the value -2 has been used to indicate an extremely negative emotion (i.e. extreme fear) and +2 an extremely positive emotion (i.e. extreme hope). In such cases an additional animation is required as explained in sections 6.1 and 7.

The Five Factors Model (FFM) is used to represent the characters' consistent personality traits that we see as the inner personality of a character, on top of which mood and emotions are built and expressed. The factors represented are agreeableness, neuroticism, extroversion, conscientiousness, and openness. Each of these variables has a different impact on the personality and is represented as a set of values in the range 0 to1.

Social cognitive factors indicate the way in which the social context determines how the character will behave socially. These analyze the environment and other characters therein against six factors: social knowledge, personal goals and standards, reflection about oneself, affective experience, expectation, and self-regulatory skill. They vary the manner in which different characters will interpret an event. Each factor is stored as a variable that can range between 0 and 1, representing the characters' ability to perform each of these cognitive skills.

4.2 Implementation

The five internal representations described above combine into a single personality model that can be distinct for each character in the environment.

The personality system as been conceived as an onion-like model where each layer takes care of one aspect of the personality. Inner layers are more stable that the outer layers to changes, and are represented in a lighter color in Figure 4.1 displaying a section of the personality onion.

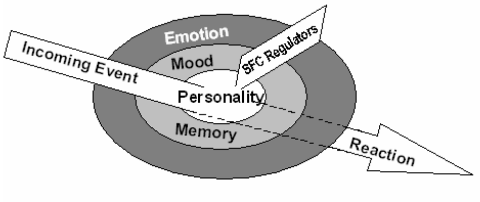


Figure 4.1 – A section of the personality onion with an incoming event passing through the layers (arrow)

The outermost layer is where emotions are formed. These vary easily according to the evolving events. Playing a game might make one happy, a discussion might make the character angry. At the second layer are placed the mood and the memory of relationships. Mood is considered as having a more permanent status than an emotion felt as a consequence of an event in the world. Mood and relationship are less subject to changes due to events in the world. Finally, the Inner layer is the personality coded by the Five Factors Model (FFM). Personality is less likely to change than emotion, memory of relationship and mood, and it is given to a character at creation, but is eventually shaped by the experiences of the world. The Social

---

[1] www.ogre3d.org

Cognitive Factors (SCF) are regulators that act across the three layers.

An incoming event is filtered through the three layers, see the arrows in Figure 4.1. An interpretation of the event is made at each layer according to the emotional impact, the relationship with the character generating the event, the mood of the character receiving the event and its personality. The incoming event is coded by OCC factors and the response to the event from the personality model is again a OCC factors combinations.

The following subsections define each of the layers and interrelations in detail.

4.2.1 Emotion and mood - The emotional state of a character in the personality model is represented by emotion and mood using the representation discussed above. Emotions and mood are respectively short to mid-term instances of a character's emotional state. The emotion is the more variable of the two, being heavily influenced by events, whereas mood is more involved in the interpretation of how an event is perceived and is a more permanent representation of emotional levels.

4.2.2 Memory - Memory is essential to the social cognitive theory of mind. Memory has an influence on how an event is interpreted, by scaling the values based upon social cognitive factors. Thus, characters that have had a previous experience with other characters will be influenced by their previous interactions. If the character following previous social interactions likes another, it will react in a more positive manner to the actions the latter performs, and vice-versa for negative past experiences. Memory gives you a preconception of the character, a guide as how to act in response to the event generated by it. Memory is implemented as a list of relationship mappings between a character and every other character that it has encountered in the world, see Figure 4.2. The relationship is represented using the OCC categories. Each time a new character is met, a new emotional set of response values is created and added to their memory of relationship. In addition, each time a character receives an event from another character, the memory is dynamically updated to refresh the opinion that the character has based upon the event received and past experience. Memory decays with time, so with the addition of new experiences the older, not updated records are deleted.

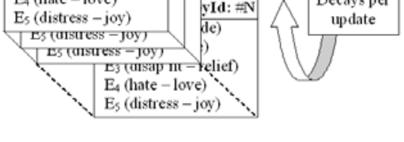


Figure 4.1 – Rapresentation of Memory)

4.2.3 Inner personality - The FFM variables describing the inner personality are given at birth and dynamically updated throughout the duration of the program (life of the character) depending upon the events received. In such manner the character's behavior is shaped by the environment, as described in section 4.3. These are the slowest to change due to external events.

Extroversion is a trait characterized by a keen interest in other people and external events, and venturing forth with confidence into the unknown [25]. In our system, it influences the weight of events in input compared with the mood and previous memory. It has an effect upon the magnitude of the reaction expressed.

Neuroticism is a dimension of personality defined by stability and low anxiety at one end as opposed to instability and high anxiety at the other end [26]. The model implements this as the speed at which the mood can be changed. A character with high neuroticism will quickly change emotion whereas a non-neurotic character will be more calm and stable.

Openness shows how willing people are to make adjustments in notions and activities in accordance with new ideas or situations [27]. The model implements this as a measure of how quickly your mood and FFM traits will changed based upon your perception of the world. A character with a low openness value will not change his personality as quickly as a character with a high openness value.

Agreeableness measures how compatible people are with other people or their ability to get along with others [27]. A character with low agreeableness is likely to have a much more adverse reaction to an event than a character that is highly agreeable.

Conscientiousness refers to how much a person considers others when making decisions [27]. In our model this again is used to determine how a character will respond to events from the environment.

The FFM variables represent long lasting personality traits, and while they have an influence on the way the system behaves, their key function is to help interpret an event in a different manner depending purely upon the character itself. This method of perception allows certain aspects of events to be focussed on by one character whereas another character could derive something completely different from the same event, thus generating diverse characters.

4.2.4 Social cognitive factors – They influence the way in which emotional and personality values interact. These factors are given to the character when they are created and are static through their life.

Reflection is a variable used to determine the degree to which emotion influences mood. When the mood is updated the emotion is weighted by reflection upon oneself. A low reflection value causes a smaller effect in the mood based upon the emotional values, whereas a high reflection will instigate a strong influence (See Eq. 4.5-4.8)

Affective experience is used to determine how much past experiences (memory) affect the current mood. Whenever an event is received, the amount of change in mood, based on the memory of the events instigator, is weighted by this value. A high value means that the memory has a large influence on the mood, a low value means that memory is largely ignored. (See Eq. 4.4 & 4.9)

Social knowledge is used to determine how much past experiences affect the character's interpretation of incoming events and how comfortable the character is in the environment. Social knowledge is used as a weight for the amount of influence that the character's previous emotional state will have on generating a reaction to an incoming event. For an example of the influence of social knowledge, see equations 4.1 and 4.2

Self-regulatory skill is used to regulate the speed at which a characters emotions, mood and memory will return to neutral over time. A character with a high regulatory skill value will have better control over their emotions than a character with a low value. In the system, this is represented as a fall-off weighting (See Eq. 4.10).

Personal goals and standards represent how much a character's personality has influence on how events are interpreted. Higher a character's personal goals and standards, the more they will allow their FFM to influence their world view depending upon their personality. A character with no personal goals and standards will not allow their FFM to have such an effect. (See Eq. 4.3)

Expectation denotes what the character would belief to happen due to a particular course of action. In our model, this is represented as changes of the character's emotional state based on the events they instigate themselves. A higher value causes a greater influence of memory in this process.

4.3 Adaptation

Time and events change a character's given personality at birth. It can change due to events received either from another character or by being in a room, or when the personality is updated by a constant logic tick indicating the passage of time. In such manner the character adapts to the environment it lives in. To facilitate the creation of the personality system, a graphical user interface, the MonkeyBrain, was created at first to simulate all possible inputs from the environment and how they change while the application is running.

There was much discussion and decision making about how the various abstract factors were to be implemented in the system, this was the most crucial point that needed careful consideration. Of course, the best way to evaluate a hypothesis is to test it with the target audience in the target environment. Since the authors aimed at the creation of a general basic model, to be able to see the effects of the different hypotheses the MonkeyBrain personality viewer was created as a means of inspecting the interaction of all the different values in the system. This viewer has been be used to simulate, test, and fine-tune all the different events that were expected to be received by the personality in the course of simulating it in a virtual environment. In the creation of a project such as this, there were several attempts at getting a plausible personality model working correctly. The first approach taken linked the 10 emotional values in a static manner to the five personality factors, in a similar way to [12]. Their approach provides a way of influencing personality based upon the emotions that the character receives. For example, if the character receives a large number of events that elicit an emotional distress, one might expect the personality trait of extroversion to be reduced, and similarly for the other personality factors. We initially drew up the table 4.1 to consider what effects the different emotions would have had on personality, where A = agreeableness, C = conscientiousness, E = extroversion, N = neuroticism, O = openness and a positive, negative or neutral versus indicates how we thought they would interact (respectively +ve, -ve, 0).

		FFM FACTORS							
		A	C	E	N	O	Total +ves	Total –ves	Total 0s
OCC FACTORS	Anger	-ve	-ve	+ve	+ve	0	**2**	**2**	**1**
	Pride	0	-ve	0	-ve	+ve	**1**	**2**	**2**
	Disapp.	-ve	0	-ve	0	-ve	**0**	**3**	**2**
	Relief	+ve	0	+ve	-ve	0	**2**	**1**	**2**
	Distress	0	0	-ve	+ve	0	**1**	**1**	**3**
	Joy	+ve	+ve	0	-ve	+ve	**3**	**1**	**0**
	Fear	0	0	-ve	0	-ve	**0**	**2**	**3**
	Hope	0	+ve	0	0	0	**1**	**0**	**4**
	Love	+ve	+ve	0	0	+ve	**3**	**0**	**2**
	Hate	-ve	-ve	+ve	+ve	-ve	**2**	**3**	**0**
	Total +ves	**3**	**3**	**3**	**3**	**3**	**15**		
	Total -ves	**3**	**3**	**3**	**3**	**3**		**15**	
	Total 0s	**4**	**4**	**4**	**4**	**4**			**20**

Table 4.1 – Potential Emotional/Personality Links

This first intuitive approach initially led to promising results. The major problem was that unless there were an equal number of emotional events received, the personality trait values tended towards the factor that is most represented. For example, joy/distress is a very simple emotional pair to identify in an event so many events will include a value of joy/distress in their emotional representation. This meant that, as the characters adapted over time, each character in the environment would up with the same personality. In Table 4.1, joy elicits three increases in the personality factors and one decrease, whereas a corresponding emotion of distress elicits one positive and one negative effect upon the personality model. This has the effect of disproportionately increasing the positive factors and they are not returned using the opposite end of the emotional scale.

Consequently the approach described in the section below was taken. Such approach does not statically link one emotion to one FFM factor. It was decided that the totals of the emotions should be considered and used such totals to change the personality traits. This meant that no matter which emotions were invoked, more often an equal change in both the FFM and the resulting interpretation of emotion could be examined over time.

The formulas that drive the personality adaptation are shown in the following sections. The way the weighting system was obtained is explained in section 4.3.3.

4.3.1 Receiving an event – When an event is received it is filtered through several steps, each having a sets of equations, as graphically shown in figure 4.1.

The first step interprets the affects of the event depending upon the FFM, social cognitive factors, mood, memory, and the current emotion. This step also uses a function that blends together the previous memories, mood, and emotion to return an instant reaction to the received event. This means that if the character encounters another it has never met before, memory plays no part in the equation. For each of the emotional affecters in the event, the following equations are used to update the personality model. If the character has no memory of previous interactions with the event source then Equation 4.1 is used, whereas if a relationship is existent then Equation 4.2 is used.

$$Ev' = Ev \cdot \left(\frac{SC_{SK}}{2}\right)\omega + (Em + M)\omega$$

Equation 4.1 - Step one A function

$$Ev' = Ev \cdot \left(\frac{SC_{SK}}{2}\right)\omega + (Em + M + Mem)\omega$$

Equation 4.2 - Step one B function

Where *Ev'* and *Ev* are the processed event and incoming event respectively, SC_{sk} is the social knowledge social cognitive factor in question, *Em* is the current emotional value, *M* is the current mood value, *Mem* is the current memory, and ω is a weighting used by the system depending upon the environment and is set empirically.

The second step uses the character's FFM values and the personal goals and standards factor of the SCF, to contribute to the characterization of an event, as shown in Equation 4.3. Four of the five FFMs have an influence upon the event by altering the values depending mostly upon neuroticism to vary the overall strength of the reaction. Personal goals and standards regulate the strengths further.

$$Ev' = Ev \cdot (FFM_N + \omega_1) \cdot (1 - FFM_O \omega_2)$$
$$+ (1 - FFM_C \omega_2) + (FFM_A - \omega_1) \cdot \left(\omega_3 + \frac{SC_{PGS}}{2}\right)$$

Equation 4.3 – Step two function

Where (excluding terms previously defined), FFM_N, FFM_O, FFM_C and FFM_A are the neuroticism, openness, conscientiousness and agreeableness factors of the FFM respectively, and SC_{PGS} is the personal goals and standards regulator.

Following this, in the third step, the event must again take into account the overall memory, this time considering the affective experience factor. This allows the model to represent how current experiences with other characters in the environment will affect the manner in which the character will perceive an event.

$$Ev' = Ev + (Mem_{Overall} \cdot SC_{AffEx}\omega)$$

Equation 4.4 – Step three function

Where (excluding terms previously defined), SC_{AffEx} is the affective experience of the character and $Mem_{Overall}$ is the overall memory.

Previously we included the memory in the reaction as a targeted relationship between the target and source. In this step, we affect the way that the character perceives the event based upon their collective memory, taking into account long-term memories.

The final step is to return the emotion to the perception as an instantaneous reaction to the event received. The OCC of the character is also updated as a result of the emotion received. Any changes to the personality model as a whole will take affect when the perception next calls the update personality function.

4.3.2 Personality update - With each invocation of the perception module, the personality is updated using the process described in this section. During the personality update, the FFM are updated based upon all the emotional values and a function is used for each of the five factors as shown in figure 4.5, 4.5, 4.7.

$$FFM'_C = (SC_{REF}\omega)FFM_E + \left[\frac{(1 - SC_{REF}\omega)}{2} + FFM_E \sum M\right]$$
$$+ \left[\frac{(1 - SC_{REF}\omega)}{4} \sum (M^2)\right]$$

Equation 4.5 – Conscientiousness update function

Where (excluding terms previously defined), FFM_E is the extroversion of the character and SCF_{REF} is the reflection upon oneself of the character. Conscientiousness tends to be increased when there are high levels of positive emotions in the mood, along with any extreme emotions. Additionally, the extroversion will influence the character to become more or less conscientious. The openness update function works as following. Depending upon the reflection upon oneself in the model, along with current openness, high values of extroversion and the rest of the FFM model will make the character more open to emotion. The same approach is used for agreeableness.

$$FFM'_O = (SC_{REF}\,\omega_1)FFM_O + (1 - SC_{REF}\,\omega_1)FFM_E \sum FFM$$
$$FFM'_A = (SC_{REF}\,\omega_1)FFM_A + (1 - SC_{REF}\,\omega_1)FFM_E \sum FFM$$

Equation 4.6 – Openness and agreeableness update functions

The final two factors in the model are extroversion and neuroticism. The way that these two factors are changed is based upon the sum of the square of emotional values of the mood. By considering these, it allows us to control the model so that the emotional and personality traits do not become too extreme in normal situations.

$$FFM'_N = (SCF_{REF}\omega_1)FFM_N + \left(\left(\frac{FFM_E(1 - SCF_{REF}\omega_1)}{2}\right)\sum(M^2)\right)$$
$$FFM'_E = (SCF_{REF}\omega_1)FFM_E - \left(\left(\frac{FFM_E(1 - SCF_{REF}\omega_1)}{2}\right)\sum(M^2)\right)$$

Equation 4.7 – Extroversion and neuroticism update functions

The next stage in the update of the personality is to update the mood. As shown in the equation 4.8 This is changed to reflect any changes that have occurred to the emotion, mood, and memory since last time the personality was updated.

$$M' = \left[\left(\frac{SC_{REF}\cdot\omega_1}{2}\right)(1 + FFM_O\cdot\omega_2)M\right] + \\ + \left[\left(1 - \frac{SC_{REF}\cdot\omega_1}{2}\right)(1 - FFM_O\cdot\omega_2)Em\right]$$

Equation 4.8 – Mood update function

The reflection about oneself element of the SCF is used to determine how much of the new emotion is to be used in the update to the mood. Next, the mood must be influenced by memory. The average memory of the character is used weighted by affective experience. This changes the way the character feels based upon what experiences it has incurred in its lifetime as shown by the equation 4.9.

$$M' = M + (\omega \cdot Mem \cdot SC_{AffEx})$$

Equation 4.9 – Memory update function

Where (excluding terms defined above) SC_{AffEx} is the affection experience of the character.

Finally, the gradual decline of the characters mood due to time is represented. Depending upon the self-regulatory skill of the character the mood tends towards zero at a different pace using the following decay function.

$$OCC' = OCC(\omega + (1 - \omega)\cdot SC_{REG})$$

Equation 4.10 – Mood function update

Where (excluding earlier terms), *OCC'* is the new mood and *OCC* is the current mood. The second decay step is then to update the emotion by substituting the self-regulatory skill of the SCF as the decay function, and the OCC values to those of the emotional state, into Equation 4.10. Finally, the memory is updated in a similar way. This time the decay function is used with the OCC' value set to the original memory, and each relationship's set of affecter values are reduced by the same percentage.

4.3.3 Weights used in the system - The Social Cognitive Factors (SCFs) of each character are set at run time, when a character is born. Characters with different SCFs respond to events, and evolve in different ways. In addition each equation has an associated weight used to vary the amount of overall influence that the calculation has in the engine. The SCFs allow us to change the behavior of the characters whereas the weights make sure that the variability of the behavior is believable and are used to calibrate the system. The weights are stored in an external file, avoiding hard-coding in the system and providing a way to tune the system without the need to re-code and recompile. Such file is called 'weights.ini'. The weights currently used have been determined with usability testing obtaining a finely balanced system. The weights descriptions are provided below covering their use. The most important weights were the values used in the update method. These were especially important, as initially it was not clear how quickly the personality was to be updated by the graphics engine and the perception module. Thus, the weights allowed to easily change the personality engine with respect to time. There are nine weights each with a different use within the system:

W_MEMORY: In the equations that concern memory, this weight varies how much previous encounters influence a reaction. This particular weight only influences equations that are used in the receive event process.

W_FFM: When an event is received, it is interpreted using the FFM. This weight has the effect of increasing or lowering the overall effect that the personality has on the interpretation of said events.

W_MOOD: When an action is received, the mood has an influence on how it is perceived. By increasing or decreasing this weight, the character will effectively become more moody.

W_MOOD_UPDATE: Every time the character is updated, there are several weights that are used to determine the amount of influence that each particular equation has in the module. This update is used to decide how much influence the mood has on the emotion every time the personality is updated.

W_FFM_UPDATE: Determined how much the FFM influences the mood every update.

Furthermore there are three regulatory weights. These are used to help determine how quickly a character's particular emotional traits return to a neutral value. The are: W_MOOD_REG: How fast the mood returns to normal over time. W_EMOTION_REG: How fast emotions return to normal over time. W_MEMORY_REG: How quickly memory deteriorates over time.

W_EVENT_RECEIVE: This decides how much of the reaction to actually return to the perception module. It represents how much the current emotion is affected by the event.

W_EXPECTATION: When you perform an event, the event has an influence on your personality. How much of an effect is influenced by the value of this weight.

5. Situation awareness

The personality model is a self-contained system that requires external events to evolve and produce emotionally coded responses. Such computational model has to be linked to a visual representation (scenario management & rendering) to show its effects and feed by an event system to inform & drive the character's social interactions by sending an event to a specific character or propagated it in a room. The event system represents the way the characters have a situation awareness. The event system is broadly composed by the scenario *rooms' global queues*, and the *character's perception module*.

Each character has a perception module that handles all incoming events providing events in response termed *reactions,* see Figure 3.1. To enable flexibility, customization and extensibility in the incoming-event/response, this relationship is scripted and independent of program compilation.

On a character update, the perception module retrieves all pertinent events from the global queue of the room in which the character is located. Each of these events is checked for *positional relevance* by checking relative position, direction and the field of view of the current character against the source of the event. If the source is a room, this test always returns success. The event is than processed though the *event* and *reaction processing* sub-modules. They inform the personality module of the new event updating its configuration accordingly and receiving an emotionally coded response that will update the scenario management & rendering.

Events have a time-out value, as an event might be only relevant for a certain period of time. This is useful to cope with cases where an event arrives before one character group is ready to process it and after such period the event is might be no longer relevant. The time-out is handled by the room global queue as it will be explained in the following section.

5.1 Global, local queues and priority order

Two types of event queue are used by the system: a global queue for each of the room in the environment and a local queue for each character.

<u>The room global queue</u> is the collection point for all the events that are relevant to that room generated by any character and the room itself (case of character leaving or entering a room described in the following section). Events added to the global queue are not immediately available for retrieval by the characters in the room, but are buffered, to prevent a scenario whereby an event is not received by all relevant characters, as shown in figure 5.1.

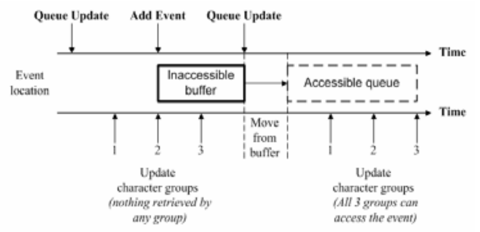


Figure 5.1 – Updating the global queue

If events were broadcast immediately, not all characters might have the chance to respond to them. Characters are updated in bursts to ensure that the rendering system is accessed frequently and with similar duration interim time differences. It must be noted that a room global event queue is effectively composed by a buffer and various separate sub-queues, one for each character that has registered with the room by entering in it. After the main loop of the program has performed updates on all the room's characters, it signals the event queue to update itself. Consequently all the events in the buffer are moved to an accessible area and reside there until the next queue update signal (broadcast), even if the timeouts of the events indicate event expiration. At the next queue update call all events that have timeout are removed. Broadcast signals are different from normal events because they have no specific target. When a broadcast signal is received by a room global queue, it inserts all relevant events into each of the character's room global sub-queues so that it is fairly processed by all characters.

Another type of queue exists in the perception module of each character, termed local queue.

<u>A character local queue</u> is coupled with the positional processing sub-module of a character and it is used as a temporary store for incoming events retrieved from the global queue that pass the positional tests. Local queues are far simpler than their global queue counterparts doing no special buffering.

Regardless of the type of queue, an order exists amongst the events based upon the priority value of each event. Priority values are: LOW, NORMAL, HIGH or CRITICAL and are assigned to an event at its creation in the script. The priority value of an event determines the insertion position of that event in the queues and consequently the processing order by the character. Events with the priority value CRITICAL have the ability to cease the processing of whatever event is currently being processed by the perception module to give relevance to the critical event. For example an explosion would be an event with a critical priority value, as it has to be reacted to immediately. Conversely, noticing that someone in the room is bored, unless relevant due to a particular relationship, would be an event with a LOW priority value.

5.2 The events types

All events have time-out value and priority value as described in the previous sections. There are seven different event types currently implemented in the system and listed below. They are implemented in an inheritance hierarchy, each deriving from an abstract base class “Event”, pointers to which are used throughout the system to facilitate polymorphism (examples found later in this section). The purposes of these events are as follows: *Null* - generated when nothing is happening externally to the character (there are no events to be processed). This is a tick of the time passing. *Gesture* - when a character performs a gesture this event informs the other characters in the same room. *Movement* - when a character begins to move to a certain location, in the current or a nearby room, this event is generated. *Mood* - used by rooms to affect the emotional status of characters residing inside it. *Delay* - allows a rest time before processing any other events. *Enter* - generated by a room when it detects that a character is entering the room, to inform all other characters in the room. *Leave* - Generated by a room when it detects that a character is leaving the room, to inform all other characters in the room.

All the events have the same field by defaults as they are inherited. Those are: event ID and animation names, timeout value, various booleans to affecting processing, values for enabling synchronicity of events, emotional affecters, priority value and destination type. Not all the fields are relevant to all event types, and the data content of each field differentiates an event from another, as it will be described in the following section.

5.3 Event processing

An event’s purpose is not just informing other characters that something is occurring, it is also used as a way to make the character firing an event do some processing. These events are termed *outgoing* events while they are *incoming* events when they are retrieved from the global event queues. The flow of the event processing system can be seen in Figure 5.2.

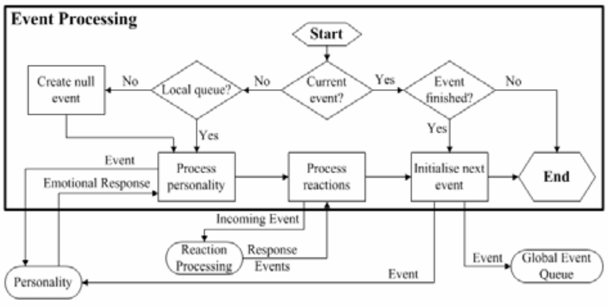


Figure 5.2 – Event processing procedure

Referring to Figure 5.2 if an event is currently being processed then it is checked to see if it is finished. The process() method is invoked and the boolean result examined. If the event has not finished then nothing more is performed in the event processing and the perception module has finished updating. Alternatively, if the event has finished the next event in the response list (if there is one) should be initialized according to its init(). Initialization and termination of events are explained in section 5.3.1. In addition, depending upon the boolean value of event fields named “global” and “personality”, the event is sent to the global event queue and personality model respectively.

In case there is no current event being processed by the perception module, the alternate path from the initial decision node is followed. A check is made to see whether there are any events within the local queue of the character and if there are any the first one is passed to the personality module for processing. If there are no events in the local queue then a *NullEvent* is generated and processed. Null events are explained in section 5.3.2 . Once an emotionally-coded response is received by the personality system, the event and the *emotional response (named incoming event* in Figure 3.1) are processed by the *Reaction Processing* sub-module that retrieve the scripted events/reactions and the first one returned is initialized.

Other important areas of the way the event system works require further explanations and are reported below.

5.31 Event initialization and termination - An event has polymorphic methods of initialization and termination checking. The standard behavior implemented in the abstract base class of the hierarchy: “Event” is an “init()” method with an empty body and a boolean “process()” method that always returns true. Most of the event classes need no extension over this default behavior however exceptions are the event types Delay, Gesture and Movement. Their “init()” and “process()” methods are depicted in Figure 5.3.

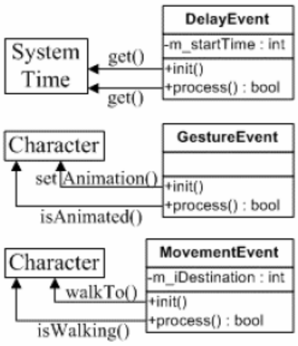


Figure 5.3 – Initialization and termination for delay, gesture and movement events

These methods are invoked for outgoing events only i.e. the response events from a reaction.

5.3.2 NullEvent – A null event (*NullEvent*) is generated when nothing is happening externally to the character. Its purpose is to provide scope for the programmer to script a reaction to nothing, that is say what the character should do when nothing else is going on. A good example of this would be a test for boredom; if the character is not being instructed to do anything, a null event is generated by the character itself and processed by its personality so that an emotional response can be obtained. If all the OCC category values lie within a certain distance from the midpoints then it could be deemed that the character is bored and trigger a boredom response e.g. walk to a random through the rooms until you meet someone else. Null events should never be sent externally from the perception module.

5.5.4 Synchronous events – A synchronous event is one that requires another character to perform the same or a different event that needs to happen at the same time in the scene. The only synchronous events implemented in our current version of the system are some gesture events. For example, if two characters were to shake hands see Figure 5.4, the animations of each should happen at precisely the same time to look believable. The synchronous gesture events use the event type's polymorphic initialization and "process()" methods. Unlike a normal gesture event, a synchronous gesture event does nothing inside its initialization method; instead, it defers setting of the gesture animation in the character until the process method is called. However, this only occurs if the target of the gesture event also has a synchronous event currently being processed. At every update a check is made against the target, as soon as the target indicates that it is processing a synchronous event, the animation is set. Two fields are used to indicated that an event is synchronous: the *sync* field that is a Boolean indicating whether or not the event is synchronous, and the *syncTimeout* field determines how long the event processing must wait before terminating the event. If synchronous events are performed, they terminate in the same way as normal gesture events.

5.4 Reaction Processing and scripting

Breaking down a human response to a certain occurrence into its constituent actions it makes it clear how such response can be described as a of pre-specified events. For example if a friend appears one might wave and to move near to talk to him/her. Thus a character appearance is a valid input event generated by the room it enters. In this case a plausible response sequence for a friendly character in the room is to wave, walk towards the new comer, shake hands and smile to such character. Conversely, an unfriendly character in the room other might walk towards the new comer, shake hands at fist, but then start arguing.

A character reaction-processing sub-module allows a series of events to be performed in turn, as a consequence of the emotional response input (see Figure 3.1). Each event of the sequence starts only when the previous one has finished. A *reaction* for the character is the incoming-event in input that requires a response, plus a list of conditional event list responses. Conditions for responses are in the form of five ranges, one for each of the OCC paired categories, and it can hold values in the continuous range from -1 to +1, indicating the normal rage of the emotion, or the extreme values -2 and +2 indicating an extreme emotion.

The incoming-event/reaction relationship used in the system are defined in scripted text, external from the source code of the actual program. This caters for customizable behavior and enhances the user's ability to provide a flexible interaction system within the simulation. In the script the incoming-events are defined as an angle-bracketed event type header, such as <GestureEvent>, followed by a list of fields and their values. The reactions are similarly defined using a <Reaction> header and then at least one pairing of a set of OCC value conditions and a list of response event IDs (corresponding to those in the scripted event file). This pairing can be repeated for as many different conditions and their associated responses as deemed appropriate.

6. Visualizing emotion

Irrespective of the quality of emotional modeling involved, it is of limited use unless this can be portrayed to the user in a believable manner. This does not necessarily mean photo-realistically, as it is clear that character empathy can be present in highly unrealistic, but believable context – Disney/Pixar films. In this section, are discussed the steps taken to empathetically portray the emotional state of the characters (or at least those that can be observed from visuals alone) within the social simulation.

Visual expression can be classified into three groups: facial expressions, postures, and gestures. The first two represent the pure emotion of the subject and are discrete, such as can be obtained from a photograph. The actual personality of the person cannot be determined in such a forthright manner, but it is portrayed through the actions and gestures performed by the character over time. Such actions/gestures are recalled from pre-stored animations rather than generated through a procedural language like in [5] [6].

6.1 Facial expressions

Due to the cartoon-like nature of our scenario, chosen in the wish to avoid the uncanny valley, facial expressions are not generated by a deformable 3D mesh, but are rendered in 2D to video memory before being used as a texture for the 3D head model. Not only does this make the rendering of the face considerably easier, it also creates a layer of abstraction so that the body model can be changed (to a human for example) without having to reform the facial mesh or rendering routine.

The definition of the face is based on Minimal Perceptible Actions [21]. Due to the 2D nature of the implementation the number of variables can be considerably reduced without losing the range of achievable expressions. We use only the following 11 facial expression factors: eyes open/closed, eyebrows raised/lowered, eyelid/eyebrow tilt, pupils horizontal, pupils vertical, mouth width, mouth smile/frown, mouth open/closed, where the last three have been defined separately for the left and right side of the face.

Prior to the model of each chimpanzee being rendered, a stack of facial features are positioned through reference to these values as shown in Fig. 6.1. Each of the 10 OCC variables has its own Expression Factor Set (EFS) representing the expression for that emotion. In addition to these there is also an eleventh EFS representing the neutral expression.

Figure 6.1 – The face stack.

Each emotional scale is, as indicated earlier, within the range -1 to 1, plus the extreme values –2 and 2 representing extreme static emotion. Beyond a magnitude of one, the values have no affect on the facial visual emotion that are therefore clamped, but when an extreme level is reached (value –2 or 2) the character performs an animation to further express the feelings. The EFS blending is done through a weighted average variation from the neutral expression (E = Expression).

$$E = \frac{E_n(1 - W_{\max}) + \sum E_x W_x}{(1 - W_{\max}) + \sum W_x}$$

Equation 6.1 – Expression function

Where E_n is the neutral EFS, E_x is the x^{th} emotional EFS, W_x is the x^{th} OCC value, and W_{max} is the maximum weighted OCC value.

6.2 Postures & gestures

The body of the chimpanzee is rigged with an Forward Kinematics (FK) skeleton for animation. Each factor of the OCC has its own animation frame defining the body posture for the emotion. At run-time, these are combined using a weighted average to produce a posture (A =animation) relative to the current emotion.

$$A = \frac{p\sum A_x W_x}{\sum W_x}$$

Equation 6.2 – Animation function

Where A_n is the x^{th} posture animation, W_x is the x^{th} OCC value, and p is the overall weighting of the total posture. The posture strength (p) remains constant unless the character is allocated a gesture to perform in which case it is reduced to allow blending with the gestural animation. Whilst this produces some hint of the emotion being experienced (besides the actual choice of gesture) the definition is small. This results in unrealistic movement, as real gestures are very heavily affected by the current mood of the instigator.

To better represent the emotional state of a character we introduce the concepts of a Gestural Affection Table, see Table 6.1. This identifies that depending on your current emotional mood, the speed and magnitude of your movements are altered. For example, an unhappy person makes small slow movements, while an angry person will be much more expressive in both magnitude and speed. These values are again averaged across the current OCC values for the character, and the animation altered by the resulting amount.

Emotion	**Magnitude**	**Speed**
Joy	+100%	+10%
Distress	-80%	-50%
Hope	-	-
Fear	-80%	+25%
Relief	+60%	-45%
Disappointment	-50%	-45%
Pride	+60%	-20%
Anger	+50%	+40%
Love	-	-
Hate	-40%	+35%

Table 6.1 – Gesture affection table

The table values have been generated through usability tests with four users, speeding up or slowing down and changing the amount a character moved during an animation until a level that was believed to be realistic for the emotion was being shown. Assumptions based on our everyday experience guided the testing, were for example anger produces a more expressive (thus higher speed and magnitude) display of energetic expressions than another emotions such as distress and disappointment. Furthermore the animations have been tested as described in section 7.

7. Scenario adaptation and impulse decision making

The prototype system created is described in this section. A scenario was defined along with a set of events and reactions to test the personality model. The code itself it is flexible and can be tailored towards any number of real-world uses. Characters in the world are born as a ‘blank page’ representing an average personality, with no relationships memories or personality tendencies. The characters’ personality can be easily be set with sliders on the control pads for mood, FFM, SCF, current emotion and relationship.

The world can be populated with an unlimited number of characters. The currently scenario created contains a neutral foyer with three attached rooms designed to instigate emotional change to those that reside in them: a disco (inducing joy), an haunted room (inducing fear), and a rather dismal pub where the bar is closed (inducing anger).

The adaptation system of the character is such that if a character is staying for a prolonged time in one of the rooms that induce a specific emotional aura having its mood modified, it reaches ad extreme emotional state performs an extreme gesture, to illustrate its status, and decides itself (with an impulse based decision) to return to the neutral foyer. At the return to the foyer the character's personalities it is altered accordingly to its past experiences.

If three or more chimpanzees are created within the central foyer, they automatically interact with each other using the events/responses scripted. In our prototype for example they wave, move closer to each other, shake hands, and then wander off into one of the three rooms around the foyer. Upon re-entrance to the foyer the characters are at a heightened emotional state, they do not quite act as cordially as at the start, while anyone else in the room, being in a normal state, waves at them. This wave induces a response according to the emotional state of the character. For example, if the character has been in the angry room it may return to the foyer and start slapping other characters for extreme anger, if it was in the disco it dances and gives the other characters a hug. Coming back from the haunted room a chimpanzees will cower in fear in response to a wave.

This short series of actions and reactions shows that the model drives the character to vary in response to their environmental situation. Depending upon what situations they have found themselves in, their behavior changes as a result. The way that their behavior changes, is consistent between different rooms, and changing the make up of their FFM and SCF characteristics further changes the manner in which their behavior varies.

8. Perception testing

To evaluate the quality of the graphical effort in portraying the emotional modeling, two tests have been performed, named static and dynamic emotion conveyance, to determine how successful the characters are at conveying empathy, thus social presence, in the viewers. The facial emotions, postures and gestures have been initially modeled based on both the six basic human facial expressions of emotions and the cartoon expressions, posture and gestures observed watching various Disney/Pixar type of characters. The expressions have been experimented with until all authors agreed on the emotion they conveyed and ready to be tested by a bigger audience. Description and results of such tests are presented below.

8.1 Static emotion conveyance

A first test evaluates static expressions and postures. Thirty volunteers where asked to participate to the test, they were all undergraduates in the computer science department, of mixed gender.

For this test the ten emotional extremes of the system (anger, pride, love, hate, joy, distress, fear, hope, relief, and disappointment) along with four mixed emotions (hope/pride, disappointment/distress, joy/relief, and fear/anger) and the neutral expression as a control, have been used. The aim was to evaluate how well the static emotions are conveyed by the character first by the facials alone, and then the impact of the inclusion of body posture on the emotion identification task.

The expressions were presented along with the five emotional pairs with a five-point scale (extreme negative emotion, moderate negative emotion, neutral emotion, moderate positive emotion, extreme positive emotion) between each of them. The user was asked to identify which of the emotions that the expression was trying to convey. The user records this by marking on each line the appropriate emotion that they think is displayed in the image of the character's face.

In the second phase of the test the volunteers were shown once more the same facial expressions (in a different order), but this time the still picture incorporated the body posture into characters emotional portrayal. The participants were again given the five 'sliders' and asked to rate the perceived emotion conveyed by each of the pictures.

In analyzing the results, an error distance measure has been used to see how far the users' perception of the emotion was, from the emotion the character was trying to convey was considered. The absolute amount that the user differed from the correct result for each of the emotions was collected. Then for each facial expression, an average of distance that the users misclassified the emotion was calculated. These were then collated and are presented in graphical form in Figure 8.1.

8.2 Dynamic emotion conveyance

A second test was performed to determine the effect of animating the different postures and expressions. The aim was to see if dynamic changes in the appearance of the character make the identification of the emotion easier. Twenty participants, once more undergraduate in the computer science department, watched the animations of one of the character in isolation on a 17" screen and were asked to identify the emotion pair being displayed from a set list. The experimenter drove the character's animation.

At first the character was moved from a neutral state to both extremes of an emotional pair, for three of the five emotional pairs, where when an extreme level is reached the character performs an animation to further express their feelings.

Second the volunteers were asked to identify which of the ten extreme emotions was being portrayed. Ten test events were created. Each of these represented one of the ten extremes of emotion, of which four were presented to each user. The results were evaluated on a purely hit or miss based method. If the user identified the correct emotion then a hit was recorded, a miss otherwise.

8.3 Discussion

8.3.1 Static emotion conveyance - The distance error shows that that overall, the inclusion of body posture makes the classification of emotions a great deal easier than just a facial expression on its own on a still pose. In most circumstances, the inclusion of body posture reduced the distance error found between the emotion that the system was presenting and the emotion that was perceived by the participants. The results are shown below in Figure 8.1.

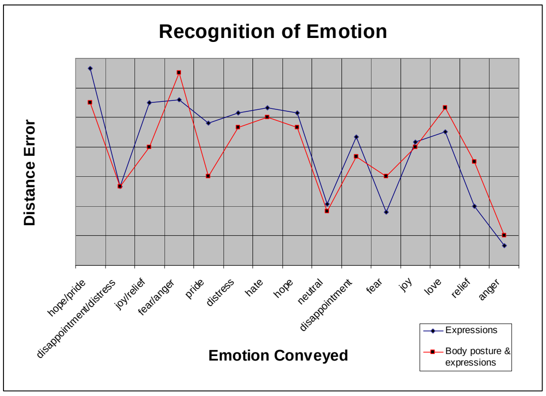


Figure 8.1 – Increase in recognition of emotion when postures are used.

The most surprising results are those that present a larger distance error, such as those of fear/anger, fear, love, and relief. What this suggests is that the user has become confused by the increased amount of information in the picture that obviously does not correspond to their personal beliefs as to how the emotion should be conveyed in an expression and body posture. However, in other cases a significant improvement in recognition has occurred.

In [Coulson04] a survey of static body postures was performed where an identification of the most appropriate static body posture that represents a particular emotion is found. If we take one of our strongest postures, fear, and compare it the posture provided in the paper (Figure 6.3) we can identify that several of the crucial elements are apparent – the outreached arms for example, although not as horizontal are significantly different to the neutral posture in our model. Although, in our model the posture was never meant to convey the full meaning of the emotion, simply a guide as to what the emotion should be. [28] results echo a similar point – the body posture can be very difficult to interpret as an emotional representation. For some emotions, with specific people, there would be no perceptible difference in posture; it can all be conveyed by the face. In contrast, there are some emotions, and some people, who are very expressive with their body movements, so will convey much more information. We feel that the key is to be consistent. If a consistent emotion is portrayed in a certain way then it will be understood given the context of the expression, the actions that resulted in that emotion, and the previous manner in which the character has acted.

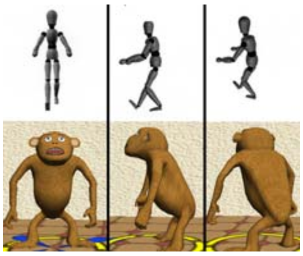


Figure 8.2 – Fear posture.

8.3.2 Dynamic emotion conveyance - In the second round of evaluation the dynamic emotions provided some interesting results. The overall correct classification of emotions improved significantly. If we examine Figure 6.4 we can see that many of the emotions are identified correctly in over 70% of cases.

The major problems occur in the identification of distress and hope. This is due to the vague nature of hope and the misclassification of distress, often with fear, although fear itself is very often identified correctly. One thing to take into account is that our character will produce an action when their emotional levels reach a certain point. This action then becomes the dominant feature of the emotion. For example, in relief, the character wipes his hand across his brow. This is a well known and obvious mannerism for someone who is relieved. Yet, if we examine some of our weaker results such as hope. The action when hope is to be displayed is very vague and perhaps does not convey the emotion of hope across very well.

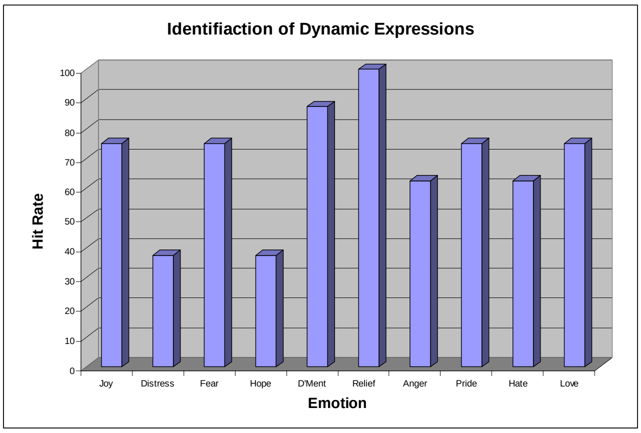


Figure 8.3 – The identification rate of dynamic expressions of emotion.

If we further consider the problem of misclassification of results we can conclude that certain emotions are classified incorrectly more often than could be expected due to simple error. The misclassification has been described pictorially in Figure 6.4. Here it can be seen that the emotions anger and hate are placed close to one another, this shows that anger and hate are often mistaken for each other. Similarly, emotions such as joy, love, and hope are often misclassified as each other by the user. In contrast, emotions such as disappointment and hate are rarely mistaken for each other in our experiments are a distance apart.

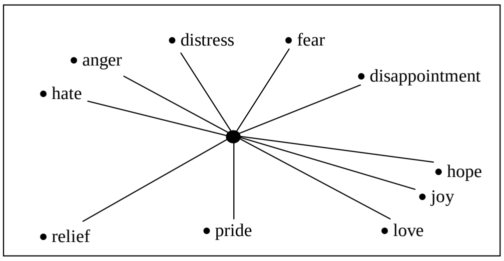


Figure 8.4 – Common emotional misclassification

In conclusion, our results show that although the actual classification of emotion into their specific type is quite poor, the users do recognize an emotion in a similar way, or at least an element of the emotion is recognized. The most common mistake of a user is to try and identify too many types of emotion in the one being portrayed.

9. Conclusion and applications

A model is capable of expressing the personality’s current emotional level through facial expression, body posture, and by performing gestures has been created and embodies into emotionally believable chimpanzees. The emotional levels of the character are displayed by taking the novel approach of mapping the OCC emotional values to a facial texture generator, and reinforcing them with posture and gestures. The model’s final expression of emotion has been validated using a series of user tests and we are satisfied that our simulation is successful in conveying emotions in a way that can be understood in a meaningful manner.

The personality model takes the approach of combining the consistent personality traits represented by the five factor model [9] and the environmental and behavioral regulators that are the social cognitive factors [Cervonne99] using an emotional model represented using a reduced version of the OCC model of emotion [8]. The OCC model is used to represent short-term emotion, long term mood, and a relationship based memory. This provides a means of representing time-dependent personality adaptation.

The system is event driven to provide situation awareness and the characters use their personality model to adapt to the environment and decide their actions following a scripted variability in response to given events depending upon their emotional levels and personality. The scripting system makes the characters respond to events that are not constrained to one particular context, making our system potentially applicable in many areas such as training, behavioral pattern analysis, and computer games.

References

[1] Biocca, F. The Cyborg's Dilemma: Progressive Embodiment in Virtual Environments. Journal of Computer Mediated Communication, 3/2. 1997

[2] Slater, M., & Usoh, M.. Representations systems, perceptual position, and presence in immersive virtual environments. Presence, 2/3, 221-233. 1993.

[3] Tammelin, M. From Telepresence to Social Presence: The Role of Presence in Network-Based Learning Environment. In Tella,S.(ed.) Aspects of Media Education: Strategic Imperatives in the Information Age. Media Education Centre. Department of Teacher Education. University of Helsinki. Media Education Publications 8. 1998

[4] Rourke, L., Anderson, T., Garrison, D.R., & Archer, W.. Methodological issues in the content analysis of computer conference transcripts. International Journal of Artificial Intelligence and Education, 12. 2001

[5] Loyall, A.B, Bates, J. Personality-rich believable agents that use language. International Conference on Autonomous Agents, *Proceedings of the first international conference on Autonomous agents*, Marina del Rey, California, United States, 106-113. 1997.

[6] Rickel, J., Marsella, S., Gratch, J., Hill, R., Traum, D., Swartout, W. Towards a new generation of virtual humans for interactive experiences. *IEEE Intelligent Systems*, 2002.

[7] G. Papagiannakis, H. Kim, N. Magnenat-Thalmann. Believability and Presence in Mobile Mixed Reality Environments, *IEEE VR2005 Workshop on Virtuality Structures*. 2005

[8] Ortony, A., Clore, G., & Collins, A. The Cognitive Structure of Emotions. Cambridge University Press. 1988

[9] Digman, J. M. Personality structure: Emergence of the five-factor model. Annual Review of Psychology, 41, 417-440. 1990

[10] Kshirsagar, S. Virtual Humans Personified, *Proceedings Autonomous Agents Conference (AAMAS)*, 356-359. 2002

[11] Dugdale, J., El Jed, M., Pallamin, N., Pavard, B. Modelling Character Emotion in an Interactive Virtual Environment. *AISB Symposioum, 29 March, 1 April,Leeds, UK.*. 2004

[12] Romano, D.M., Wong, A.K.L. Personality Model of a Social Character. *Empathy in Human Computer Interaction Workshop, HCI 2004, 6 Sept., Leeds, UK*. 2004

[13] Ortony, A. On making believable emotional agents believable. Trapple, R. P. ed.: Emotions in humans and artefacts. MIT Press, Cambridge, USA. 2003

[14] Louchart S., Romano D.M., Aylett R., Pickering J. Speaking and Acting - Interacting Language and Action for an Expressive Character. *Intelligent Virtual Environments and Virtual Agents Workshop*. Mexico City, Mexico. 2004

[15] Cervone, D, Social-Cognitive Theories and the Coherence of Personality. In D. Cervone & Y. Shoda (Eds.), The coherence of personality: Socialcognitive bases of consistency, variability, and organization, 3-37. 1999

[16] Rudomín, I., Perez, F., Millan, E. Groups and Crowds with behaviors specified in the environment. *Intelligent Virtual Environments and Virtual Agents Workshop*. Mexico City, Mexico. 2004

[17] Delgado-Mata, C. Characterising Emotion Signalling and Affected Group Behaviour in Virtual Mammals. *Intelligent Virtual Environments and Virtual Agents, IVEVA 2004, ITESM CCM, Mexico, CEUR Workshop*, 97. 2004

[18] Mori, M.. The Buddha in the Robot. Charles E. Tuttle Co. ISBN 4333010020. 1982

[19] Paiva A., Dias J., Sobral D., Aylett, R., Zoll C., Woods, S. Caring for Agents and Agents that Care: Building Empathic relations with Synthetic Agents. *Third International Joint Conference on Autonomous Agents and Multiagent Systems*, 194-201. 2004

[20] Parke, F.I.. A Parameterized model for facial animation, IEEE Computer graphics and applications, 2/9, 61-68. 1982

[21] Kalra P, Mangili A, Magnenat Thalmann N, Thalmann D. 3D Interactive Free Form Deformations for Facial Expressions, *Compugraphics, Siesembra, Portugal*. 1991

[22] Glanz, K., Rimer, B.K. & Lewis, F.M. Health Behavior and Health Education. Theory, Research and Practice. San Fransisco: Wiley & Sons. 2002

[23] Blumber, B. Learning in character: building autonomous animated characters that learn what they ought to learn. In O. Balet, G. Subsol, and P. Torguet (Eds.): ICVS 2001, LNCS 2197, 113–126. 2001.

[24] Blumberg, B., Downie, M., Ivanov Y.A., Berlin, M., Johnson, M., P., Tomlinson, B. Integrated learning for interactive synthetic characters. *SIGGRAPH 2002:* 417-426

[25] Ewen, R. B. Personality: A topical approach. Mahweh, NJ: Erlbaum. 1998.

[26] Pervin, L. A. Personality: Theory and research (5th ed.). New York: Wiley. 1989.

[27] Popkins, N. C. The Five-Factor Model: Emergence of a Taxonomic Model for Personality Psychology, http://www.personalityresearch.org/papers/popkins.html. 1998.

[28] Coulson, M. Attributing Emotion to Static Body Postures: Recognition Accuracy, Confusions, and Viewpoint Dependence. Journal of Non-Verbal Behavior, Kluwer, 28/2. 2004.

Virtual encounters. Creating social presence in net-based collaborations

Sabine Rüggenberg, Gary Bente, Nicole C. Krämer
University of Cologne
{sabine.rueggenberg@uni-koeln.de, bente@uni-koeln.de, nicole.kraemer@uni-koeln.de }

Abstract

The paper discusses the experience of "social presence" as a relevant effect dimension of avatar-mediated net-communication. Special attention is paid to measurement issues combining subjective verbal reports of social presence with objective behavioral data relating nonverbal activity, visual attention and person perception. Data will be reported from a study comparing the effects of different real time communication modes (text, audio, video and avatar) in a shared collaborative workspace. Results point to a significant difference between text and all other communication modes, indicating that audio, video and avatar systems worked similarly well in creating an experience of social presence. Analyses of behavioral data yield similar levels of visual attention for both video and avatar conferencing modes, which is however decreasing over time. The data raises critical questions about the added value of avatar systems and the specific requirements those systems have to meet to prove superior to mere audio/video transmissions.

1. Introduction

Technologies for computer mediated communication (CMC) are advancing rapidly, overcoming early restrictions of text-based communication via the internet. Shared workspaces and collaborative virtual environments (CVEs) allow for real time information interchange and the synchronization of distributed working efforts over distance. Especially in the field of net-based collaboration and cooperative learning these developments were mainly driven by one goal: To improve work effectiveness by creating shared knowledge and coordinating problem solving activities. The potential limitations of this primarily task-oriented perspective on mediated collaborations have only recently been noticed. In this line Redfern and Naughton [1] state: “By focusing solely on work effectiveness, we risk missing out on social richness – this has indeed been a problem with technologies such as video conferencing, which typically provide spaces for interaction but not social places as meaningful platforms for communication” (p. 207).

Aiming at a more personalized and emotional communication via the net Redfern and Naughton [1] (p. 206) refer to empirical evidence that the use of avatars can play an important role: „CVEs can provide richness of expression and personality, as well as “identity persistence” via appropriately detailed and customizable avatars. By fostering users’ interest in one another’s characters we will support the development of sociability and community”. Avatars allow to overcome constraints of mere text-based or audio communication by including nonverbal communication channels. In contrast to video conferencing systems avatar platforms provide additional communication bandwidth without loosing specific degrees of freedom which we much appreciate in CMC, i.e. avatars can convey nonverbal cues without necessarily disclosing the person’s identity or triggering prejudices based on physical appearance (e.g., gender, culture, age, attractiveness). At the same time avatars as embodied representations allow people to allocate themselves in a shared virtual space [2] and simultaneously handle the shared virtual objects and thus are expected to create an experience of co-presence [3].

Based on such observations the concept of “social presence” has emerged as a central variable in evaluating possible socio-emotional effects of virtual encounters. Biocca, Harms and Burgoon [4] comment: „The assessment of satisfaction with entertainment systems and with its productive performance in teleconferencing and CVEs is based largely on the quality of the social presence they afford“. Social presence is here broadly defined as a “sense of being together”, based on „mediated representations of humans via text, images, video, 3D avatars and in artificial representations of humanoid or animal-like intelligence including virtual humans, agents, computers, and robots” (p. 3). Based on previous research we conceptualize social presence as a basic experience of spatial co-location, emotional closeness and social relatedness [5, 6] on which more specific interpersonal effects which are considered relevant for successful net-based collaboration, such as “interpersonal trust” [7] can built up.

The dimensional structure of the complex psychological variable as well as its relation to the other concepts, and its impact on net-based collaborations however has not been explored in detail so far. Also, we lack systematic data on the influence of avatar-mediated nonverbal signals on social presence as compared to other means of computer-based real-time communication. The current study aims to provide empirical clarification regarding the dimensionality of social presence, its interrelation to other relevant social psychological variables in CMC, and the influence of different communication modalities (text, audio, video and avatar conferencing) on social presence and collaborative behavior. In contrast to the existing approaches our study used behavioral process measures (nonverbal activity and visual attention) to complement subjective verbal assessments of social presence.

2. Method

142 participants (68w/ 74m) collaborated in a management assessment centre task, making a decision with respect to selecting the right applicant for a predefined job. The participants interacted in same-sex dyads using a low immersive open desktop communication system consisting of the shared workspace “Cool Modes” [8] and a real time communication window. “Cool Modes” interactions were performed by means of a graphic tablet allowing to make notes, place, edit or remove shared information needed to solve the collaborative task. Cyber Gloves, Polhemus motion trackers and a high resolution eye-tracking system were used to collect nonverbal data: (head and upper body movement, hand and finger movements eye movements and gaze direction). Nonverbal data was stored for all subjects for later analysis. In the avatar conditions the nonverbal data was used to animate the virtual representatives of the communication partners in the communication window in real time. Two types of avatars were available: a cartoon-like low fidelity avatar (LFA), which was rendered in our own rendering software and an anthropomorphic high fidelity avatar (HFA), which was rendered by a commercial 3D animation tool (Kaydara Filmbox©). The participants were randomly assigned to one of five possible communication settings: (1) text only, (2) audio only, (3) audio + video, (4) audio + low fidelity avatar (LFA), and (5) audio + high fidelity avatar (HFA). In the audio mode the communication window was empty, in the text mode it served as a chat window. In the video and in both avatar modes it was used to display the nonverbal behaviour of the vis-à-vis. Figure 1 shows the experimental setting and screen shots of the different interface modalities.

Social presence was measured by means of 58 five-point Likert scale items based on the dimensions introduced by Biocca et al. [4], Kumar and Benbasat [9], and Nowak [10] and Tu [11]. Further a 20 item questionnaire was created to measure interpersonal trust as a two dimensional construct as described by Kanawattanachi and Yoo [7] and Nowak [10]. At last a set of 25 bipolar items was used for the measuring of mutual person perception and a set of 21 items (5 point Likert scale) for the aspect of perceived communication effectiveness. All items sets have already been evaluated and proved their internal consistencies during previous mediated communication studies [5, 12]

3. Results

3.1. Results of the principal component analysis and internal consistency tests of the questionnaires

Principal component analysis (Varimax rotation) and internal consistency tests were conducted across the four item sets. For the aspect of social presence the analysis yielded a four factor solution explaining 52.14% of the total variance (*closeness, co-presence, contingency, attention*). As expected for interpersonal trust a two factor solution was found, explaining 50,43% of the variance (*cognition based trust, affect based trust).*

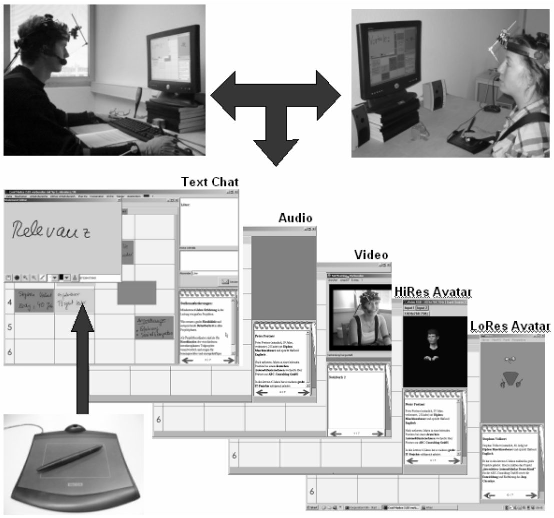


Figure 1 Technical setup and experimental conditions for the media comparison study.

Three components resulted from principal component analysis for the person perception items, explaining 46,77% of the total variance (*immediacy, assertiveness, competence*). At least the principal component analysis of the perceived communication effectiveness items resulted in a four component solution explaining 55.9% of the variance (*satisfaction, clarity, impression management, relevance*). The Cronbach´s alpha values for all resulting scales were good to excellent. Table 1 shows the factors and the consistency measures.

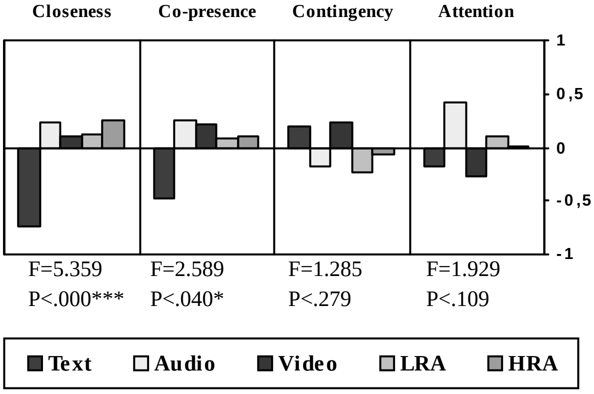


Figure 2 Technical setup and experimental conditions for the media comparison study

Table 1: ANOVA comparison of five communication modalities (mean factor score)

	Text	Audio	Video	LRA	HRA	F	P	(Scheffé Test)	Cronbach's Alpha
Social Presence									
Closeness	-.730	.233	.103	.126	.248	5.359	.000	1>2 (p = .015) 1>3 (p = .027) 1>4 (p = .030) 1>5 (p = .003)	.907
Co-presence	-.481	.259	.217	.087	.098	2.589	.040		.804
Contingency	.200	.180	.232	-.240	-.066	1.285	.279		.734
Attention	-.169	.426	-.277	.102	.018	1.929	.109		.653
Interpersonal trust									
Cognition based trust	-.207	.119	.049	.028	.031	.398	.809		.878
Affect based trust	-.626	.408	.018	.202	.088	4.331	.003	1>2 (p = .009) 1>4 (p = .043)	.831
Person perception									
Immediacy	-.830	.433	.048	.126	.248	7.511	.000	1>2 (p = .000) 1>3 (p = .015) 1>4 (p = .010) 1>5 (p = .001)	.904
Assertiveness	-.223	-.003	-.007	.311	-.033	.922	.453		.822
Competence	.014	.339	.087	-.170	-.025	.835	.505		.664
Perceived communication effectiveness									
Satisfaction	-.928	.031	.219	.267	.330	9.818	.000	1>2 (p = .009) 1>3 (p = .000) 1>4 (p = .000) 1>5 (p = .000)	.875
Clarity	-.150	.434	-.205	.076	-.033	1.585	.182		.814
Impression Management	.053	.126	-.095	.062	-.083	.257	.905		.814
Relevance	-.062	-.529	.145	.130	.156	2.120	.082		.659

3.2 Media differences in social presence, interpersonal trust, person perception, and perceived communication effectiveness

ANOVAs with ex-post Scheffé tests were conducted for all factors to determine the differential influence of the five communication modalities on the social presence, interpersonal trust, person perception, and perceived communication effectiveness aspects.

For the social presence scales significant differences were only found for the factors "closeness" and "co-presence" (see figure 2). With respect to "closeness" the text condition proved to be significantly different from all the other modalities, indicating that the provision of an analogous real-time channel alone – be it audio, video or avatars - was sufficient to increase the experience of emotional closeness, immediacy, and mutual understanding.

The significant difference in the "co-presence" factor revealed by the ANOVA was not reflected in the pairwise post-hoc comparisons. However, the text condition also proved to be the one that scored lowest on co-presence (see table 1).

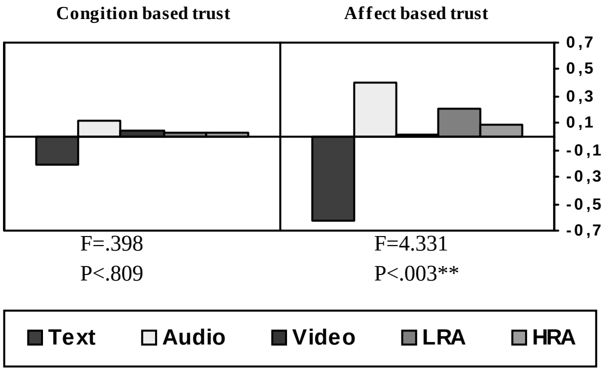


Figure 3 Media differences in the trust factors

For the aspect "cognition based trust" no significant results could be found. For "affect based trust" the results attained significance, indicating that - in contrast to all other modes - the text did not produce positive levels of trust. Only the difference between text chat and both audio and high resolution avatar reached significance (see figure 3 and table 1).

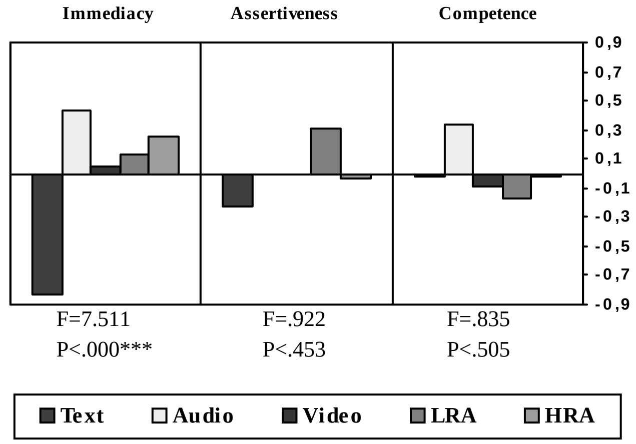


Figure 4 Media differences in person perception

The different communication groups varied significantly with regard to the aspect of "immediacy/social evaluation". However, direct comparisons show that this is only due to the text condition, as this is the only condition that differs from all other settings (see figure 4 and table 1).

The different communication groups varied only with regard to the aspect of "perceived interaction effectiveness". Direct comparisons show that again only text was significantly different from the other four communication technology settings (see figure 5 and table 1).

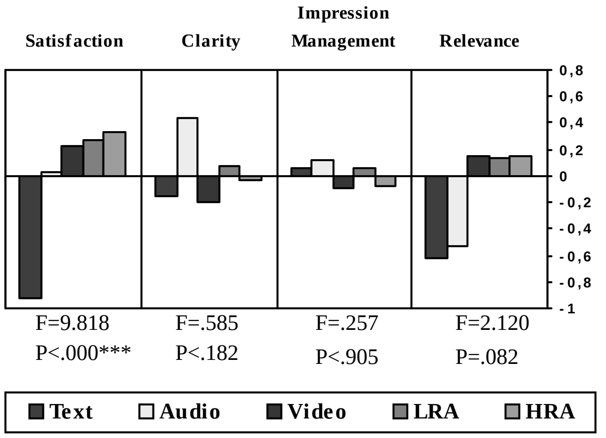


Figure 5 Media differences in the communication effectiveness factors

3.3 Media differences in nonverbal activity and visual attention

As behavioural indicators of social presence we analysed the parameters head orientation (exposing the own face), gaze direction (being attentive to the vis-à-vis's appearance and nonverbal activity) and nonverbal effort (defined as overall movement complexity). It was assumed that the relevance of nonverbal information as provided by the audio-video conditions (video, avatar) would lead to higher levels of exposure (head orientation towards the communication window, higher levels of visual attention to the communication window (directed gaze) and higher levels of nonverbal activity (movement complexity, time spent moving)

To separate the communicative use of the visual channels from orientation responses towards moving stimuli and curiosity effects (innovation effects) comparisons of the media were done for three consecutive time segments at the beginning of the interactions. A decline in visual attention and nonverbal activities could thus indicate a loss of interest over time which would mean that the nonverbal signals were only relevant for first impression checks, or due to orientation responses, which habituated over time or just to the newness of the medium. As illustrated by figure 8 the *number of directed gazes* towards the communication window was highest in the text mode (reading activity) and lowest for the audio mode (window had no display function). This data was in line with our expectations. For video as well as for the LFA condition the communication window reached nearly the same level of visual attention as in the text mode. HFAs however attracted only about 50% of the attention of the other AV modes. ANOVA result for the whole observation time was highly significant (F=11.317, p<.000). Post-hoc tests showed that the audio mode was significantly different from chat as well as from video and LFA, which reached the highest levels of visual attention during the first 6 minutes of interaction. There was however a significant drop of visual attention over time, indicating a loss of interest after the first inspection. Visual attention dropped by 10% in the video and low resolution avatar condition in the third 3-minute sequence. The decrease in visual attention in the three audio-visual modalities indicates that nonverbal behaviour was not primarily used for interaction fine-tuning, but for first impression formation. In the HFA condition the drop of visual attention was stronger (20%) and already occurred in the second sequence. It has to be mentioned that the average time the gaze was directed towards the communication windows in the first three-minute sequence was highest for the HFA, indicating an increased level of curiosity and a more persistent visual inspection in the beginning. After this, however, the total number of gazes as well as average duration of directed gaze dropped significantly:
The data suggests that there might have been higher expectations with respect to the social information provided here than could be met by the avatar. One possible reason for this could be that the lack of facial expressions could be attributed to the technology in case of the low resolution avatar but not in case of the HFA.

The statistical analysis of head orientation indicated that the heads of the interlocutors were more upright (F = 13.697, p <.000) and right oriented (F = 6.475, p < .000) - i.e. in direction of the communication window: upper right corner - in the audio-visual conditions as compared to the text or audio condition. Figure 7 shows the deviations from the mean position for all conditions, indicating that the face was more exposed to the partner in the audio-visual conditions as compared to text or audio.

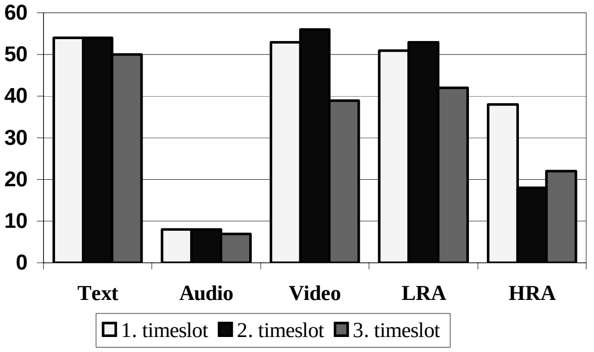


Figure 6: Number of gazes towards the communication window in three subsequent 3-minute timeslot

Significant differences were also found in the degree of nonverbal activity as reflected in the parameter movement complexity (F=18.397, p < .000). Post-hoc comparisons revealed significant differences between text and all other modes (p < .000). Typing activity in the text mode determined the higher levels of activity here. A significant difference was also found for the comparison between audio and video. Although not reaching conventional levels of statistical significance there was a clear tendency for all avatar modes to induce more nonverbal activity than the audio mode. As this behavior was persistent over time, figure 8 only shows the average values for the whole observation time (9 minutes).

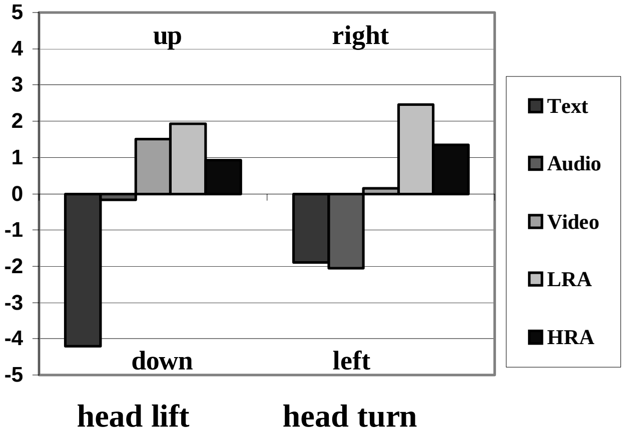


Figure 7 Sagittal and rotational head positions as deviations from the mean (right side indicates the direction of the communication window)

Discussion

The results of our study indicate that socio-emotional effects of the provided communication modalities are clearly reflected in the subjective verbal reports of the participants. General satisfaction with interaction outcome, the feeling of being co-present and the experience of emotional closeness as a relevant dimension of social presence as well as the affective component of interpersonal trust and the evaluative component of mutual person perception (liking) all seem to benefit from the provision of real-time audio or audio-visual channels for communication. However, post-hoc tests showed that significant ANOVA results were mainly due to the differences between the text mode and all the other modes, i.e. audio, video and avatar platforms did equally well in producing these desirable interpersonal results. This is notable, given that in earlier studies the most significant differences were found when comparing all types of computer-mediated communication to f2f interactions [13]. Now within CMC we find a clear distinction between text and all other modes [14]. Also, the behavioural data point to similar patterns of nonverbal communication and visual attention in the video and the avatar conditions revealing a certain loss of interest over time. This data challenges the common assumption that social presence and the related concepts are to be conceptualized as a continuum on which media characteristics and psychological effects of media use are co-aligned and on which social presence can be quantitatively tuned.

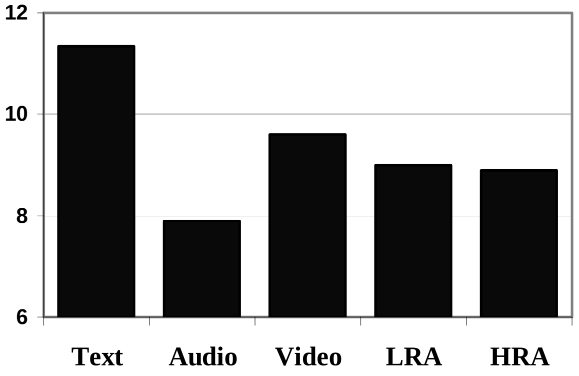


Figure 8 Nonverbal activity (movement complexity = number of movement dimensions involved in action) in five media conditions

Alternatively, qualitatively distinct processes could be posited which refer to different levels of social verification [15]. The first level would be the mere construction a social situation by entering any kind of interpersonal communication or by being addressed by another social entity. Second level verification would refer to the distinction between face to face or mediated communication. This distinction is most relevant with respect to any kind of mutual bodily impact (touch, interpersonal distance). In f2f situations actions and reactions cannot be temporarily or spatially buffered or decoupled. Responses have to take into account social adequacy and possible consequences. Third level verification is concerned with distinctions within mediated encounters. The special effects of text might be due to the fact that written speech is discrete (non-continuous) and is based on an arbitrary code. There is a natural turn taking structure induced by the syntax of the language. This is different for all the other modes where always at least one continuous and analogous channel is open simultaneously for the interlocutors, i.e. for audio, video and avatars. Similarities between these modes again could be due to the salience and predominant role of speech – as for example posited by Nass and Gong [16]. But while Nass and Gong [16] see speech generally and against an evolutionary background as the most decisive component, we hypothesize that under specific conditions - e.g. in case relational or socio-emotional information cues becomes more important – nonverbal aspects will gain additional salience and other than within the current study visual aspects of behavior will not suffer from a loss of interest over time. To test this hypothesis, subsequent experiments will imply systematic variation of the importance of socio-emotional aspects of interaction and mutual person perception.

All these modalities are still experienced as mediated, i.e. not constituting the experience of co-presence in the sense of expecting immediate bodily consequences. However, while audio and video are limited, avatar platforms offer new possibilities to overcome many of these restrictions. In the words of Foster and Meech [17] the challenge could be defined as to overcome the experience of “here and there” and to create a new shared experience of “elsewhere” (p. 212). Virtual worlds and avatars could thus be seen more as a means to contextualize social interaction and to foster the salience of nonverbal information, rather than just to provide high fidelity transmission channels for visual cues. They are in this sense not just virtual equivalents of a video conferencing system but a possibility for active filtering and contingency management systems. While current psychological research focuses mainly on the measurement of social presence, future research will have to address psychological as well as neurobiological knowledge concerning the specific demand characteristics of highly immersive virtual social realities as derived from human social information.

References

[1] S. Redfern, N. Naughton. Collaborative virtual environments to support communication and community in internet-based distance education. *Journal of Information Technology Education, 1(3), 201-211*. 2002.

[2] J. Steuer. Defining virtual reality: Dimensions determining telepresence. *Journal of Communication, 42 (4),* 73-93. 1992.

[3] E. Prasolova-Forland. Supporting awareness in education: overview and mechanism. *In Proceedings of the International Conference on Engineering Education.* August 18–21, Manchester, U.K. 2002.

[4] F. Biocca,, C. Harms, J. Burgoon. Criteria and scope conditions for a theory and measure of social presence. *PRESENCE 2001, 4th Annual International Workshop May 21-23,* Philadelphia, PA. 2001. URL: http://nimbus.ocis.temple.edu/~mlombard/P2001/papers.html

[5] G. Bente, S. Rüggenberg, N. C. Krämer. Social presence and interpersonal trust in avatar-based, collaborative net-communications. *7th Annual International Workshop Presence 2004*. Valencia: Editorial Universidad Politécnica de Valencia. 2004.

[6] F. Biocca, J. Burgoon, C. Harms, M. Stoner. Criteria and scope conditions for a theory and measure of social presence. *E. Lansing, MI: Media Interface and Network Design (M.I.N.D.) Lab.* 2001. URL: http://nimbus.ocis.temple.edu/~mlombard/P2001/Biocca1.pdf

[7] P. Kanawattanachai & Y. Yoo. The dynamic nature of trust in virtual teams. In *Journal of Strategic Information Systems, 11(3), 187-213*. 2002.

[8] Bollen, H.U. Hoppe, M. Milrad & N. Pinkwart. Collaborative Modelling in Group Learning Environments. In P. I. Davidsen, E. Mollona, V. G. Diker, R. S. Langer & J. I. Rowe (ed.), *Proceedings of the XX. International Conference of the System Dynamics Society*, Palermo (Italy). 2002.

[9] N. Kumar, I. Benbasat. Para-social presence: a re-conceptualization of „social presence“ to capture the relationship between a web site and her visitors. *In Proceedings of the 35th Annual Hawaii International Conference on System Sciences (HICSS-35.02), IEEE Computer Society*. URL: http://csdl. computer.org/comp/proceedings/hicss/2002/1435/01/14350013b.pdf . 2002.

[10] K. Nowak. Conceptualizing and measuring the sense of access to another mind. *Paper presented at the Presence 2001 Conference, Philadelphia, PA.* 2001. URL: http://astro.temple.edu/~lombard/P2001/Nowak1.pdf .

[11] C.-H. Tu. The impacts of text-based CMC on online social presence. *In The Journal of Interactive Online Learning, 1(2), 1-24*. 2002.

[12] A. Petersen, Interpersonale Kommunikation im Medienvergleich. Münster: Waxmann. 2001.

[13] A. Petersen, G. Bente, N.C. Krämer. Virtuelle Stellvertreter: Analyse avatar-vermittelter Kommunikationsprozesse. In G. Bente, N.C. Krämer, A. Petersen (eds.), *Virtuelle Realitäten, 227-253.* Göttingen: Hogrefe. 2002.

[14] S.G. Straus, J.E. McGrath. Does the medium matter? The interaction of task type and technology on group performance and member reactions. In *Journal of Applied Psychology* 79, *1, 87-97*. 1994.

[15] J. Blascovich, J. Loomis, A. Beall, K.R. Swinth, C. L. Hoyt, J. N. Bailenson. Immersive virtual environment technology as a methodological tool for social psychology. In *Psychological Inquiry, 13, 146-149*. 2001.

[16] C. Nass, L. Gong. (2000). Speech interfaces from an evolutionary perspective. *Communications of the ACM, 43 (9), 36-43*. 2000.

[17] D. Foster, J. Meech. The Social implications of virtual reality. In K. Carr (ed.) *Simulated and Virtual Realities: Elements of Perception*. Taylor and Francis. USA. 1

Non-verbal Communication for Correlational Characters

Marco Gillies and Mel Slater
Department of Computer Science, University College London, London WC1E 6BT, UK
{m.gillies, m.slater}@cs.ucl.ac.uk

Abstract

Social interaction is a key element of modern virtual environments. This paper discusses how non-verbal communication (or body language) is vital to real world social interaction, and how it is important to carry it over to virtual environments. It is not sufficient for a character to passively exhibit non-verbal communication; non-verbal communication should be a genuine interaction between a real and virtual person. To this aim the behaviour of the character should correlate realistically with that of the real person. We hypothesise that this sort of correlational non-verbal behaviour enhances presence and outline work in progress to investigate this hypothesis. We present a virtual character that exhibits this type of correlational behaviour in an immersive virtual environment.

1. Introduction

Perhaps the most interesting virtual environments for participants are social ones, where participants commonly share the VE, both with other real people, each represented by their own graphical character, or avatar, and with completely virtual people, that are entirely computer controlled. Since humans are social animals these other inhabitants of the virtual environment become a focus of interest, and VEs become a venue for social interaction. This means that such social interaction is a vitally important issue for presence research.

Though most social interaction among humans takes the form of conversation, there is a large sub-text to any interaction that is not captured by a literal transcription of the words that are said. Tone of voice can transform the meaning of a statement from angry, to sarcastic or playful. Posture can indicate keen engagement in the subject of discussion or bored disengagement, by leaning forward or slumping in a chair. Gestures can help clarify a path to be taken when giving directions. Facial expression can be smiling, and encouraging or indicate displeasure at what is being said. How close people stand to each other can indicate a lot about their relationship.

All of these factors go beyond the verbal aspects of speech and are called Non-Verbal Communication (often referred to by the popular term “body language”). Non-Verbal Communication (NVC) is a key element of human social interaction. Certain aspects of communication such as the expression of emotion or of attitude toward, and relationship, with other people are much more readily expressed non-verbally than verbally. Communication that lacks non-verbal elements can be limited and ambiguous, as demonstrated by the problems of interpreting the emotional tone of emails. In particular virtual characters that do not display NVC during conversation are less likely to be judged as realistic or to elicit presence.

However, it is not enough to display realistic postures, gestures, facial expressions etc, if these do not represent a genuine interaction with participants. In a recent review of the literature Sanchez-Vives and Slater[14] defined presence in a VE as successful replacement of real by virtually generated sensory data. Here ‘successful’ means that the participants respond to the sensory data as if it were real, where response is at every level from physiological through to cognitive. One element in this is the response of the environment to behaviours of the participant, and suggests that one of the most important factors in eliciting presence is form of interaction, particularly whole body, natural interaction. It is therefore important that social interaction occurs through natural bodily interaction, i.e. through NVC. This should be a true interaction, not merely a real and virtual human independently producing NVC.

Under what circumstances are people likely to find themselves responding to virtual characters as if they are real? Our hypothesis is that this would occur if the virtual characters respond to people *as if they are real*! Specifically what this means is that a kind of correlational dance is established in which actions of one person are refected in the actions of the other, which are reflected in the actions of the other, and so on. Moreover, people naturally attempt to find correlations between their own behaviour and that of their environment. This is particularly true of interaction with other people, people naturally interpret the behaviour of others in terms of their own actions and state. This occurs even when interaction with virtual characters whose behaviour is pre-recorded, and therefore is not related in any way[13] . This leads us to the Correlational Presence hypothesis, that presence is enhanced by producing this type of correlation between a person’s behaviour and that of the VE, and will therefore be enhanced if correlations are included as part of the environment. This work focuses on correlational presence during social interaction with virtual characters. This entails creating characters who not only autonomously produce behaviour, but behaviour, and in particular NVC, that is correlated realistically with full body behaviour of real participants.

Thus we come to the central hypothesis of this paper: correlational NVC is a key determining factor for presence during social interaction with virtual characters, or mediated via avatars. The remainder of this paper describes current work in progress to test this hypothesis, and in particular to create characters that display correlational NVC. Our characters have been created to run in a Cave-like immersive virtual reality system[4] , which allows natural interaction with a life-size virtual character.

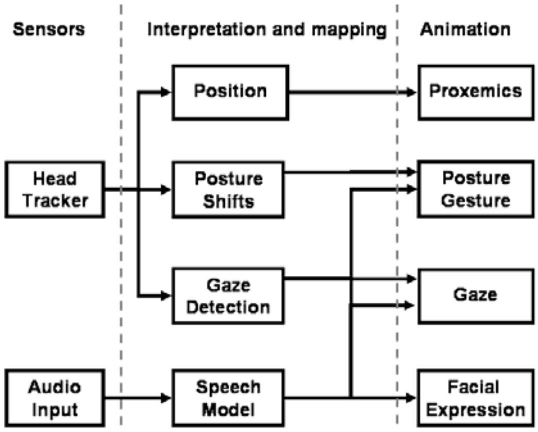


Figure 1: Mapping between sensor data and animation

Figure 2: An example character

Virtual characters require three basic elements in order to display NVC that correlates with a participant, as shown in figure 1. The first is an animation system that is able to generate realistic non-verbal behaviour, this is described in section 3. The character must also be able to sense the behaviour of the user. In our current system we have chosen to use the sensors commonly available in immersive virtual reality systems, particularly Cave-like systems. Thus we have restricted ourselves to a single head tracker, and to audio input via a microphone. In the future it would be interesting to look into more complex tracking systems, but this would reduce the general applicability of this work. Mediating between these two elements is a module that interprets the sensor data and maps the results to behaviour. The sensing, interpretation and mapping aspects of this work are described in section 4.

2. Non-verbal Communication

As described in the introduction non-verbal communication takes many forms, or modalities. Argyle[1] lists the following modalities of NVC: “facial expression; gaze (and pupil dilation); gestures, and other bodily movements; posture; bodily contact; spatial behaviour; clothes, and other aspects of appearance; non-verbal vocalizations, and smell”. This work is restricted to modalities that involve bodily movements, avoiding non-bodily modalities such as vocalizations or smell, and static modalities such as appearance or clothing. We therefore use five main modalities: posture, gestures, facial expression, gaze and proxemics (spatial behaviour, personal space).

Our work on correlational NVC builds on a large body of work on animating NVC, for example Cassell *et al.* [3] , Guye-Vullième *et al.*[8] and Pelachaud and Bilvi[1] . We use the Demeanour framework[6] [7] to generate animated non-verbal behaviour. Demeanour consists of a number of animation modules that display the behaviour (described below), and a declarative behaviour language for specifying rules for what behaviour should be displayed. The behaviour language is used to specify mappings from input variables to output behaviour. The input variables come from sensing the user, and other contextual factors and described in section 4.The general aim of the behaviour generated is to give a generally favorable and friendly impression of our character (shown in figure 2). Thus most of the behaviour will display a generally friendly attitude towards the participant. The rest of this section will describe the modalities we use.

2.1 Posture and Gesture

Posture is the long-lasting static pose of the body whereas gestures are more transitory movements, mostly of the arms and head that commonly accompany speech. While people always have a posture, gestures are a purely conversational phenomenon, and seem intimately connected with speech, people gesture while talking on the telephone even though no one can see them.

Though posture and gesture are distinct communicative phenomena they use the same body parts, and as such there is a single animation module for both. Postures and gestures and generated from a set of basis poses (which are static) and animations (which are body movements). New postures or gestures are generated by a weighted interpolation over these bases. In order to vary the postures or gestures in response to the participant’s behaviour while maintaining a large variety of behaviour, we group the bases into different types. Different types of behaviour are generated depending on the participant’s behaviour, but each type can exhibit a variety of different behaviour by choosing different interpolation weights for the members of that type.

2.2 Facial Expression

The facial animation module is based on morph targets. The face is represented as a mesh, and each facial expression is represented as a set of displacements from this mesh (a morph target). The face is animated by giving weights to the morph targets. The displacements of each morph target are scaled by its weight and added to the face mesh, generating a new facial expression. The facial animation module works in the same way as the body animation module, having a number of bases which are interpolated to produce new animations. The bases can either be static facial expressions (morph targets, for example a smile) or facial animations (time varying weights over the morph targets, for example open and closing the mouth for speech). As with body motions the facial bases are grouped by type. Facial expression is not currently used to react to the behaviour of the participant, we always use a friendly smiling expression (see figure 2). Facial expression is also used to represent speech and blinking.

2.3 Gaze

The gaze animation module determines where the character is looking. At any given time the character is looking at a single gaze target, which might be the participant, an object in the environment or a location. The character moves its eyes, head and body to look at the target. It looks at the target for a set duration and after the end of that duration a new target is determined based on rules as described in section 4.

2.4 Proxemics

Proxemics are spatial relationships between people. People tend to maintain a comfortable distance between themselves. This distance depends on a number of factors such as culture and the relationship between the people. The proxemics animation module maintains this comfortable distance. If the distance between the character and participant is too large the character steps towards the participant and vice versa. The distance itself can be varied to make it a comfortable distance for the participant, or an uncomfortably distance (too close, for example) in order to elicit a behavioural response from the participant.

3. Interaction

For truly correlational behaviour the character must be able to detect the behaviour of a real person in order to react to it. The work is targeted at standard Cave-like systems and other similar immersive systems. As such, participant sensing is limited to the types of sensor that are normally available on this type of system. In fact, we only use two sensors, a 3-degrees-of-freedom head tracker (InterSense IS900) and audio input from a standard radio microphone. We attempt to extract enough information from these limited sensors to give a strong sense of correlation. The use of these limited sensors has the obvious advantage that they are relatively cheap but also that they are less intrusive and bulky than full body tracking. It is important to avoid overly intrusive trackers as they can be uncomfortable for the user and reduce the naturalness of their behaviour. This is particularly true of the subtle behaviours that make up non-verbal communication. The rest of this section describes how the sensor information is mapped to the character's behaviour, figure 1 gives an overview of this process.

3.1 Head Position

The most basic information that can be obtained from the head tracker is the current position of the participant. This is used by the proxemics module to determine the current distance of the character to the participant. In order maintain a comfortable distance as described in section 2.4. The head position is also used by the gaze module to enable the character to look appropriately at the participant.

3.2 Interactional synchrony

It is also possible to obtain more complex information from the head tracker. Kendon [10] has shown that when people engage in conversation and have a certain rapport, their behaviour will tend to become synchronised, an effect he calls 'interactional synchrony'. This is particularly true of a listener synchronizing their behaviour with a speaker. This can take many forms, two of which we simulate. The first is that a listener will tend to move or shift posture at the same type as the speaker (but not necessarily have the same posture). This can be implemented very simply using a single head tracker. We detect the participant's posture shift when the tracker moves above a threshold. When a shift is detected the character will also perform a shift. The other form of interactional synchrony noted by Kendon that we simulate is a listener synchronizing their movements with important moment in the speaker's speech. As we detect when the participant is speaking (see section 3.4) it is possible can detect the start and end of their speech. The character performs a posture shift at these two important moments in the conversation.

3.3 Head orientation

The head tracker also gives the orientation of the head. This can give an approximate direction of gaze for the participant. This is used to implement gaze following. A powerful cue for social understanding is that a one person will look in the same direction as another[9] . This displays shared attention, that they both share an interest in a common object, and they both understand that the object is important to the other and to the conversation. Thus a character that follows the gaze of the participant gives a powerful cue that they are understanding the participant's conversation and that they empathise, to some degree, with the participant. This only works when the participant is looking at something relevant, so the character cannot follow the participant's gaze arbitrarily. Otherwise the character will appear to be constantly looking at irrelevant objects, and seem stupid. To avoid this problem certain objects in the environment and defined to be *salient objects*, when the participant appears to be looking at one of these the character will follow gaze, but not otherwise.

3.4 Speech

As this work deals mostly with social behaviour a good model of speech and conversation, is needed. This model depends on a conversational state, which can have one of three states: *character talking*, *participant talking* and *neither*. The character's own conversation is handled in a wizard-of-oz manner, a number of audio clips, can be triggered by a confederate. It is thus trivial to know if the character is talking. The participant has a radio microphone

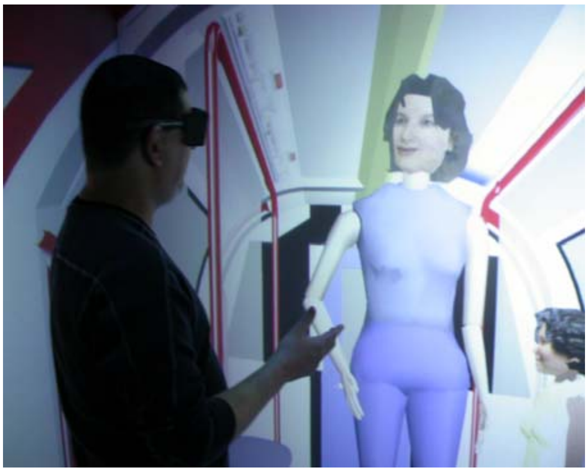


Figure 3: Social Interaction between a real and virtual human

which is used to detect when they are talking (simply based on a threshold for the amplitude of the signal). The behaviour associated with speech is consists in gesture, gaze and posture shifts (describe in section 3.2).

Gesture behaviour is intimately connected with speech. There are two basic types of gesture, normal gestures that accompany speech, and "back channel" gestures that occur when listening, and aim to encourage the talker. Normal gestures are modeled based on a number of basis gestures as described in section 3.1, and only occur in the *character talking* state. The characters mouth is also animated during the *character talking* state to show that they are talking. The most common back channel gestures in western culture are head nodding to show agreement and encouragement, and shaking the head to show disagreement. As the character's behaviour is designed to be favorable towards the participant, only head nodding is shown.

The character's gaze is driven, based on speech, by a model by Garau *et al.*[5] Vinayagamoorthy *et al.*[15] and Lee, Badler and Badler[11] , which are ultimately based on the work of Argyle and Cook[2] . In this model the character looks either at their conversational partner (the participant) or at other locations in the world. The length of any look is determined at random based on mean lengths determined from data from observation of conversations. The mean length of looking at the participant is greater when listening than when talking (as is consistent with numerous Argyle's observations of conversations). When the character is not looking at the participant then the locations chosen are determined based on statistics by Lee, Badler, and Badler.

5. Conclusions

This paper has described work in progress in developing correlational non-verbal behaviour in virtual characters. The aim of this work is to enhance presence in social interactions with virtual characters by simulating a key element of real human social interactions. We are currently planning a study to test the effects of this work. The study will involve the subjects holding a conversation with a character controlled by the behaviour model described, compared with a character that exhibits the same beahviour but without it being correlated to the behaviour of the user. . The scenario chosen is one of a London Underground train, with the character being a tourist asking directions (the environment and character are shown in figures 2 and 3).

Acknowledgements

This work is supported by the Presencia EU FET project.

References

[1] Argyle, M. Bodily Communication. Routledge. 1975

[2] Argyle, M., and Cook, M. 1976. Gaze and Mutual Gaze. Cambridge University Press,

[3] Cassell J, Bickmore T, Campbell L, Chang K, Vilhjàlmsson H, Yan H Embodiment in Conversational Interfaces: Rea. *Proceedings ACM SIGCHI*, 520-527: ACM Press. 1999

[4] Cruz-Neira, C., Sandin, D.J., DeFanti, T.A. Surround-Screen Projection-Based Virtual Reality: The Design and Implementation of the CAVE. *ACM Computer Graphics*, 27/2 , 135-142. 1993.

[5] Garau, M., Slater,M., Bee, S. and Sasse, M.A. (2001). The impact of eye gaze on communication using humanoid avatars. *Proceedings SIG-CHI conference on Human factors in computing systems*, 309-316. 2001

[6] Gillies M, Ballin D. Integrating autonomous behaviour and user control for believable agents. *Proceedings Third international joint conference on autonomous agents and mult-agent systems.* New York USA. 2004

[7] Gillies M, Crabtree B, Ballin D Expressive characters and a text chat interface. *Proceeings AISB workshop on Language, Speech and Gesture for Expressive Characters* (Olivier P, Aylett R, eds). University of Leeds, UK 2004

[8] Guye-Vuillème A, T.K.Capin, I.S.Pandzic, Magnenat-Thalmann N, D.Thalmann Non-verbal Communication Interface for Collaborative Virtual Environments. *The Virtual Reality Journal* 4, 49-59. 199

[9] Johnson, S.C. Detecting agents. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* 358, 549-559. 2003

[10] Kendon, A. Movement coordination in social interaction: some examples described. *Acta Psychologic*. 32, 100-125. 1970

[11] Lee S. P., Badler, J. B. and Badler, N. I. Eyes Alive, *ACM Transactions on Graphics, (Proceedings of ACM SIGGRAPH 2002)* 21/3, 637-644, July 2002

[12] Pelachaud C, Bilvi M. Modelling gaze behavior for conversational agents. *Proceedings Intelligent Virtual Agents*, pp 93-100. 2003

[13] Pertaub DP, Slater M, Barker C. An experiment on public speaking anxiety in response to three different types of virtual audience. *Presence-Teleoperators and Virtual Environments* 11, 68-78. 2002

[14] Sanchez-Vives MV, Slater M. From Presence to Consciousness Through Virtual Reality. *Nature Reviews Neuroscience* 6, 8-16. 2005

[15] Vinayagamoorthy V, Garau M, Steed A, Slater M An eye gaze model for dyadic interaction in an immersive virtual environment: Practice and experience. *Computer Graphics Forum* 23, 1-11. 2004

Issues of Law and Ethics in the Design and Use of Virtual Environments

Woodrow Barfield PhD, JD, LLM

Abstract

Recent advances in the technologies associated with wearable computing, virtual environments, and computer-mediated realities has led to interesting legal, policy, and ethical issues and concerns. We are now at the stage in technological development where we can begin to build on-line virtual or computer-mediated communities, and where people can spend significant amounts of time. In these worlds it is pertinent to ask, should there be any rules, laws, policies, or ethics to govern human interactions? For example, what will happen when the real and virtual merge and become indistinguishable, or when cyberspace spills out into a computer-mediated reality? What will be the role, if any, of the government in regulating and setting policy for conduct in virtual communities or computer-mediated spaces, and what will be the role for the Courts in interpreting and carrying out the law? Will the traditional roles and functions performed by the government and courts in real-world environments, by analogy be transferred to virtual communities? Can cyberlaw reach outside someone's electric eyeglasses and into the real world that he or she is computationally mediating? Should virtual space be treated like real space? Should real space be treated like virtual space? Should existing property, contract, tort, and criminal law wash across the dissolved boundaries of computer-mediated perception? And should one even be able to own a piece of virtual space? The above set of questions are timely and interesting given current technological developments in virtual and computer mediated reality displays, and law scholars have already begun the discussion of whether interactions in virtual space should be governed in a similar manner as real space, but the results to date are inconclusive. However, since governmental bodies are already prominent in creating statutes to govern on-line electronic commerce, it may not be much of a leap to assume that such bodies may also see it within their province to begin the active process of codifying and creating policy for the full range of human activities occurring in virtual environments. This talk will review basic issues of law for virtual environments and will attempt to serve as a warning bell for the "presence" community that now is the time for the community to get involved in setting the policy that will guide interactions in virtual environments.

Session 4

22nd September, 2005

Meta Presence

10.00-10.30 *Sharing and Analysing Presence Experiments Data*
Doron Friedman[1], Andrea Brogni[1], Angus Antley[1], Christoph Guger[2] and Mel Slater[1]
[1] Department of Compute Science, University College London, UK
[2] g.tec - Guger Technologies OEG, Austria.

10.30-10.45 *The Big Picture: Gaining Perspective by Examining Patterns in the (Tele)Presence Literature*
Matthew Lombard and Matthew T. Jones
Department of Broadcasting, Telecommunications and Mass Media, Temple University, USA

10.45-11.00 *Are we persuaded by feeling "part of the action"? Exploring the similarities between the transportation imagery model and presence*
Cheryl Campanella Bracken
Department of Communication, Cleveland State University, USA

Sharing and Analysing Presence Experiments Data

Doron Friedman[1], Andrea Brogni[1], Angus Antley[1], Christoph Guger[2] and Mel Slater[1]
[1] Department of Compute Science, University College London, UK
[2] g.tec - Guger Technologies OEG, Austria.

Abstract

Presence research heavily relies on empirical experiments involving subjects in mediated environments. Such experiments can be extremely resource intensive and produce very large amounts of data. As the presence community matures, we would like to suggest that data collected in experiments will be publicly available to the community. This will allow the verification of experimental results, comparing results of experiments carried out in different laboratories, and evaluating new data-analysis methods. In this paper we present the complete data set from a large-scale experiment that we have carried out in highly-immersive virtual reality. We describe the data we have gathered and give examples of the types of analysis that can be made based on that data.

Keywords--- **methodology, presence, virtual reality, physiology, GSR, ECG.**

1. Introduction

Engineering practice places emphasis on reusable components and standardization. In Presence research technical standardization and reusable components, such as reusable virtual environments (VE) would be useful, but moreover, we can strive towards standardization in data collection and data analysis.

This will serve the following goals:

- Researchers can test new methods of analysis without carrying out time-consuming experiments.
- Researchers who have experiments with huge amounts of data can rely on the community to assist in the analysis.
- This will ensure high quality of experiments and publications.
- This will promote progress in presence methodology, by allowing a comparison of methodologies and research techniques.
- It would make it possible to compare results among different experiments, and even among experiments carried out in different labs.

In this paper we present the complete data for an experiment we have carried out in a Cave-like system[1]. The experiment was large scale in that it included several types of measurements, including both quantitative data such as physiological measurements and qualitative data such as semi-structured interviews. The hypothesis and the results for this specific experiment are not of interest here. We describe the data itself and point to the analysis techniques we have used to analyze it. We suggest other techniques to analyze the same data, and some other types of data that may be used in the future.

The IPQ group have already published data from presence experiments and encouraged other researchers to use it[2]. However, this only includes data from the IPQ questionnaire; we encourage publishing and sharing all types of data.

Analyzing data is, of course, not unique to presence research. Research on evaluating VR usability in general may be relevant here (e.g., [2]). The unique characteristics of our research, as well as the research of many others in the presence community, is that we have different types of data generated by technical devices and computers, which are mostly detailed and accurate, and ultimately, their integration can allow us to “reconstruct” the subjective experience of the subject. One area which is similar in many aspects, and in which such reconstruction of experience is considered critical, is post-mission debriefing of air-force pilots (e.g., see [18]).

The data for the experiment, as described in this paper, can be downloaded from: http://www.presencedata.info/. The data is organized in online tables. In the paper below we refer to the data by mentioning the table number; these can be accessed from the main URL above.

2. The Experiment

As mentioned earlier, in this paper we are not concerned with the specific hypothesis or with the results of the experiment; these are described elsewhere [12, 14, 25]. Rather, we describe the types of data collected and the ways to analyze them. In this section we provide an overview of the experiment goals and procedure.

The overarching goal of the experiment was to investigate presence as a multi-level construct ranging from lower-level involuntary responses to higher-level subjective responses. Specifically, this experiment was designed to find physiological correlates to breaks in presence (BIPs) [3, 23, 24, 26]. However, note that the techniques mentioned in this paper should be appropriate for most presence experiments.

Upon arrival, participants were given an instruction sheet describing the experimental procedure and the possible risks associated with using virtual reality equipment (including simulator sickness). They were asked

---

[1] CAVE™ is a trademark of the University of Illinois at Chicago. In this paper we use the term ‘Cave’ to describe the generic technology as described in [4], rather than to the specific commercial product.

[2] http://www.igroup.org/pq/ipq/data.php

to fill out a consent form and a pre-questionnaire covering their age, gender, occupation, and previous experience with VEs and computer games.

They were then led though to the Cave, where they were shown how to connect the electrocardiogram (ECG) and respiration sensors. Galvanic skin response (GSR) sensors were attached to their non-dominant hand, and they were asked to stand still in the Cave for a baseline reading. During this time, no images were displayed on the Cave walls.

Next, participants were asked to complete a brief exercise in a virtual "training" room designed to make them comfortable moving around the Cave. Once they felt comfortable, they were told that in a few moments they would find themselves in a bar, where they were asked to spend a few minutes until we told them it was time to come out. It was explained that they were free to explore the bar as they wished, and that afterwards we would be asking them questions about the experience. They remained in the virtual bar for the duration of two songs, approximately five minutes. Note that the virtual space of the bar was not much larger than the physical space of the Cave; this means subjects moved around the bar by walking rather than navigating with a wand.

The bar contained five virtual characters: one barman, one couple standing near the bar on the right, and another couple seated on the left of the room. When approached by the participant, the characters would utter phrases suggesting that a celebrity was about to arrive.

At four points during the experience, the walls of the Cave were blanked out, leaving participants in a completely white room for approximately 2 seconds. Two experimental minders observed them throughout, noting their bodily and verbal responses to the whiteouts. Participants' autonomic responses were also monitored throughout. Figure 1 shows a participant in the bar environment, wearing the physiological monitoring equipment.

Figure 1: Participant wearing the bio-sensors and VR goggles in the Cave.

The experiment included two conditions. The main condition included 20 subjects and the goal was to try to detect, for these subjects, whether there is a physiological "signature" to the BIPs. A second condition included 10 other subjects. They went through the same procedure as described above, but in addition they were given an explanation about BIPs, or "transitions to real". They were trained to click a wireless mouse whenever they experienced a transition from the virtual world to the physical reality of the laboratory. During the experiment phase, they were asked to click the mouse whenever they had such a transition. The goal was to find out if it was possible to detect a physiological signature to these self-reported BIPs.

Immediately after the experience, and before taking off the equipment or leaving the Cave, participants were asked to answer two questions concerning their immediate impressions regarding their overall sense of "being in" and "responding to" the bar.

Next, they were shown the video of themselves in the bar, and were asked to comment on anything that they remembered while watching the video. A semi-structured interview was conducted afterwards.

The experiments were carried out in a four-sided Cave-like system [4], which is driven by an Onyx IR2 with 4 graphics pipes. Subjects were wearing Intersense IS900 wireless trackers. The application was written on top of Dive [6, 27]. Physiological signals were measured using ProComp Infiniti by Thought Technology Ltd.

3. The Data

Analysis methods are typically classified into quantitative and qualitative methods. We do not undermine this distinction, but in this paper we find it useful to make another distinction: between data that is temporal and data that is not.

3.1. Temporal Data

It has been argued that rather than being a stable constant throughout the mediated experience, presence may vary over time [1, 16, 20]. Generally, we would like to be able to measure how presence varies over the duration of the experience, and how it is affected by specific events in the environment. Specifically, we encourage studying presence by looking at physiological data. One of the first studies to show that presence can be studied as an objective, measurable response, based on GSR and heart rate, was carried out by Meehan et al. [19].

Ideally, all data could be placed on the same timeline, and visualized together. In this section we present these types of data independently, and in later sections we discuss possible ways to cross-analyze them.

Most of the temporal data are generated digitally; the main challenge is synchronization. Accurate synchronization is critical for event-related responses, such as discussed in section 4.2. In our lab, we use the Virtual Reality Peripheral Network (VRPN)[3] to synchronize among the data and the VR system. VRPN is an open software platform that is designed to implement a network-transparent interface between application programs and the

---

[3] http://www.cs.unc.edu/Research/vrpn/index.html

set of physical devices (trackers, etc.) used in a VR system. VRPN was recently extended to support the ProComp physiological recording device. Using VRPN, all data generated during an experiment can be synchronized, sent over a network, and stored with uniform timestamps for later analysis.

3.1.1. VE Events: In VR, and in fact in any type of digital media experience, it is possible to keep accurate logging of most meaningful events that take place during a session.

First, events and actions carried out by the system can be easily logged by the application. In our experiment we recorded all the instances in which the virtual characters spoke. The data is included in Online Tables 1A and 1B.

Second, events carried out by the participant typically involve some type of interaction device; such events are easily tracked as well. In our experiment, we have allowed, in one condition, for subjects to indicate breaks-in-presence, using a wireless mouse device. The data is included in Online Table 1C. Typically, VEs would allow interactions of subjects with the VE; such events would similarly be tracked and logged. We did not record when participants speak, but this could be done in principle.

The analysis of this type of data is typically useful for detecting event-related responses. For example, one can look at the physiological state of the participant whenever something happened in the VE; examples are given in section 4.2.

3.1.2. Tracker data: In VR the participants are, typically, head tracked. This provides extremely useful information about their position and head direction at any moment. While theoretically they can be looking sideways, we expect this gives us a good approximation of what they were looking at, without the need to perform eye tracking, which is difficult in a Cave environment.

The tracker data from our experiment is included in Online Table 2. Examples of this analysis are spatial analysis (see section 4.3) and event-related analysis (see section 4.2). We did not use head-tracker data to reconstruct what the subjects were looking at; this should be possible since the trackers include orientation information as well as position.

3.1.3. Galvanic Skin Response: GSR, also sometimes called galvanic skin conductivity or Electro Dermal Activity (EDA), is measured by passing a small current through a pair of electrodes placed on the surface of the skin and measuring the conductivity level. In our experiment GSR was sampled at 32 Hz, and the signal was obtained from electrodes on two fingers.

The GSR data for the two experimental conditions appears in appendices 3A and 3D. More details about GSR, and about analyzing GSR data from this experiment, can be found in Slater et al. [25]. Specifically, they show that the GSR parameters predict the occurrence of breaks in presence, using a method based on continuous wavelet transforms of the GSR signal.

3.1.4. ECG: Several parameters can be extracted from ECG recordings. In addition to the obvious one – heart rate – the heart-rate variability (HRV) can be used to describe the physiological behavior of the participant, and an event-related heart-rate response may be useful to study the reaction of the subject to an event (such as a BIPs).

The ECG data for the two experimental conditions is provided in appendices 3B and 3E. The sampling rate is 256Hz. Slater et al. [25] and Guger et al. [14] provide an analysis of the ECG, including a comparison of the training and experimental phases, comparison of social phobic and non-social phobic participants, and event-related ECG.

3.1.5. Respiration: The respiration signal measures the inhalation and exhalation phases of the human subject. The signal can be used to extract the deepness and frequency of the respiration. The first step is to low-pass filter the signal with 10 Hz to remove noise components and movement artifacts. Then each zero crossing of the bipolar respiration signal is detected in order to calculate the frequency. Event-related respiration changes around a BIP can be investigated. For example, it is possible to detect a change in deepness and frequency after the BIP. It is very common that the subjects hold their breath for a few seconds when the BIP occurs.

The respiration also modulates the ECG signal with a frequency of about 0.1-0.2 Hz. This modulation effect must be considered when the ECG is analyzed; details can be found in Florian et al. [5].

The respiration data for the two experimental conditions appear in appendices 3C and 3F.

3.1.6. Video: The whole experiment session was videotaped for all subjects. In our experiment we used a Cave system where the projection takes place on three walls and on the floor. The camera was placed outside the Cave so that it captures the whole area of the Cave. This is useful to observe the subject’s motion throughout the physical space of the Cave, and also allows analyzing their main body gestures and postures. However, the subject is typically shown from the back. Generally, it would be difficult to pick up the subject’s facial expressions, given the relative darkness in the Cave and the fact that subjects wear VR goggles. We still recommend placing another camera that picks up the subject from the front; e.g., in our Cave setting, we could eventually place one on the top of the front Cave screen.

The video can be used for testing hypotheses, for providing the experience to researchers who were not present in the experiment, and for later analysis of body language. A sample of the videos can be found in Online Table 4 and a copy will be sent upon request.

3.1.7. Video interview: Following the experiment, the subject watched the video together with the experimenter and reflected on his or her experience. We have used this video interview to gain some insights for later exploration during the post-experiment interview. Ideally, this interview by itself should be recorded and provided with the data, because it provides a potentially insightful glance into the subject's experience when it is still fresh, and in a way that allows the subjective impressions to be temporally aligned with the experience.

3.1.8. Additional measurements: In the future we hope to explore additional types of temporal data. Some experiments involve conversation, either among multiple subjects in a multi-user experiment or between a subject and a confederate. Recording such conversation and synchronizing it with the other types of data can be extremely useful. This is a specific rich type of VE events, as discussed in Section 3.1.1.

Other types of physiological data can also be used. Our system now includes electroencephalograph (EEG) measurement as well; this was used for a brain-computer interface [10], but may also be used for post-experiment analysis. Similarly, it should be possible to analyze muscle activation in the form of electromyogram (EMG) recording. Such measurements are extremely useful for measuring emotional states[4], and may be especially useful in VR where the subject's face is obstructed by the VR goggles, which undermines video-based facial expression analysis.

It is also possible to track body parts in addition to the head. Most VR systems include a wand device that is tracked. The VE of the bar room was such that no navigation using wand was required; walking in the Cave was enough. Thus, hand-tracking data for this experiment is not available. For some experiments it may be useful to include the wand, even as a simple tracking device; it may be possible to partially analyze hand and arm gestures. Naturally, full-body tracking is highly useful for experiments that may involve body language and non-verbal communication. If such tracking devices are not available, it is still possible to utilize experts in body language who can observe the subjects and interpret their behavior; this can be done after the experiment by watching the video[5]. Freeman et al. studied postural shifts in response to motion stimuli [8], as an indication of presence.

In addition to documenting the experiment sessions by video, it would be useful to be able to record the virtual environment. Such recording of interactive environments, although not a new idea, is still not straightforward and is not provided by any of the standard VR toolkits.

There are a few systems that allow users of VR/VE systems to review sessions [13, 15, 17], and Steed et al. [26] actually used such a system in their experiments. They describe a system that records the full Dive session and allows the experimenter to play it back within Dive and experience it as a first person view. It is also possible to use intelligent tools that create movie summaries from interaction sessions. Such tools may allow one to view the interaction from various angles, and to focus on specific events within a session [9].

3.2. Non-Temporal Data

In this section we discuss data that is collected after the experiment, and thus cannot be temporally aligned with data collected during the experiment.

3.2.1. Questionnaires: It has been pointed out several times that questionnaires are problematic in the context of measuring presence: for example, they are unstable, in the sense of being very sensitive to prior experience [7], they may not be able to distinguish reality from virtual reality [29], and they can shed no light on whether 'presence' actually exists as a uniquely identifiable brain activity during the course of the experience to which it is meant to relate [22]. Questionnaires may be made more useful and reliable if their results are integrated with qualitative results and with physiological data, such as suggested in this paper.

Even if the community does not converge on one questionnaire, experimenters could let the subjects fill in more than one questionnaire, thus allowing cross-community comparisons.

In addition to presence questionnaires, we suggest administrating psychological tests, such as personality tests. Again, these tests could be controversial independently, but could have valuable contribution when crossed with other data. As an example, Slater et al. [25] found a correlation between a test for social phobia and ECG. The results for the psychological test that elicits the degree of social phobia [30] are given in Online Table 5C.

In the bar experiment only the subjects in the second condition filled in presence questionnaires; the questionnaire appears in Online Table 5B and the results are available in Online Table 5C. We are now refining a methodology of evaluating and measuring presence based on a combination of questionnaires, interviews, and physiological responses; thus, in the future, we do plan to include presence questionnaire data with our experiments.

Each participant also completed a questionnaire prior to their immersion that gathered basic demographic information and other background information regarding their use of computer games. The questionnaire is included in Online Table 5A, and the results in Online Table 5C.

3.2.2 Immediate question: Immediately after the experience, and before taking off the equipment or leaving the Cave, participants were asked to answer two questions concerning their immediate impressions regarding their overall sense of "being in" and "responding to" the bar.

The purpose of these two questions was to capture participants' immediate subjective response to the experience in a way that was as far as possible unclouded by post-hoc rationalizations. Afterwards, they were able to expand on their answers in the semi-structured interviews.

The responses to the immediate questions are given in Online Table 6C.

---

[4] Hugo Critchley, UCL Institute of Neurology, private communication.

[5] Note that we ensure that the subjects are completely separated from the surrounding lab by covering the Cave with curtains. This means the experimenters can only observe the subjects during the experiment by watching the video feed from the camera. Thus, there is in practice no difference between analyzing the experiment (video) during the experiment or afterwards.

3.2.3 Presence graph: During the interview, participants were asked to draw a graph describing the extent to which they felt they were in the bar versus being in the laboratory throughout the experience. A sample graph is shown in Figure 2.

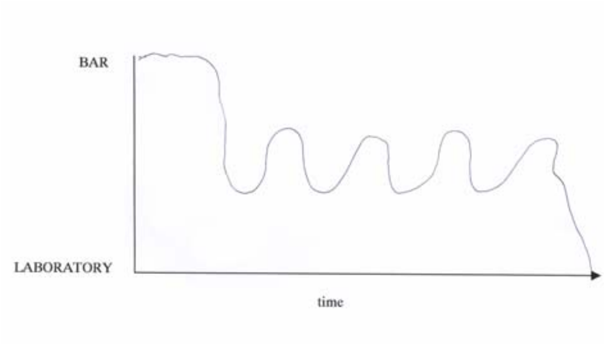


Figure 2: Presence graph illustrating BIPs (P8 female)

.

While the diagrams provide some temporal information, they cannot be aligned precisely with the temporal data, and thus are not considered here to be temporal. Ideally, they could serve as a link between the interview and the temporal data. For example, while drawing the diagram, the subject can point to certain extreme points in their presence function and describe how they relate to their interview.

The presence graphs were abstracted and classified into four types [12]; the data is provided in Online Table 6C.

3.2.4 Post-experiment interview: this is often very useful in providing hypotheses for future research. Such interviews typically contain a lot of fascinating insights, which often gets lost because, due to their subjective nature, they are difficult to analyze in a rigorous manner. Garau [11] and Thomsen [28] used Grounded Theory extensively for analyzing interviews in the context of presence research. By including the interview transcriptions with the data we hope other researchers can get an insight into the subject's experience, and perhaps suggest methods of analyzing this data in a systematic way.

In our experiment each interview was conducted using a semi-structured interview agenda, to ensure that it did not stray from the research questions in which we were interested. The interviews were audio taped and then transcribed verbatim. In the future, it may be useful to videotape the interviews, for post analysis of the interview itself. Garau et al. [12] discuss the interview techniques and the results obtained for this experiment. The transcriptions for the two experimental conditions are included in appendices 6A and 6B.

4. Compound Analysis

In the previous sections we described the individual data types and the analysis we have carried out with them. In this section we describe analysis of two or more elements together. Again, the intention here is to explain what types of analysis are possible, rather than to focus on specific results from this experiment.

4.1 Event-Related Analysis

It is of great interest to see if we can detect measurable responses to events in the experience. Our experiment was specifically designed to find out if we can detect a physiological "signature" to BIPs; such evidence was found using a wavelet transform of GSR parameters as reported in Slater et al. [25]. Also, using the same analysis technique, they have found a significant physiological response to events in which the virtual characters spoke to the subject.

One possibility that we are examining is whether the stiffening stabilizing reaction that subjects have to a sudden change in their visual field is detectable as a loss in height that can be seen in the head tracker data. In the future we could combine this with EMG data from the Soleus muscle in the lower leg to detect when a subject is experiencing a BIP.

An example of one subject's height following a BIP appears in Figure 3. The subject pulls down nearly two centimetres after the BIP. Due to high variance in peoples' standing height our results are so far inconclusive. However, given that many emotions, such as stress, are manifested as muscular tension in the body, looking at the results of this muscular activity whether through EMG or postural change is a promising method to analyse response to virtual reality.

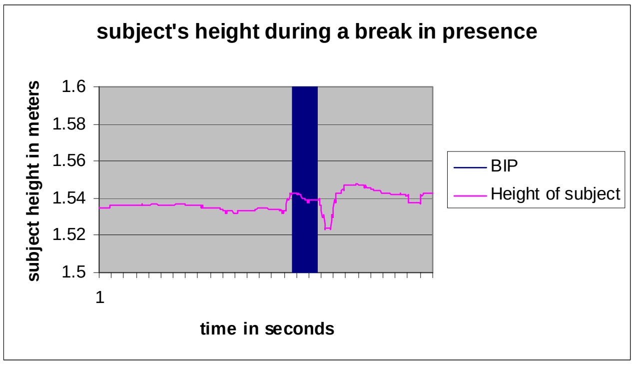


Figure 3: A graph of one subject's head height, as measured by the head tracker, following a BIP event.

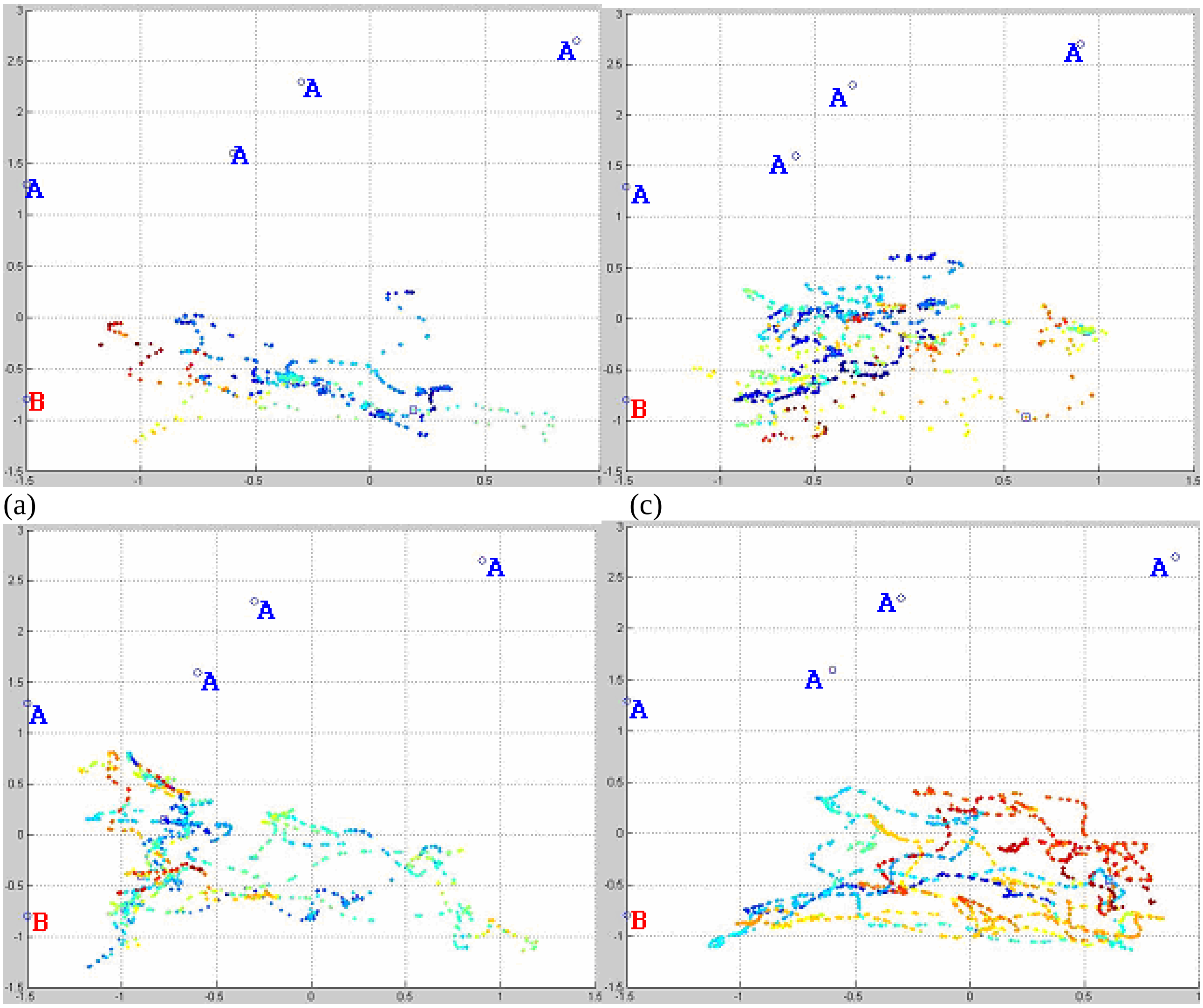
4.2 Spatial Analysis

A spatial representation of time-variant signals is a very useful tool for the experimenter. A quick glance may allow detection of areas of the VE where the signal has extreme values, and this may provide clues for further analysis.

Specifically, an interesting approach in the data analysis is linking the physiological values with the position of the subject in the virtual space at the same time. The resulting graph shows how the signal spatially changes over the VE, and it can be useful to detect whether there is a difference in the way different areas affect the subjects' physiology.

While such plots may not qualify as conclusive evidence by themselves, they could be a useful starting point for further analysis of physiological responses and proximity to virtual characters. Figure 4 illustrates this point for a few subjects in the bar experiment: Figures 4a and 4b show data for subjects who had stronger GSR values next to the barman, whereas figures 4c and 4d show subjects with the opposite trend. Of course, many subjects did not show a clear pattern at all

This technique is typically more useful when the VE is large, and the exploration of the VE is of interest in itself. In the case of the bar experiment, the room was spatially limited by the Cave's walls, and movement was restricted. In this case we provide an example of the response to the virtual characters, but it is also possible, of course, to study the response in the vicinity of virtual objects.



(a) (c)

(b) (d)

Figure 4: Plots of GSR levels in space from four different subjects. The positions of the five virtual characters are given by letters: the barman is denoted by the letter B and the other characters by the letter A. GSR values are denoted by more red colors, and lowest GSR values are towards blue.

Interestingly, the qualitative analysis of the interviews seemed to reveal that the subjects responded differently to different areas in the virtual space, as related to the spatial organization and to the virtual characters [12]. It could be interesting to compare this qualitative evidence with the objective physiological measure. Our tracker data definitely seems to indicate that almost all subjects spent more time on the side of the barman than in the left side of the bar.

4.4 Merging Temporal and Non-Temporal Data

Subjective descriptions provided by subjects can be insightful; examples from the experiment discussed here appear in [12]. However, how do we cross them with temporal data? We would like to have an equivalent of the interview, which is obtained during the experiment.

One such option may be to ask the subject to verbalize their subjective experience out loud, during the experience. When teaching drama this is often used as an exercise. At first the students report that it feels unnatural for them to speak out loud, but they quickly get used to it. Is it possible to train subjects to verbalize their thoughts and feelings, during an experience, to an extent that it does not interfere with their experience? We do not know, but we hope to explore this option.

Another option is possible if the VE scenario includes a well-defined narrative. If so, subjects can be encouraged, in the interview, to describe their feelings as related to certain events. For example, in a modified bar experiment, subjects can be encouraged to describe their feelings when a character tells them something intimate. Even though a few minutes pass from the time of the real experience to the time of the reconstruction, the information gathered in this way may be useful. Such recollection of the experience may be done during the video interview, as mentioned in Section 3.1.7, or with a replay of the VE events as suggested in Section 3.1.8.

4.5 Inter-Experiment Comparisons

Ideally, it should be possible to compare experiments carried out by different researchers in different labs, even if only part of the data is available. This is especially true for qualitative data such as physiological data and questionnaires, which could be assumed to be universal.

Obviously, there would be many differences in the settings and the contexts among different labs, and it would not be simple to compare the data. For example, if one group finds a much stronger physiological response to a virtual character than another group, we would need to carry out further work to understand why this happens: is this because of the technical setting of the lab, the software used, the specific task and scenario in the experiments, or something different. Still, we believe such comparisons would be insightful, especially when there are large differences among research groups.

We hope to compare the results of this experiment with others. In our lab there have been experiments carried out with a similar experimental procedure but with a different VE. More generally, we can compare the physiological responses in this experiment with the physiological responses obtained in other experiments using the same Cave setting.

5. Discussion

In this paper we detail the types of data that we collected during one experiment, and the techniques we used to analyze this data. There is still a long way to set standards for data sharing and analysis in the presence community.

We recognize that presence is a complex, multi-dimensional concept, which needs to be studied with multiple techniques in multiple levels. Thus, we expect that if presence research matures we will be faced with ever growing1 amounts of data, of different types, which will need to be analyzed.

In order for such data sharing to become widespread, there is a need for standards in data representation, and a standard set of tools and utilities that will allow converting the data into the commonly used tools. In this paper we do not yet suggest such standard. We believe it is too early to suggest complex mechanisms (e.g., using XML to annotate the data), but we hope that the next step would be for researchers to define data formats and provide generic utilities that import, export, and analyze data using this formats.

Finally, in order to avoid abuse of data, we would need to suggest copyright mechanisms, probably in the lines of GPL[6].

6. Conclusions

We encourage other researchers to use their methods in analyzing this data. There are large parts of our data that have not yet been analyzed, or only partially analyzed, and we welcome other researchers to apply other techniques to the data that we have already analyzed.

We encourage other researchers to publish their data, in addition to their results, as we have done in this paper. This will allow the community to analyze and compare experiments as a shared effort, assigning credit where due, of course.

Once a corpus of data is available and arranged in a systematic way, we can strive towards additional analysis methods. In particular we encourage researchers to investigate analysis of physiological data. The integration of the different data types presents an interesting challenge. We hope to address it in the future, using visualization and possibly data mining. We feel this would allow presence research to be established as a genuine scientific discipline by such eventual data publication and sharing[7].

Finally, we encourage researchers to use this methodology, which relies on large amounts of synchronized recorded data in a mediated experience, beyond presence research; we expect our setting to be useful for researchers in many areas of psychology.

Acknowledgements

This work has been supported by the European Union FET project PRESENCIA, IST-2001-37927. We would like to thank Anthony Steed for his comments on this paper.

.

---

[6] http://www.gnu.org/copyleft/gpl.html

[7] For example, as part of the Visible Human Project, the U.S. National Library of Medicine has recently made digitized datasets of male and female human cadavers available for research and education.

References

[1] F. Biocca, J. Burgoon, C. Harms, M. Stoner. Criteria and scope conditions for a theory and measure of social presence. In *Fourth Annual Int'l Workshop: Presence 2001*, Philadelphia (PA), USA, May 2001.

[2] D.A. Bowman, D.B. Johnson, L.F. Hodges. Testbed evaluation of virtual environment interaction techniques, *Proc. ACM symp. on Virtual reality software and technology*, pp 26—33, London, United Kingdom, 1999.

[3] A. Brogni, M. Slater, A. Steed. More breaks less presence. In *The 6th Annual International Workshop on Presence*. Aalborg, Denmark, 2003.

[4] C. Cruz-Neira, D. J. Sandin, T. DeFanti. Surround-screen projection-based virtual reality: The design and implementation of the CAVE, in *Proceedings of Computer Graphics (SIGGRAPH) Proceedings, Annual Conference Series*, 135-142, 1993.

[5] G. Florian, A. Stancak, G. Pfurtscheller. Cardiac response induced by voluntary self-paced finger movement. *Int'l J. Psychophysiology*, 28(3):273-83, 1998.

[6] E. Frecon, G. Smith, A. Steed, M. Stenius, O. Stahl. An overview of the COVEN platform. *Presence-Teleoperators and Virtual Environments*, 10:109-127, 2001.

[7] J. Freeman, S.E Avons, D. Pearson, W.A. Ijsselsteijn. Effects of sensory information and prior experience on direct subjective ratings of presence. *Presence-Teleoperators and Virtual Environments*, 8:1-13, 1999.

[8] J. Freeman, S.E. Avons, R. Meddis, D.E. Pearson, W. IJsselsteijn. Using behavioural realism to estimate presence: A study of the utility of postural responses to motion stimuli. *Presence: Teleoperators and Virtual Environments*, 9(2): 149–164, 2000.

[9] D. Friedman, Y. Feldman, A. Shamir, T. Dagan. Automated creation of movie summaries in interactive virtual environments. In *Proc. IEEE Virtual Reality 2004*, pp191--198, Chicago, IL, 2004.

[10] D. Friedman, R. Leeb, A. Antley, M. Garau, C. Guger, C. Keinrath, A. Steed, G. Pfurtscheller, and M. Slater. Navigating virtual reality by thought: First steps. *Proc. 7th Annual Int'l Workshop on Presence, PRESENCE 2004*, pages 160—167, Valencia, Spain, October 2004.

[11] M. Garau. The Impact of Avatar Fidelity on Social Interaction in Virtual Environments. *PhD thesis, University College London,* 2003.

[12] M. Garau, H. Ritter-Widenfeld, A. Antley, D. Friedman, A. Brogni, and M. Slater. Temporal and spatial variations in presence: A qualitative analysis. *Proc. 7th Annual Int'l Workshop on Presence PRESENCE 2004,* pages 232—239, Valencia, Spain, October 2004.

[13] C. Greenhalgh, M. Flintham, J. Purbrick, S. Benford. Applications of Temporal Links: Recording and Replaying Virtual Environments. In: *Proc. of IEEE VR 2002 (Virtual Reality),* pp 1-8, November, 2001.

[14] C. Guger., G. Edlinger, R. Leeb, G. Pfurtscheller, A. Antley, M. Garau, A. Brogni, D. Friedman, M. Slater. Heart-Rate Variability and Event-Related ECG in Virtual Environments. In *Presence 2004: The 7th Annual International Workshop on Presence*, pp. 240-245, 2004.

[15] J.C. Hart. On recording virtual environments, *Proc. IEEE Visualization '93 Symposium on Research Frontiers in Virtual Reality*, pp. 80-83, Nov. 1993.

[16] W. Ijsselsteijn, J. Freeman, H. de Ridder. Presence: Where are we? *Cyberpsychology and Behavior*, 4(2): 179– 182, 2001.

[17] T. Imai, A. Johnson, J. Leigh, D. Pape, T. DeFanti. Supporting transoceanic collaborations in virtual environment, In the *Proc. of the 5th Asia-Pacific Conf. on Comm. / 4th Optoelectronics and Comm. Conf. - APCC/OECC '99*, pp. 1059-1062, Beijing, China, October 19-21, 1999,.

[18] R. Lipshitz, M. Popper, N. Ron. Post-flight reviews in the Israeli-Defense-Force air force as an organizational learning mechanism. Presented at *the 3rd Int'l Conf. on Organizational Learning*, Lancaster University, June, 1999.

[19] M. Meehan, B. Insko, M. Whitton, F.P. Brooks. Physiological measures of presence in stressful virtual environments. *Acm Transactions on Graphics*, 21:645-652. 2002.

[20] T.B. Sheridan. Musings on telepresence and virtual presence. *Presence: Teleoperators and Virtual Environments*, 1:120–125, 1992.

[21] M. Slater. Measuring presence: A response to the Witmer and Singer presence questionnaire. *Presence-Teleoperators and Virtual Environments*, 8:560-565. 1999.

[22] M. Slater. How colorful was your day? Why questionnaires cannot assess presence in virtual environments. *Presence-Teleoperators and Virtual Environments*, 13:484-493. 2004.

[23] Slater, M., Brogni, A., Steed, A. (2003). Physiological Responses to Breaks in Presence: A Pilot Study. *Presence 2003: The 6th Annual International Workshop on Presence.*

[24] Slater, M and Steed, A. A virtual presence counter. *Presence: Teleoperators and Virtual Environments*, 9(5): 413–434, 2000.

[25] M. Slater, C. Guger, A. Antley, A. Brogni, D. Friedman, M. Garau, H. Ritter- Widenfeld, and A. Steed. Analysis of Physiological Responses to a Social Situation in an Immersive Virtual Environment, Submitted 2005.

[26] A., Steed, V. Vinayagamoorthy, and A. Brogni, "Using Breaks in Presence to Identify Usability Issues", To be presented at *the 11th International Conference on Human Computer Interaction*, 22-27, Las Vegas, July 2005.

[27] A. Steed, J. Mortensen, E. Frecon. Spelunking: Experiences using the DIVE System on CAVE-like Platforms. In *Immersive Projection Technologies and Virtual Environments*. Springer-Verlag/Wien, pp. 153-164, 2001.

[28] M. Thomsen. Discovering Mixed Reality: inventing design criteria for an action-based Mixed Reality, *PhD thesis, University College of London*, 2004.

[29] M. Usoh, E. Catena, S. Arman, M. Slater. Using presence questionnaires in reality. *Presence-Teleoperators and Virtual Environments*, 9:497-503, 2000.

[29] D. Watson, R. Friend. Measurement of Social-Evaluative Anxiety. *J. of Consulting and Clinical Psychology*, 33:448-&, 1969.

The Big Picture: Gaining Perspective by Examining Patterns in the (Tele)Presence Literature

Matthew Lombard, Matthew T. Jones
Temple University
lombard@temple.edu, matjones@temple.edu

Abstract

This paper reports initial findings from an analysis of 1,098 scholarly works identified as constituting the literature on (tele)presence. The value and implications of such a macro-level analysis and future research possibilities are discussed.

1. Introduction

In some important ways, the work we all do related to (tele)presence[1] is akin to the work of a team of builders constructing a building. Both involve large, diverse groups of people using a variety of materials, specialized tools, and expertise across many areas to create something they hope will be of lasting value. And to be successful, both endeavors require those involved to have a vision of, and to monitor, not just their immediate or micro-level activities and goals but the longer term or macro-level ones as well – i.e., the big picture.

In this paper we explain why it's important to examine the scope, evolutionary trends and demographic patterns in the (tele)presence literature, and describe and report initial findings from such an examination.

2. Reasons to examine the Big Picture of (tele)presence scholarship

There are (at least) four important reasons to examine (tele)presence scholarship from a macro level:.:

2.1. Efficiency

Just as builders need to understand and coordinate their diverse activities (e.g., foundation building, plumbing, electrical, masonry, roofing, cabinetry, painting, etc.) to complete the project as quickly and efficiently as possible, scholars working on a diverse set of questions and problems can accomplish their collective goal – building knowledge about a set of related phenomena – more efficiently if they understand the larger context of their work. We need to know what concepts, terms and definitions others are using; what micro-level problems and phenomena they're studying; and where to find reports of work that relates to our specific areas. While we each conduct literature reviews for our focused topic areas, without this macro perspective, we're likely to slave away in our offices and laboratories inventing and reinventing individualized and needlessly redundant theories, terminology and tools that slow our progress toward a comprehensive understanding of (tele)presence.

2.2. Identifying gaps and plotting new directions

Although efficiency is an important benefit of gaining perspective on a literature, it should not (and need not) come at the expense of creativity. The macro perspective can reveal new directions for research and theory related to (tele)presence.

Overall, examining patterns in the literature of any field can help identify gaps in that work, questions that still need to be examined, applications that should be considered, etc. Managers who oversee the building of physical structures similarly use checklists and closely examine the plans and progress across all aspects of the project to avoid overlooking important elements in the building process. In the case of scholarly knowledge building, a broader perspective on the literature provides a larger collection of the data (questions, findings, etc.) of individual studies to inductively develop new questions and avenues for exploration.

2.3. Tracking the evolution and current health of the field

The macro perspective on a literature, especially a new, interdisciplinary and rapidly changing literature such as ours, allows scholars to better understand the collective history of work in the field as well as the current health (or lack thereof) of the academic area. While building with bricks and mortar generally takes less time and is less abstract than building knowledge, both require regular review and assessment to insure that the structure being built is sound. And both require specific and objective measures of progress to satisfy stakeholders (in the academic context these are university administrators and granting agency officers). Presence scholars, especially those outside Europe, look forward to the day when a (tele)presence study is viewed as part of a valid and recognized tradition just as a study in psychology or biology.

---

[1] This term is used here because of the continuing confusion regarding the distinction between presence and telepresence; (tele)presence refers to either or both of these concepts.

2.4. Interest value and uncertainty reduction

Finally, gaining a ‘big picture’ view of a literature and field is inherently interesting (just as watching a building being constructed is to many people; see, e.g., [1]). Aside from learning about others in one’s community of scholars and the work that they’ve done, it should be satisfying to reduce our sense of uncertainty (and thus perhaps anxiety) about the outlines, merit and constituency of one’s academic area.

3. Attributes to observe in a macro-level analysis of the (tele)presence literature

The following sets of attributes provide a basis for a macro level analysis that would fulfill the goals described in the previous section.

3.1. Locations of works

Because (tele)presence phenomena occur in diverse contexts, relevant scholarship is conducted across several disciplines and is likely to appear in a diverse set of journals, books and database indices. Examining which and how many journals, databases, and other sources contain the (tele)presence literature provides a gauge of the level of dispersion and impact of (tele)presence concepts across academic disciplines and, conversely, information about which disciplines have had an impact on the study of (tele)presence.

3.2. Demographic and other characteristics of works

Patterns in basic reference information across (tele)presence publications should provide valuable perspective. Changes in the numbers of publications across time demonstrates the rate of growth of (tele)presence scholarship. The relative proportions of journal articles, books, book chapters, dissertations, etc. (and changes in these proportions over time) provide insight into how knowledge is being built and synthesized in the field, including the level of ‘maturity’ of the field. The use of “presence,” “telepresence,” and related terms (and changes in these over time) suggest the degree of centrality of the concept(s) as well as which terminology is favored. In addition, the proportions of works that report data-based research findings;. synthesis, review and theory building; and descriptions of evolving technology provide further insights regarding maturity and epistemology in the field.

3.3. Characteristics of authors of works

The gender and national/geographic and institutional affiliations of the authors of (tele)presence publications tell us a lot about our community. The number of authors of the publications also provides insight into the amount of collaborative work in the field.

4. Research Question and Method

Based on the above discussion, we conducted a study to address this general research question: What are the salient macro level attributes of the literature on (tele)presence? This involved identifying the literature, collecting the works within it, and coding specific variables.

4.1. Identifying the (tele)presence literature

The first step toward obtaining a macro perspective on the (tele)presence literature of course is to identify the literature. We have identified the (tele)presence literature (as of February 2005) and report the details elsewhere [2]; the complete list of references is available from the authors (and will be available online soon). The identified works are 1,098 academic or scholarly articles, books, etc. that are generally available to researchers (e.g., via libraries, full text databases, etc.) Thus, dissertations are included but conference papers (unless they appear in generally available proceedings), unpublished master's theses, and specialized or technical reports are not. The works are drawn from the following sources:

1. IJsselsteijn, Lombard, and Freeman’s 2001 article, “Toward a core bibliography of Presence” [3].

2. Searches in the citation databases ComAbstracts (Communication), Computer Abstracts, PsychInfo and ISI Web of Science for these keywords:

   telepresence, tele-presence, (tele)presence
   spatial presence
   social presence
   parasocial
   computers are social actors
   copresence, co-presence
   subjective presence
   virtual presence
   sense of presence
   perceived realism
   perceived reality
   perceptual realism
   social realism

   (Other popular (tele)presence-related terms such as “immersion,” “being there,” and, in fact, the term “presence” itself were omitted for the primary reason that they are in such general use that it would be impossible to use them in a keyword search and have any hope of sifting out irrelevant citations.)

and

3. The journal *Presence: Teleoperators and Virtual Environments*, its online, blind peer-reviewed supplement *Presence-Connect*, and the journal *Cyberpsychology & Behavior*; works containing the key words above and/or

the term “presence” used in its scholarly context are included.

A complete list of the 1,098 identified works is available at http://ispr.info/p-lit_index.htm

4.2. Collecting works in the literature

We have begun compiling electronic and printed copies of all 1,098 works in the literature. Many of the works are available at no cost online while others present economic or other challenges (e.g., the length and bulk of books and dissertations). To date we have acquired and printed copies of 286 of the identified works; we will obtain most of the others by the PRESENCE 2005 conference.

4.3. Coding variables

Using an Excel spreadsheet, we coded the following variables for each work in the literature: the database(s) and/or other source(s) that listed or contained the work; the full name of journals and proceedings in which works appeared; the year of publication; the format (article, book, book chapter, dissertation, proceedings); the appearance of any variant of the term “presence” in the title; whether the work presented data-based research findings (i.e., contained “method” and “results” headings and presented quantitative and/or qualitative findings), synthesis and theory building, or descriptions of technology; and the gender, institutional and national/geographic affiliation of the author (in cases of multiple authors we recorded this information for the first author and recorded the number of authors and whether their institutional affiliations were different or not).

While it would be valuable to examine additional variables as well, these most fundamental aspects of the literature should be considered first and will likely suggest additional variables for future analyses.

5. Results

5.1. Locations of works

Over half (55.6%, n=611) of the works in the (tele)presence literature appear in the ISI Web of Science database and over 4 in 10 (41.6%, n=457) can be found in PsychInfo, with smaller but substantial numbers of the works in the other databases and sources (see Table 1). Nearly three quarters (72.7%, n=797) of the works appear in only one of the eight sources, with another 16.5% (n=181) appearing in two of them (see Table 2).

Table 1. Sources in which (tele)presence works are found.

Source	Number of works	Percent of works
IJsselsteijn, Lombard, and Freeman (2001) [3]	95	8.6
ComAbstracts (Communication)	73	6.6
Computer Abstracts	124	11.3
PsychInfo	457	41.6
ISI Web of Science	611	55.6
Presence: Teleoperators and Virtual Environments	105	9.6
Presence-Connect	9	0.8
Cyberpsychology & Behavior	54	4.9

Table 2. Number of sources in which (tele)presence works can be found.

Number of sources	Number of works	Percent of works
1	797	72.7
2	181	16.5
3	107	9.8
4	12	1.1
5	0	0.0
6	0	0.0
7	0	0.0
8	0	0.0
Total	1098	100.0

Of the 930 (84.7%) of the works that appear in journals or proceedings, the greatest number (105) appear, not surprisingly, in the MIT Press journal *Presence: Teleoperators and Virtual Environments*; an additional 54 appear in *Cyberpsychology & Behavior* and 8 other journals contain 11 or more of the works (see Table 3). The works appear in 442 different journals (427) and proceedings (15), and 312 (33.58%) of these venues contain only one of the works (see Table 4). The journals come from diverse disciplines including art, business, communication, computer science, education, engineering, linguistics, medicine, music, nutrition, oceanography, philosophy, physics, psychology, religion, social work, sociology, and more.

Table 3. Journals and proceedings that contain the most (tele)presence works (total n=930).

Journal	Number of works
Presence: Teleoperators and Virtual Environments	105
Cyberpsychology & Behavior	*54*
Communication Research	14
BT Technology Journal	13
Human Communication Research	13
IEEE Transactions on Circuits and Systems for Video Technology	12
International Journal of Human-Computer Studies	12
Psychological Reports	12
Journal of Broadcasting and Electronic Media	11
Journal of Communication	11

Table 4. Number of journals and proceedings that contain different numbers of works in the (tele)presence literature.

Number of works in journal	Number of journals	Percent of journals
105	1	11.30
54	1	5.81
14	1	1.51
13	2	2.80
12	3	3.88
11	2	2.37
10	0	0.0
9	3	2.91
8	2	1.72
7	1	0.75
6	3	1.94
5	5	2.69
4	15	6.46
3	26	8.40
2	35	13.99
1	312	33.58
Total	930	100.0

5.2. Demographic and other characteristics of works

The earliest publication in the literature is from 1937, Mauge's. book *La destinée et le problème du vrai : L'esprit et le réel perçu*. [Destiny and the problem of truth. The mind and perceived reality] (published by Alcon). A clear growth trend can be seen in the number of publications in the subsequent years (see Figure 1). Three quarters (78.1%, n=858) of the works were published in the decade between 1995 and 2004.

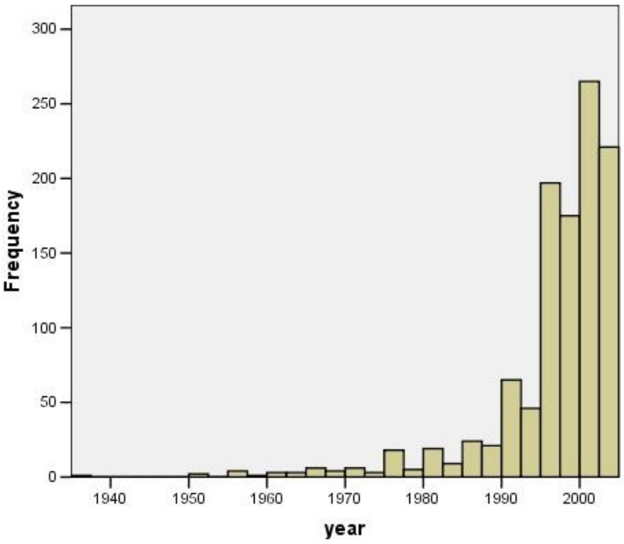


Figure 1. Publications per year.

Over 80 percent (n=911) of the works are articles, with smaller numbers of dissertations, book chapters, books and papers from proceedings (see Table 5).

Table 5. Formats of works in the (tele)presence literature.

Number of sources	Number	Percent
Articles	911	82.9
Proceedings	21	1.9
Books	44	4.0
Book chapters	49	4.5
Dissertations	72	6.6
Other	1	0.1
Total	1098	100.0

"Presence" and/or variants of that term appear in the titles of 314 (28.6%) of the works, with "telepresence" the most common variant (see Table 6).

Table 6. Frequency of “ Presence” terms in titles of works.

Terminology	Number	Percent
Presence	149	13.6
Telepresence	120	10.9
Social presence	34	3.1
Copresence	11	1.0
Virtual presence	5	0.5
Immersive presence	1	0.1
Perceived presence	1	0.1
Spatial presence	1	0.1
Subjective presence	1	0.1

As of this writing we’ve acquired and compiled printed copies of 286 or 26% of the 1,098 works in the literature and coded 77 of these regarding the nature of the work [All of the works will be acquired and coded before PRESENCE 2005.]. Nearly 6 in 10 (58.4%, n=45) are reports of data-based research findings, another 3 in 10 (28.6%, n=22) are focused on review, synthesis and theory-building, and the rest (7.7%, n=10) are descriptions of technologies and their application.)

5.3. Characteristics of authors of works

Among the initial sample that has been coded, male authors (67.5%, n=52) outnumber females (24.7%, n=19); the gender of a few authors (7.8%, n=6) cannot be determined from the work.

The (first) authors of the 77 coded works are, or at least were at the time the works were published, affiliated with 65 different institutions; Eindhoven (5), Michigan State (3) and Temple (3) universities are, so far, most represented in this small sample.

Eleven nations are represented in the institutional affiliations of the 77 coded works: England, France, Germany, Israel, Italy, Korea, The Netherlands, Scotland, Spain, Sweden and the U.S. The U.S. (45%, n=35) and England (14.3%, n=11) were the most frequent affiliations.

Of the 58 multi-authored works coded, the authors of 37 (63.8%) have different institutional affiliations, indicating substantial collaboration across geographical boundaries.

Nearly 4 in 10 (37.8#, n=415) of the entire collection of published works have only a single author; another 3 in 10 (26.9%, n=296) have two authors, and the remaining works have 3 or more authors (see Table 7).

Table 7. Number of authors of (tele)presence works.

Number of authors of work	Number of works	Percent of works
1	415	37.8
2	296	26.9
3	169	15.4
4	106	9.6
5	41	3.7
6	25	2.3
7+	42	3.8
No author identified	4	0.4
Total	1098	100.0

6. Discussion

These results provide an initial macro-level view of the (tele)presence literature and field. So what do we see from this vantage point?

The fact that three quarters of the works in the (tele) presence literature can be found in only one of eight key research databases and other sources and no source contains more than 56% of the works, suggests opportunities to increase the efficiency with which we explore (tele)presence concepts and phenomena. If and when a central database becomes available for our field – PresenceInfo? – it would likely facilitate the exploration of all of these sources in conjunction with one another. In turn, this would allow us to move ahead more productively and efficiently because we’d find within and among them unexpectedly relevant theories, findings, and descriptions of evolving technology.

The big picture perspective demonstrates that the (tele)presence field is young, highly interdisciplinary, and growing. While the oldest publication in the literature is from 1937, three-quarters of the works are from the last 10 years. Despite the fact that *Presence: Teleoperators and Virtual Environments* and *Cyberpsychology & Behavior* are the predominant fora, (tele)presence works appear in a startlingly large number of journals and proceedings across dozens of disciplines and subdisciplines. The number of publications in the field, including articles, books, book chapters and dissertations, is increasing rapidly. The literature contains about twice as many data-based research reports as works in which the focus is on synthesis, and a smaller number of works that describe developing technologies. People can reasonably differ regarding what the appropriate balance of these is, but clearly all are valuable and required in a healthy discipline.

On the other hand, the terms “presence” and “telepresence” appear in the title of less than 30% of the works and it’s not clear that they’re being used to represent distinct concepts; in addition, the use of the word presence, because of its other common usages, represents a challenge to scholars searching for relevant literature.

Hopefully, compiling and analyzing this and other macro level information about (tele)presence scholarship will help us learn more (and reduce our uncertainty) about the “presence” community and the work of all of its members.

6.1. Future research

The big picture view should be updated regularly in order to monitor the evolution of (tele)presence scholarship, including changes in its dispersion across disciplines, databases, venues, and formats, as well as demographic trends among its members. Tracking changes in the percentage of works that include different key conceptual labels (primarily “presence” and “telepresence”) should provide a good indicator of the centrality of (tele)presence in the literature, and thus one aspect of the field’s maturation.

Comparative data from other fields (especially regarding collaboration, format and focus of works) would provide additional useful perspective on the (tele)presence literature and field.

Many more detailed analyses of the literature, e.g., of the conceptual and operational definitions of (tele)presence, and the specific topics and conclusions within the works, would help identify gaps in the literature and suggest new research questions to pursue.

An example of a tool that could be useful in the analyses is RefViz, available from Thomson™. RefViz literature visualization software permits the user to obtain “at-a-glance” summaries of major themes as well as conceptual relationships and correlations between concepts in a set of works. Certain practical obstacles (such as the lack of database abstracts in some of the literature sources) would need to be overcome to use this tool, but the results are likely to lead to many of the benefits described at the beginning of this paper.

Just as builders of a physical building must have a macro level view of their work and their progress toward their goals, so must builders of knowledge. This project provides this kind of big picture perspective for builders of knowledge regarding what we consider the important and fascinating topic of (tele)presence.

References

[1] Kimmel Center, Inc. Building the Center: Time-Lapse Construction Video. Retrieved June 8, 2005 from http://www.kimmelcenter.org/building/timelapse.php

[2] Lombard, M. & Jones, M. T. (2005). Identifying the (Tele)Presence Literature. *Cyberpsychology & Behavior* (submitted).

[3] IJsselsteijn, W. A., Lombard, M., & Freeman, J. (2001). Toward a core bibliography of presence. *Cyberpsychology & Behavior, 4*(2), 317-321.

Are we together? Exploring the similarities between the concepts of transportation imagery model and presence and their possible impact on persuasion

Cheryl Campanella Bracken
Cleveland State University
{c.bracken@csuohio.edu}

Abstract

While the concept of presence has been investigated for over 10 years, there have been few studies investigating presence and persuasion. Work in the area of consumer psychology has recently begun to explore the possible connections between aspects of presence and persuasion and have developed a new concept called – transportation imagery model. This paper attempts to demonstrate the similarities between these two concepts (Presence and Transportation imagery model) and the benefits of integrating this new concept into the Presence fold.

Keywords - ***Presence, Transportation Imagery Model, Persuasion.***

1. Introduction

Many commercials use stories to entertain and sell their products. Advertisers are aware that people can and often expect to be entertained by commercials . Some television commercials are mini-movies can be considered "mini-movies" that are designed to meet this types of consumer expectations. Commercials are a type of television content that has not yet been explored by Presence researchers.

The extent to which media users are "lost in", "immersed in", have a sense of "being there", or "connected to" media has been explored by Presence researchers in a variety of disciplines. These types of experiences have been recently identified and examined within the area of consumer psychology research [1-2]. These scholars are referring to this phenomenon as "transportation" and have developed the "transportation imagery model" (TIM) [2-3]. To presence scholars the phenomenon they describe will sound quite familiar, but these researchers have not connected their conceptualization, operationalization, or findings to the larger (tele)presence literature that has been developing since the early 1990s. This paper discusses presence, introduces TIM, and discusses the relevant findings of both concepts as they relate to persuasion within media. Further this paper attempts to integrate TIM into the presence literature and along they way illustrate how these areas of research can be mutually beneficial.

2. Presence

The concept of presence or (tele)presence has been defined as the "Illusion of non-mediation" [4]. Since the introduction of this conceptualization several types of presence have been identified and studied: physical or spatial presence (the sense of being there) [5], social presence (having a sense of another personality) [6], and co-presence (a sense of being together with another person either in a mediated space) [8]. This paper address the similarities between spatial/physical presence and TIM.

2.1 Presence and Entertainment Media

What do we know about presence and entertainment media? While the bulk of presence research has focused on the creation of and media users' responses to highly immersive environments [8-11], there are some researchers who have also studied entertainment media and presence. The use of media to induce presence-like experiences is not a new phenonomenon. For a detailed discussion of presence and media history, see [12]. Other entertainment media and presence researchers have found that television audiences have presence experiences [13], and film makers intentionally encourage audiences to "stay in the film" [14]. More recent research on video games has shown presence is not only experienced by video game players [15-16] but that some researchers have argued presence should be considered a moderating variable [15] when designing video game studies. The findings of these studies unanimously found that presence can be experienced by media users in far less immersive environments and in the case of television and film, less interactive media that virtual reality environments.

2.2 Encouraging Presence in Entertainment Media

Researchers in the area of entertainment and presence have studied some of the variables that have been linked to encouraging presence sensations – these include both form and content variables. Previously, two types of form variables were identified [4]: user controlled and producer controlled. Examples of user controlled variables are screen size, home theatre (surround sound), image quality, etc… Findings related to user controlled variables demonstrate that larger screen size (or field of view) [17-18], improved quality of both the sound [19], the image [20] increase the likelihood that the media user will experience a sense of presence. Similar findings were documented with increased image quality and video games [16].

Producer controlled variables include pace, camera shots, cuts, etc…The findings for this type of form variable have not been individually studies by presence researchers, however, the literature suggests that faster, quick cutting programs are likely to decrease a media users sense of presence (because it makes the experience less like-like). Producer controlled variables have not been as widely studied in regards to television and presence, but have been investigated with videogames. Video games have also been found to evoke a sense of presence [21]. Recently, there have been several studies that explored the use of narratives (or storyline) in videogames and whether it has any impact on gamers' sensations of presence. There appears to be consensus that the use of narrative increases players' level of enjoyment and sensations of presence [22-23].

Another area of entertainment and presence research has explored the influence of content on media users' sensations of presence. The type of content viewed or the type of game played has an impact on the level presence experienced. Examples include arousing versus calming content [24] with viewers who saw the calming content being reporting higher level of presence. Point of view movement in television has been found to induce a sense of presence in television viewers [13] and first-person shooter games [23].

Together these studies on presence and entertainment media make a strong case that presence as an experience that is not limited to highly immersive media systems. This point is important to acknowledge because it is the use of these types of media that most people are likely to use and experience (with the exception of simulation rides and arcade VR systems).

2.1 Presence and persuasion

There are only a few studies exploring the impact of presence on persuasion [25-28]. In a series of experiments exploring the role of presence on consumers' responses to 3-D advertising, researchers [27] found that presence influenced consumers' feeling of physical presence and engagement and their purchase intentions. While others [25, 28] found that presence impacts the processing route by which the information is processed.

An explanation for why presence should increase advertising effectiveness is offered by [29] who state

> "One thing interactivity is thought to increase is the sense of 'presence,' and presence is thought to lead to a variety of effects which include enjoyment and persuasion, primary goals of advertising. Therefore presence, and research and theory concerning presence, may serve as a useful guide to understanding and marshaling the use of interactivity in advertising to maximum effect" (paragraph 18).

Presence has also been found to impact persuasion in less immersive environments, specifically [26], found that television viewers who reported a sense of physical presence influenced purchase intentions and consumer confidence. It is interesting to note that the authors found differences between the two dimensions of physical presence they measured

- "arrival" or being present in the mediated environment
- "departure" or a feeling of separation from the physical environment.

Specifically, arrival had a positive impact on consumer confidence in a product/brand, while departure had a positive impact on purchase intentions. Together these results indicate that presence can have an impact on the process of persuasion, specifically advertising, but possibly on other types of persuasion in a mediated setting (i.e., product placement, edutainment, etc…). The following section will detail another concept, Transportation Imagery Model [2] and its role in persuasion.

3. Transportation Imagery Model

This particular model has emerged within the areas of Psychology and Consumer Psychology. The model is concerned with predicting the extent to which consumers are persuaded by advertisements. The model holds that many advertisements contain drama or narrative stories. The basic premise of this model is that narrative structures are able to "involve, captivate, and entertain consumers" [1, p. 267]. It has been argued that narratives or dramas "draw the viewer into the action it portrays [30, p. 335]. The same authors also assert that the audience becomes "lost" in the narrative and empathizes with the characters (p. 335). The Transportation Imagery Model (TIM) has been developed from these prior assumptions, within narration, as well as reader response theory, and diegesis [31].

The TIM model adapts the term "transportation" from [32] who refers to a "traveler" as a media user (though Gerrig is only concerned with readers of text-based materials) who make a mental journey to a distant location (typically the reality of the text) and returns a "changed" person. Green and Brock assert that transportation can have a physical and psychological dimension for the traveler or transported individual. Even though they borrow heavily from the narrative or print tradition, they argue that "transportation is not limited to the reading of written material. Narrative worlds are broadly defined with respect to modality; the term "reader" may be construed to include listeners, viewers, or any recipient of narrative information" [2, p. 702].

The model is seen as having a high to low continuum of "transportingness" with a high level being more likely to be persuasive [3, p. 334]. Persuasion is felt to be influenced by two factors: 1) imagery ability and 2) absorption propensity. Imagery ability [33] is the ability of a person to visualize or mentally rehearse, but can occur without the physical stimulus. While absorption is seen the disposition of individuals to have their attentional resources completely engaged in the activity [34]. However, research on the impact of these two individual differences and their impact on TIM.

3.1 TIM and information processing

The mechanism by which TIM leads to persuasion is by "reduced negative cognitive responding, realism of experience, and strong affective responses" [2, p. 702]. Green and Brock [2] assert that "Transportation into a narrative world is conceptualized as a distinct mental process, which may mediate the impact of narratives on beliefs (p. 324). One of the most interesting assertions of the TIM is that this type of persuasion occurs via a separate route from the dual process models of attitude change (.i.e., Elaboration Likelihood [35] and Heuristic-Systematic Persuasion Model [36]).

The TIM is different argue Green and Brock (2002) because personal relevance is not a necessary condition for the desirable central processing (in Elaboration Likelihood Model) or persuasion. In fact the researchers have empirical evidence demonstrating strong held assertions are just as influenced as weak held assertions when exposed to a persuasive narrative [37]. It also noteworthy that need for cognition appears to play no part in TIM.

4. Presence and TIM

There are obvious similarities between the concept of presence and the transportation imagery model. Aside from the specific application of TIM to persuasion, both focus on the perceptual process by which media users are willing to ignore or transcend the technology they are using to access the content. There has been previous overlap of some the dimensions identified within these concepts. Specifically, individual differences, such as absorption, has been identified as playing a moderating role on the amount of transportation (and in turn the level of persuasion) and was included in the "presence equation" [38].

The language and terms used by TIM researchers to describe the experience media users have when they are transported sounds very like the early presence literature. Some examples include: "immersion into a text", "lost", "absorbed", and "transported." Even the term "transportation" has a direct link to the presence literature. There were six dimensions of presence identified by [4], one of which is transportation.

However, it is here that the concepts being to differ. Lombard and Ditton use the term "transportation" to refer to a group of three types of feelings: "You are there," (user is transported to a different location); "It is here," (people/things come to user); and "We are together," (use and others are together in shared space). Please note, this is one more than was discussed earlier where the different sensations were referred to as arrival and departure. However, these are similar types of experiences [26].

The TIM only uses this first dimension (the feeling that a media user is either physically or mentally in a different local than the physical one they occupy) when conceptualizing and measuring what they refer to as transportation.

Another and perhaps more interesting difference is the literature which these concepts have been built. The (tele)presence literature predominantly developed out of computer science, and social science, while the TIM grew out of reader response and narrative theory. It has also only been tested with television commercials and print ads (non-immersive environments) However, what is missing from the TIM literature is the mechanism for how this process occurs. The presence literature has come a bit further in making claims about the process (see [39] for a full discussion). Another area where these two concepts may inform each other is the "book problem" [41-42].

4.1Presence and TIM research questions

These two concepts share several obvious research questions, including:

- Does experiencing a sense of presence also increase the likelihood a media user will be persuaded within highly immersive environments?
- Can presence and/or TIM be used an as explanation for the effectiveness of product placement in film, TV shows, or video games?
- Does presence offer a stronger theoretical grounding to TIM?
- Will high-definition television be more persuasive than the current standards?
- Can TIM and/or presence be used to encourage healthier or pro-social messages conveyed through entertainment?

Conclusion

The concepts of presence and transportation imagery model share a common interest, the investigation of feeling of being connected to or in an artificial or virtual environment (though perhaps TIM would claim it was only in one's imagination). The similarities are strong enough to consider TIM a new area of presence research. Researchers using TIM may benefit from both the breath and depth of existing presence research. It seems that presence and TIM may be a way of exploring the persuasiveness of entertainment content, such as "edutainment" – which includes an on going narrative, and for which there is empirical evidence that it is persuasive and little about the mechanism [40]. Presence may illuminate the processes by which audiences are in a state to be persuaded.

References

[1] Escalas, J. E. (1998). Advertising narratives: What are they and how do they work? In B.B. Stern (Ed.). *Representing Consumers: Voices, views, and visions.* New York: Routledge.

[2] Green, M.C., & Brock, T. C. (2000). The role of transportation in the persuasiveness of public narratives. *Journal of Personality and Social Psychology, 79*(5), 701-721.

[3] Green, M. C., & Brock, T. C. (2002). In the mind's eye: Transportation-imagery model of narrative persuasion. In M. C., Green, J. J. Strange, & T. C. Brock (Eds). *Narrative Impact: Social and Cognitive Foundations.* New York: Lawrence Erlbaum.

[4] Lombard, M., & Ditton, T. B. (1997). At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication, 13*(3). Available: http://www.ascusc.org/jcmc/vol3/issue2/lombard.html

[5] Minsky, M. (1980, June). Telepresence. *Omni,* 45-51.

[6] Biocca, F., Harms, C., & Burgoon, J. K. (2003). Toward a More Robust Theory and Measure of Social Presence: Review and Suggested Criteria. *Presence: Teleoperators & Virtual Environments, 12,* 456- 481.

[7] Zhao. S. (2003). Toward a taxonomy of copresence. *Presence: Teleoperators & Virtual Environments, 12,* 445-455.

[8] Bystrom, K.E., Barfield, W., & Hendrix, C. (1999). A Conceptual Model of the Sense of Presence in Virtual Environments. *Presence: Teleoperators & Virtual Environments, 8,* 241-245.

[9] Mania, K., & Chalmers, A. (2001). The Effects of Levels of Immersion on Memory and Presence in Virtual Environments: A Reality Centered Approach. *CyberPsychology & Behavior, 4,* 247-265.

[10] Schubert, T., Friedmann, F., Regenbrecht, H., (2001). The Experience of Presence: Factor Analytic Insights. *Presence: Teleoperators & Virtual Environments, 10,* 266-282.

[11] Slater, M., Sadagic, A., Usoh, M., & Schroeder, R. (2000). Small-Group Behavior in a Virtual and Real Environment: A Comparative Study. *Presence: Teleoperators & Virtual Environments, 9*, 37-52.

[12] IJsselsteijn, W. (2003). Presence in the past: What can we learn from media history? In G. Riva, F. Davide, & W. IJsselsteijn (Eds*), Being there: Concepts, effects, and measurement of presence in synthetic environments*. Amsterdam: Ios Press.

[13] Lombard, M., Reich, R. D., Grabe, M. E., Bracken, C. C., & Ditton, T. B. (2000). Presence and television: The role of screen size. *Human Communication Research, 26*(1), 75-98.

[14] Marsh, T. (2003). Presence as experience: Film informing ways of staying there. *Presence: Teleoperators & Virtual Environments, 12*(5), 538-549.

[15] Lee, K.M. (2005). *Presence as a mediator for psychological effects of computer games.* Paper presented to the Communication & Technology Division of the annual International Communication Association, New York

[16] Bracken, C.C., & Skalski, P. (2005, May). *Presence and video games: The impact of image quality and skill level.* Accepted for presentation to the Information Systems division of the annual conference of the International Communication Association, New York, NY.

[17] Bracken, C. C., & Atkin, D. (2004). How screen size affects perception of television: A survey of presence-evoking technology in our living rooms. *Visual Communication Quarterly, 59*(4), 23-28.

[18] IJsselsteijn, W., de Ridder, H., Freeman, J., Avons, S. E., & Bouwhuis, D. (2001**).** Effects of Stereoscopic Presentation, Image Motion, and Screen Size on Subjective and Objective Corroborative Measures of Presence. *Presence: Teleoperators & Virtual Environments, 10,* 298 – 312.

[19] Lessiter, J., Freeman, J., & Davidoff, J. (2001). *Really Hear? The Effects of Audio Quality on Presence.* Paper presented to the fourth annual Workshop on Presence, Philadelphia, PA.

[20] Bracken, C. C. (2005). Presence and image quality: The case of high definition television. *Media Psychology, 7*(2), 191-205.

[21] Eastin, M. S., Griffiths, R., & Lerch, J. (2005). *Beyond Shooter Games: How Game Environment, Game Type, and Competitor Influence Presence, Arousal and Aggression.* Paper presented to the Communication & Technology Division of the annual International Communication Association, New York.

[22] Lee, K. M., Jin, S. A., Park, N., & Kang, S. (2005). *Effects of Narrative on Feelings of Presence in Computer-Game Playing.* Paper presented to the Communication & Technology Division of the annual International Communication Association, New York.

[23] Schneider, E. F, Lang, A., Shin, M., Bradley, S.D. (2003). Death with a Story How Story Impacts Emotional, Motivational, and Physiological Responses to First-Person Shooter Video Games. *Human Communication Research, 30*, 361 - 376.

[24] Dillon, C., Keogh, E., Freeman, J., & Davidoff, J. (2001). *Presence: Is Your Heart In It?* Paper presented to the annual meeting of the Presence Workshop, Philadelphia, PA.

[25] Grigorovici, D. (2003). Persuasive effects of presence in immersive virtual environments. In G. Riva, F. Davide, & W. IJsselsteijn (Eds*), Being there: Concepts, effects, and measurement of presence in synthetic environments*. Amsterdam: Ios Press.

[26] Kim, T., & Biocca, F. (1997). Telepresence via television: Two dimensions of telepresence may have different connections to memory and persuasion. *Journal of Computer-Mediated Communication, 3*(2), Available at: http://www.ascusc.org/jcmc/vol3/issue2/kim.html

[27] Li, H., Daugherty, T., & Biocca, F. (2002). Impact of 3-D advertising on product knowledge, brand attitude, and purchase intention: The mediating role of presence. *Journal of Advertising, 31,* 43-57.

[28] Skalski, P., & Tamborini, R. (2005). *Vividness, Social Presence, and persuasion: Reconsidering the influence of modality on attitude formation.* Paper presented to the Information Systems Division of the annual International Communication Association, New York.

[29] Lombard, M., & Synder-Duch, J. (2001). Interactive advertising and presence: A framework. *Journal of Interactive Advertising, 13*(2). Available at www.jiad.org.

[30] Deighton, J., Romer, D., & McQeen, J. (1989). Using drama to persuade. *Journal of Consumer Research, 16,* 335-343.

[31] Burch, N. (1979). *To The distant observer: Form and meaning in the Japanese cinema.* London: Scolar Press.

[32] Gerrig, R. J., (1993). *Experiencing narrative worlds.* New Haven, CT: Yale University Press.

[33] Sheenan, P.W. (1967). A shortened form of the Bett's questionnaire upon mental imagery. Journal of Clincial Psychology, 23, 386-389.

[34] Tellegen, A., & Atkinson, G. (1974). Opennes to absorbing and self-altering experiences ("absorption'), a trait related to hypnotic susceptibility. *Journal of Abnormal Psychology, 83*(3), 268-277.

[35] Petty, R. E. and Cacioppo, J. T. (1986) *Communication and persuasion: Central and peripheral routes to attitude change,* New York: Springer-Verlag.

[36] Chaiken, S., Wood, W., & Eagly, A. H. (1996). Principles of persuasion. In E.T. Higgins and A. Kruglanski (Eds.), S*ocial psychology: Handbook of basic mechanisms and processes.* New York: Guilford Press.

[37] Wheeler, S. C., Green, M., C., & Brock, T. C. (1999). Fictional narratives change beliefs: Replications of Prentice, Gerrig, & Bailis (1997) with mixed collaboration. *Psychonomic Bulletin and Review, 6,* 136-141.

[38] Sas C. & O'Hare G.M.P. (2003). Presence Equation: An Investigation into Cognitive Factors Underlying Presence. *Presence: Teleoperators & Virtual Environments,12*, 523-537.

[39] Vorderer, P, Wirth, W., Saari, T., Gouveia, F. R., Biocca, F., Jäncke, F., Böcking, S., Hartmann, T., Klimmt, C., Schramm, H., Laarni, J., Ravaja, N., Gouveia, L. B., Rebeiro, N., Sacau, A., Baumgartner, T. & Jäncke , P. (2003). *Constructing Presence: Towards a two-level model of the formation of Spatial Presence.* Unpublished report to the European Community, Project Presence: MEC (IST-2001-37661). Hannover, Munich, Helsinki, Porto, Zurich.

[40] Stitt, C. R., & Nabi, R. (2005). *The persuasive impact of narratives: A comparison across message types and modalities*. Paper presented to the Information Systems Division of the annual International Communication Association, New York.

[41] Schubert, T., & Crusius, (2002). *Five theses on the book problem: Presence in books, film and VR.* Paper presented to PRESENCE 2002- Fifth annual Workshop on Presence, Porto, Portugal.

[42] Gysbers, A., Klimmt, C. , Hartmann, T., Nosper, A. & Vorderer, P. (2004). *Exploring the Book Problem: Text Design, Mental Representations of Space, and Spatial Presence in Readers*. Paper presented to Seventh annual Workshop on Presence, Valencia, Spain.

Session 5

22nd September, 2005

Individual and Cultural Effects

11.30-12.00 *Individual Differences in the Sense of Presence*
Ivan Alsina Jurnet, Cristina Carvallo Beciu and José Gutiérrez Maldonado
Grupo de investigación sobre aplicaciones de la realidad virtual en Psicología clínica, University of Barcelona, Spain

12.00-12.30 *The Impact of Personality Factors on the Experience of Spatial Presence*
Ana Sacau[1], Jari Laarni[2], Niklas Ravaja[2] and Tilo Hartmann[3]
[1] University Fernando Pessoa, Portugal
[2] Helsinki School of Economics, Finland
[3] Department of Journalism and Communication Research, Hanover University of Music and Drama, Germany

12.30-13.00 *Culture Matters - A Study on Presence in an Interactive Movie*
Jun Hu and Christoph Bartneck
Department of Industrial Design, Eindhoven University of Technology, The Netherlands

Individual Differences in the Sense of Presence

Ivan Alsina Jurnet, Cristina Carvallo Beciu and José Gutiérrez Maldonado
University of Barcelona
{ivanalsinajurnet@lycos.es, cristina.carvallo@lycos.es, jgutierrezm@ub.edu}

http://www.ub.es/personal/rv/rv1.htm

Abstract

There is a lack in the literature of studies that investigate the human factors involved in the engagement of presence. The present study is addressed to investigate the influence of five user's characteristics (spatial intelligence, personality, cognitive style, computer experience and test anxiety) on the sense of presence. This study is the first one to investigate the relationship between spatial intelligence and presence, and it's also pioneer in investigating the influence of personality characteristics on the sense of presence in an immersive virtual reality system. Results suggest that spatial intelligence, introversion, and anxiety influence the sense of presence experienced by the user.

1. Introduction

One of the core features in the virtual reality treatment of psychological disorders is the sense of presence. Although there is not a common definition of presence and it is a rather difficult concept to define and measure, there is a consensus to define it as a multi-component construct influenced by technological factors on the one hand, and by human factors on the other hand. The aim of this study is the research of the human factors (spatial intelligence, personality, cognitive style, computer experience and test anxiety) that can determine the engagement of the sense of presence in environments designed to treat students with test anxiety. This is the first study that is conducted to investigate the influence of spatial intelligence in the sense of presence, also it is pioneer using an immersive virtual reality system to investigate the relationship between the personality characteristics of the users (measured through the EPQ-RS questionnaire) and presence.

According to Steuer (1), determinant factors affecting presence can be grouped into three dimensions: vividness, interactivity and user characteristics. Steuer defines vividness as the representational richness of mediated environments defined by its formal features, that is, the way in which an environment presents information to senses. Interactivity is defined as the extent to which users can participate in modifying the form and content of a mediated environment in real time. Finally, the influence of user's characteristics derives from the individual differences in the sense of presence when subjects are confronted with the same virtual environments. In a similar way, Lombard (2) pointed that presence is determined by media characteristics and user characteristics. Media characteristics are divided into: media form, that includes the properties of a display medium (such as the extent of information presented, user's ability to modify the aspects of the environment, etcetera); and media content, that includes the objects, actors, and events represented by the medium. User's characteristics refer to the range of individual differences (for example age, gender, user's perceptual, cognitive or motor abilities, and so on).

Most of the studies that attempt to specify the determinants of the sense of presence have focused in the media form, concluding that some of the factors which can influence the subjective experience of presence are: the field of view (3,4,5,6), the foreground/background manipulations (7), the update rate (8), stereoscopy (5,9, 10), geometric field of view (9), pictorial realism (11,12), image motion (5), the use of a CAVE versus a desktop VR (13) or a HMD (14), spatial sound (15), the number of audio channels (16), tactile (17,18) or olfactory cues (18), the use of head tracking (15, 19), the feedback delay (11), the possibility to interact with the virtual environment (11, 20) or the body movement (21).

Relating to the influence of media content on presence, Hoffman (22) encountered that the sense of presence of experienced chess players was enhanced when chess pieces were positioned in a meaningful way,. In this line of research, EMMA project is aimed to study if emotions may enhance presence (23). Thus, if an environment is able to produce anxiety, sadness, joy, etc. it will be more probable that the user feels present in that environment. The results obtained at this moment show that presence could be influenced by the emotions that the environment is able to provoke to the user (24, 25).

Although presence is a psychological phenomenon (26), little research has been done about the user's characteristics involved in its engagement, thus previous discussions have been typically based on informed conjecture rather than research. Despite this fact, several authors (1, 2, 27, 28) state that presence is not only a direct function of the characteristics of the system but also of the human factors. In words of Schubert (29): "Stimuli from a VE are only the raw material for the mind that constructs a mental picture of a surrounding world, instead of a mental picture of pixels on the display in front of the eyes".

An interesting issue to point here is the difference between the terms "immersion" and "presence". In this line, Slater (26) distinguished between the terms immersion (an objective description of aspects of the system such as field of view) and presence (a subjective phenomenon such as the sensation of being in a VE). Three years later, Kalawsky (30) argue that presence is a cognitive parameter whereas immersion essentially refers to the physical extent of the sensory information, being a function of the enabled technology.

The pioneer study investigating the relationship between user's characteristics and sense of presence was conducted by Slater and Usoh (31) who distinguished between exogenous and endogenous factors responsible for determining the extent of presence. These authors use the therapeutic technique known as NeuroLinguistic Programming (NLP) to characterize the user's psychological representational and perceptual systems. This model claims that subjective experience is encoded in terms of three main representation systems: Visual, Auditory and Kinesthetic, and people usually prefer one system over the others. Furthermore, the experiences and remembers of a given individual are encoded in one of these perceptual positions: first (egocentric standpoint), second (from a standpoint of another person) or third (from a non personal view). In their study, 17 students were assigned to either an experimental group (n=9) or a control group (n=8). The control group was endowed with a disembodied arrow cursor, and the experimental group had a virtual body that responded to the participant movements. All the students were exposed to the same virtual environment (VE): a corridor with six doors each leading to a room that exercised a feature of a VE–person interaction. Results suggested that the greater the degree of visual dominance, the higher the sense of presence, whereas those who preferred the auditory representational system experienced the lower presence. The use of a kinesthetic system correlated with high presence in the experimental group but correlated negatively in the control group. The level of presence also increased when the subjects preferred the first perceptual position.

In a similar study Slater, Usoh and Steed (32) exposed 24 subjects to a VE and found the same results encountered in the previous study; participants who preferred the visual or the kinesthetic representational system (in this case if a virtual body was included) experienced the higher levels of presence. Finally, in a study performed with 8 participants, Slater, Usoh and Chrysanthou (33) found that including dynamic shadows in a virtual environment only derived in a higher degree of presence in those individuals who preferred the visual representational system.

In 1998 Witmer and Singer (34) developed the ITQ (Immersive Tendencies Questionnaire) to measure the tendencies of individuals to become immersed in a virtual environment. The ITQ is composed by three subscales: Involvement (propensity to get involved passively in some activity like reading books), Focus (ability to concentrate on enjoyable activities, and ability to block out distractions) and Games (frequency with which the subject plays games and the level of involvement in these games). The authors encountered a significant correlation between the ITQ and the sense of presence measured through the Presence Questionnaire (PQ). In the same year, Bangay and Preston (35) found in a study with 355 participants that subjects between ages 35 and 45 tended to provide lower scores in presence than the participants between ages 10 and 20.

One year later, in the University of Valencia, Baños (36) found that absorption (defined as the tendency to become involved or immersed in everyday events or the tendency to totally immerse oneself with the attentional objects) was positively correlated with presence in virtual environments. This study found that the individuals who experienced more anxiety during the exposure, had higher scores on presence.

Corina Sas (37), in a later study, performed with 15 students, investigated the relationship between four cognitive factors (absorption, creative imagination, empathy and cognitive style), and the sense of presence experienced by the users in a non immersive virtual reality system. The results show a significant correlation between presence score and creative imagination on the one hand, and presence and empathy, on the other hand. Presence also correlates highly, but not significantly, with the absorption scale. Due to the limited size of the sample, the results that concern the cognitive style are of limited value, but it seems that participants with perceiving or feeling type experienced a higher degree of presence. Furthermore, no differences were found in the sense of presence in function of the participant's gender. One year later, Sas (38) studied the effect of cognitive styles (measured through the Myers-Briggs Type Indicator) upon the sense of presence. In the line of the anterior study, Sas encountered that subjects who scored higher in feeling type or sensitive type, experienced a higher level of presence. Also, without being significant, introverted individuals tended to experience higher presence.

Finally, in a study conducted in 2005, Schuemie (39) exposed 41 participants to a virtual environment that contain height situations. No correlation between presence and absorption, gender, computer experience or the level of acrophobia was found. Despite this, a positive correlation was found between presence and age.

This study was conducted to evaluate the impact of five individual characteristics (personality, spatial intelligence, degree of test anxiety, cognitive style and experience in the use of computers) upon the sense of presence. Schubert's (29) spatial-functional model suggests that two cognitive processes are responsible of the engagement of the sense of presence. These are representations of bodily actions as possible actions in the virtual environments, and the suppression of antagonistic sensory information. This model claims that users need to construct a mental model of the virtual space, where the location of the own body is constructed as being contained in the space rather than looking at it from outside. Once users have developed such model they are able to play an active part and take control over their actions. Although becoming immersed in a virtual environment leads to a greater sense of presence, as pointed by Schubert (29), users need to perceive that they are capable of taking on the role they are governing within the virtual environment. Users who place themselves in the virtual space by navigating and interacting with the objects are more likely to experience presence as they mentally remove themselves from the real world to the virtual world. Thus it is expected that higher levels of spatial intelligence and more computer experience will facilitate the navigation and interaction within the virtual world, and, consequently, lead to a greater sense of presence.

In order to construct the mental representation of the virtual space, users have to suppress conflicting sensory inputs such as the stimuli of the hardware or the stimuli of the real world. The suppression of conflicting stimuli and the allocation of attention to the virtual stimuli can lead to the engagement of the sense of presence (29, 34). Introverts have been suggested to have a narrower focus of attention than extraverts. This narrower range of attention leads to a less extensive processing of the stimuli not related to the primary task, and, consequently, task-irrelevant or distracting information should be more easily ignored (40). Hence, due to the greater capacity of introverts to allow their attention to the main task, it may be suspected that the most introverted subjects should experience higher levels of presence in virtual environments.

Although several authors state that two of the main dimensions of presence are the sense of physical space and the selective attention to the virtual environment (2,29,34,41), this is the first attempt to investigate the influence of spatial intelligence on presence. Furthermore, to our knowledge, this is the first study utilizing an immersive virtual reality system to investigate the relationship between personality characteristics and presence. Other studies, mentioned before (37, 38), used non immersive virtual reality systems.

Cognitive style is another factor that can influence the sense of presence. Cognitive style refers to an individual preferred and habitual approach to organizing and representing information (42). One approach is the distinction between verbalizers and visualizers. Visualizers use images as a form of thinking, whereas verbalizers operate mostly in a world of words and verbal thoughts, ideas and structures. Thus, visualizers prefer visual information, and verbalizers prefer verbal or written information. Virtual reality environments offers mainly visual information to the participants, with a lower auditory stimulation, for this reason we formed the hypothesis that visualizers are likely to experience a higher degree of presence than verbalizers. Furthermore, people whose cognitive style is mainly visual, probably can construct a more accurate mental model of the virtual environments, this mental model, according to the spatial-functional model of Scubert (29), is necessary to experience presence.

Finally, the relationship between emotion and presence is investigated. Huang & Alessi (43) pointed out that various mental health conditions, such as depression, anxiety, or psychotic disorders, are likely to influence the sense of presence, since they are known to have a clear effect on how people experience the world around them. We consider that presence is influenced by emotions. In mental health applications of virtual reality, numerous have demonstrated that emotions are especially important in order to generate and enhance the sense of presence (44,45,46,47). From this line of thinking, in this study it may be expected that high test anxious students will obtain higher levels of presence than low test anxiety students.

This study is part of a broader reseach in which the exposure to virtual environments will be used to evaluate and treat test anxiety in students. In a first stage of this project a study was conducted to explore the effectiveness of virtual environments in producing emotionally significant responses in students with high degrees of test anxiety. This study concluded that the virtual environments were able to provoke higher levels of subjective and state anxiety, and higher levels of depressed mood in high test anxiety students than in low test anxiety students. Later, a pilot study (48) showed benefits in the treatment of test anxiety by using the technique of exposure to these virtual environments, obtaining a reduction in the levels of test anxiety of the participants and an increase in their academic performance. These students also diminished their ratings of avoidance to exams.

2. Method

2.1. Subjects

The initial sample, recruited via an on-line course on test anxiety, comprised 306 university students. The Test Anxiety Inventory (49) was administered to assess subjects' degree of test anxiety. Students who presented extremely high or extremely low scores were contacted. Those with scores in or above the 75th percentile (direct punctuation > 55) on the TAI were provisionally recruited for the high test anxiety group and students with scores in or below the 25th (direct punctuation < 36) for the low test anxiety group.

Finally, 26 students agreed to take part in the study, 16 with high test anxiety and 10 with low test anxiety. Twenty-two were women (84.6%) and four were men (15.4%), with a mean age of 22.85 years (S.d.: 3.21 range 18 – 34). The high test anxiety group comprised 16 women with a mean age of 23.06 years (S.d..: 3.45), the low anxiety group comprised six women and four men with a mean age of 22.5 years (S.d..: 2.9).

2.2. Instruments

Hardware:

The virtual environments were developed on a Pentium IV, 2 GHz, Windows 2000, 768 Mb RAM, 60 Gb hard disk, 19'' monitor, Hercules 3D Prophet 9700 PRO graphics cards with 128 MB DDR and AGP 8X. An *I-visor* DH-4400VP virtual personal display was used with a resolution of 800 X 600 pixels and a visual field in diagonal of 31º, connected to a *Tracker* Intersense 3-DOF (*degrees of freedom*) which measured the position and movement of the head.

Software:

To develop the virtual environments, tools of two kinds were used:

- Modeling and animation tools: the scenarios, virtual elements and animated 3D objects were constructed with

3D Studio Max 6. The Poser 4 program was used to design the characters, which were animated with Character Studio 4.0. Adobe Photoshop 6.0 was used to create the textures and images.

- Interactive development applications: Virtools Dev 2.5 was used to combine the objects and characters created with the different graphic design tools, and to integrate them with textures and sound. It was also used to make the environments interactive and to facilitate browsing.

Virtual scenarios:

The virtual environments were prepared in chronological order: the student's home, representing the day before and the morning of the examination, then the metro, and finally, the corridor and lecture-hall where the examination takes place. The situations and elements that comprise the environments were selected on the basis of a survey administered following a procedure that will be described later. Also a training room was created in which the students can familiarize with the technology.

-Training room:

This scenario represents a room composed by different elements like tables, chairs, a sofa, switches, and so on. In this room the students can learn to navigate in the environment, use the head tracking to look at the different directions, and interact with the objects (sit, switch on the light, etcetera).

- Home:

The scenario includes a flat, with a bedroom, (figure 1), a corridor, bathroom, dining-room (figure 2), kitchen and hall. The first scene shows the student's bedroom at 11 o'clock on the night before the examination. In the room there is a desk with a textbook, and there are signs reminding him/her that there is an examination the next day. To increase the level of presence and to provoke the same emotional and cognitive reactions as in real situations, the students are able to carry out the same actions as s/he would carry out on the day before a real examination: s/he can turn the lights on and off, open the windows, put on music, lie down on the bed, eat or drink, study, go to the bathroom, brush their teeth, have a shower, and so on. There are also clocks all over the house so that the student knows how much time there is left to study, or can decide when to go to bed.

This scenario is also used to represent the start of the examination day. The alarm clock rings at 7.30 am. As in the previous scenario, the students do all the things they would normally do; in addition, they now dress, prepare the belongings that they will take to the university, have breakfast, and so on.

Figure 1. View of the bedroom

Figure 2. View of the dining-room

- Metro:

This scenario represents part of the Barcelona underground system (figure 3). The initial view shows the station entrance. Ahead of the student are the steps leading to the platform. Once there, the student hears the conversations of groups of other students waiting for the train. After a minute's wait the train arrives and the student gets on and sits down (figure 4). During the journey, which lasts three stops, the student can study while other students talk about the examination they are about to take.

Figure 3. The metro station

Figure 4. Inside the metro.

- University:

There are two scenarios at the university. In the first (figure 5) the student is waiting in the hallway, outside the lecture-room where the examination will take place. During the wait, s/he is surrounded by other students talking about the subjects, the examination, how they have prepared for it, and so on. After five minutes the lecturer arrives with the examinations and tells the students they can go in. The second scenario presents the lecture-room where the examination will take place (figure 6). The student is now seated and waits as the lecturer hands out the examinations. After the lecturer's instructions, the examination appears on the student's desk. Students have to answer 25 general knowledge questions. The format is multi-choice, with four possible answers for each question.

Figure 5. The hallway in the university

Figure 6. Inside the examination room

- <u>Evaluation:</u>

• TAI (Test Anxiety Inventory) (49).

A self-report questionnaire designed to measure test anxiety as a situation-specific personality trait. The questionnaire comprises 20 items in which the student must indicate how often they experience the symptoms of anxiety before, during and after the examinations, on a 1 to 4 point Likert scale (1= hardly ever; 4 = almost always). The TAI contains two sub-scales, of eight items each, which assess *worry* (cognitive aspects) and *emotionality* (physiological aspects).

• EPQ-RS (Eysenck Personality Questionnaire Short Revised version) (50).

A self-report questionnaire designed to measure the personality characteristics. The EPQ-RS consists of 48 items, each answered on a yes-no basis, that asses the Eysenckian dimensions of extroversion, neuroticism and psychoticism. It contains three sub-scales: Extraversion, Neuroticism, Psychoticism and Social Conformity or Lie.

• Solid Figures Rotation (51).

It is a 21 item self-applied instrument designed to measure the aptitude to recognize and interpret objects in the space. In each item five different solid figures are presented. Each figure displays a three dimensional solid block. The person must decide which figure matches a given model figure seen from another perspective.

• IPQ (Igroup Presence Questionnaire) (29).

A self-report questionnaire designed to measure the sense of presence in virtual reality environments. It comprises 14 items rated on a seven point Likert Scale. IPQ contains three subscales, each one of them composed by four items, which asses Involvement (the awareness devoted to the VE), Spatial Presence (the relation between the VE and the own body), and Realness (the sense of reality attributed to the VE). Also contains one item that asses the "sense of being here".

• VVQ (Verbalizer-Visualizer Questionnaire) (52).

The VVQ is the most used measure of the relative reliance on verbal and visual code in habitual modes of thinking. It is a self-administered questionnaire constituted by 15 true-false items. The results raise a single value, the higher scores indicate visual preference while low scores are indicative of a verbal preference.

- CO (Computer experience) (53).

This instrument asseses subjects' experience with 3D games and computers. It consists in a 5 item scale rated from 1 to 5, where 1 = *very bad/never* and 5 = very good/often.

2.3 Procedure

To obtain information on the elements needed to make the environments clinically significant, we asked to a sample of 240 undergraduate students of the University of Barcelona which examination-related situations and thoughts generated the higher levels of anxiety (48). After analysis of their responses, 22 specific categories related to specific situations were established. The most frequent ones were: the comments of classmates, studying the day before the exam, bed time, waiting in the hallway, the morning of the examination, sitting in the examination room, the day before the examination, and so on. All these situations were incorporated in the three previously designed environments. Furthermore, nine categories related to anxiety thoughts were obtained, in this case the most frequent ones were the negative evaluate of owns capacities, perfectionism, worry about extern factors, worry about the negative consequences of failure, etcetera. These thoughts were inducted to the students through the conversations maintained by the virtual students in the virtual environments

In the present study, the virtual environments were presented to the two groups of students (high and low test anxiety). Exposure to the virtual environments was individual. Subjects visited all the environments in a single session (the mean duration of the sessions, including exposure to the environments and administration of the questionnaires, comprised 120 minutes). The procedure was double blind, that is, the researcher who administered the environments was unaware of the subject's TAI score, and students did not know their score or the aim of the research; they were told only that the study was designed to obtain information on students' behaviour in exam situations, in order to prepare a treatment program. Before starting the session, the participants were told that they would be shown a series of virtual environments simulating what students go through before and during an examination, starting with the previous evening and finishing with the examination itself. They were told that the exam consisted of a general knowledge test, which would be graded. They were asked to act as they would normally prior to and during an examination and they were told what it involved, and what tasks they could perform. Before starting the exposure to test anxiety environments, each participant was exposed to the training room with the objective that can familiarize with the virtual worlds and the virtual reality technology, and was administered the EPQ-RS, the VVQ, the Solid Figures Rotation and the CO questionnaire. After seeing each test anxiety environment each participant was administered the IPQ questionnaire.

3. Results

Multiple regression analysis were conducted to fit linear models relating the dependent variable to the independent variables. The EPQ-RS, VVQ, Solid Figures Rotation, TAI, and CO questionnaire scores were included as predictors in a linear regression with the IPQ scores in the house, metro, university, and the average score of the IPQ obtained in the three virtual environments as dependent variables. The method used in all cases was step wise.

In the regression model performed to predict the score of the IPQ in the house, only the score obtained in the Solid Figures Rotation was included in the regression equation (R=0.503, p=0.009), explaining the 25.3% of the presence's variance. As shown in table 1, the correlation between these variables was significant (p=0.004). The analysis also showed a significant negative correlation between IPQ and the extraversion sub-scale of the EPQ-RS (p=0.046). Also, no significant correlation was found between the student's test anxiety and the IPQ scores (p = 0.125). These results indicated that the higher scores in spatial intelligence lead to experience a higher degree of presence. Also, the most introverted students tend to feel more present in the virtual house than the extroverted.

In the second virtual environment (the metro) only the score obtained in the TAI questionnaire (R=0.475, p=0.016) was included in the regression equation, this variable explains the 22.6% of the variance of the IPQ, also the correlation between these variables was found significant (p=0.008). As in the anterior environment, extraversion correlated negatively with the score obtained in the IPQ (p=0.024). No more significant correlations were found in this environment (table 1), although the analysis also showed a small trend (p<0.10) toward a positive correlation between the psychoticism sub-scale of EPQ-RS and the score obtained in the IPQ questionnaire. These results show that individuals who scored higher in test anxiety experience the highest degree of presence. Also, introverts tend to obtain higher scores on presence.

In the regression analysis performed to predict the IPQ scores in the virtual university, none of the predictors was included in the regression equation. Despite this, the results showed two marginally significant correlations (table 1). On the one hand, the IPQ score correlate positively with the Solid Figures Rotation (p=0.050), and, on the other hand, the IPQ correlated positively with the level of test anxiety of students (p=0.061). Thus, in this environment, the individuals with highest scores on TAI, but mainly on the Spatial Figures Rotation, had a tendency to obtain higher scores on presence.

Finally, a composed measure of presence in the three environments was calculated as the average of the IPQ scores obtained in the house, the metro and the university. None of the predictor variables was included in the regression equation, although significant positive correlations between presence and the scores obtained in the Solid Figures Rotation (p=0.033) and TAI (p=0.028) were found. Also, a negative significant correlation was found between the IPQ scores and the extraversion sub-scale of the EPQ-RS (p=0.043). These results indicate that the participants tend to feel more present in the environments if they are introverted or if they have a high degree of test anxiety or spatial intelligence.

Table 1. Pearson's correlation coefficients (and significance) between the scores of IPQ and EPQ, Solid figures rotation, VVQ ,CO and TAI.

	IPQ House	**IPQ Metro**	**IPQ University**	**IPQ Total**
EPQE	-0.338 (0.046)	-0.400 (0.024)	-0.216 (0.144)	-0.350 (0.043)
EPQN	0.089 (0.333)	0.228 (0,136)	0.091 (0.329)	0.156 (0.229)
EPQP	0.238 (0.121)	0.293 (0.077)	0.073 (0.362)	0.224 (0.140)
EPQD	0.003 (0.494)	0.106 (0.307)	-0.083 (0.344)	0.014 (0.473)
Solid Figures Rotation	0.503 (0.004)	0.220 (0.146)	0.330 (0.050)	0.373 (0.033)
VVQ	-0.044 (0.415)	0.046 (0.413)	0.095 (0.321)	0.035 (0.434)
CO	-0.218 (0.142)	-0.003 (0.495)	-0.129 (0.264)	-0.131 (0.434)
TAI	0.233 (0.125)	0.475 (0.008)	0.312 (0.061)	0.386 (0.028)

4. Discussion

Virtual reality therapy is based on the assumption that people feel present in the virtual environment. Despite individual differences can moderate presence, little research has been conducted in this line. Research into these individual moderating traits will be of value because it may enhance the number of patients that can benefit of virtual reality therapy and can help to understand why some patients don't respond to this form of therapy. Exploring these variables can help to explain why the 20% of the patients treated by Max North (28) showed little or no reduction in agoraphobic symptoms, or why half of the participants of the study conducted by Walshe (54) didn't feel present in virtual reality environments. These studies suggest that it doesn't exist a direct relationship between the sense of presence and the media characteristics. In that case, all subjects should respond in the same way when confronted with the same virtual reality system, whereas, in fact, usually, identical media form and content produce different degrees of presence in different individuals.

The results of this study suggest that test anxiety, spatial intelligence and extraversion have an important influence in the level of presence experienced by students exposed to virtual environments in order to treat their test anxiety problem, and neither verbalizer-visualizer cognitive style nor the experience with computers have a significant impact on it.

The individuals with higher spatial intelligence tend to feel more present in the virtual environments. For a better understanding of this relationship, we must take into account the sense of presence experienced in each particular environment. The results suggest that in the university, and mainly in the house, there is a high relationship between spatial intelligence and presence, but no relationship was found in the metro.

Each environment requires a different degree of interaction. The house was designed as the environment with the highest interaction degree because the participant can navigate freely and at his own pace through the different rooms, and she can interact with a great number of elements. In the virtual university the navigation is restricted; the students can only navigate freely in the hallway while they are waiting for the lecturer to arrive. Despite this, the environment requires a high degree of interaction to respond to the exam. Finally, the metro is the environment with fewest interaction opportunities (the student can only pick up his/her briefcase for study) and with less free navigation (during the journey the student is sitting). This is the first study that evaluates the influence of spatial intelligence on the sense of presence, and it seems to play an important role in interaction with the characteristics of the environments.

Regarding the personality characteristics of the users, introverts tend to experience a higher degree of presence. This relationship could be explained by the fact that introverted people are more able to select relevant information from the stimuli they are exposed to, Thus, introverted participants probably were more able to suppress the distracting stimuli and focus their attention to the virtual environments, increasing the sense of presence (29,34). Furthermore, as suggested by Gutierrez-Maldonado et al., it must be pointed that introverted people, due to their tendency to reflection and their low impulsivity, are more comfortable when interacting in a computer mediated environment, where they can control the rhythm of the interaction (55). In the virtual environments they can navigate and interact at their own pace, so they feel more comfortable, and this can lead to a higher degree of presence.

This relationship was found in the house and in the metro, not in the university. These results are consistent with those found by Sas (38), who found that, in a non immersive virtual reality system, individuals who are more introverted are likely to experience higher degrees of presence (without being statistically significant). Thus, it seems that independently of the degree of immersion, introverts tend to score high in presence.

Regarding test anxiety, the results point towards a correlation with presence in every environment, excluding the virtual home. A very strong relationship is found in the virtual metro. The high correlation obtained in the virtual metro can be due to the fact that some of the high test anxiety students had some degree of specific anxiety to the metro too. This can lead to an enhanced sense of presence of these students, and provoke, in part, the high correlation between test anxiety and presence found in this environment.

The environments represent habitual situations of exam (the participants can study, do an exam, etc.), situations that are experienced by the students since the elementary school. For these reason, it can be considered that these environments represent a meaningful situation for them. Virtual Reality needs personal relevance to achieve involvement and high presence (56). Probably the meaningfulness and personal relevance of test anxiety environments for students can lead to experience high levels of presence in most of them. Furthermore, an exam situation typically causes some degree of anxiety in students, and represents a stressful situation. This emotional activation, even higher in test anxiety students (48), can lead to experience a high degree of presence in most of them. Thus, the meaningfulness and the emotional activation of test anxiety environment can moderate the effect of the degree of test anxiety upon the sense of presence. These results suggest that there is a relationship between presence and emotions.

No relation was found between the verbalizer-visualizer cognitive style of the students and the sense of presence in none of the virtual environments. These results are discordant to those found by Slater (31, 32) who encountered that people whose preferred representational system was the visual one were likely to have a higher degree of presence than those whose primary system was auditory or kinaesthetic. Probably the differences between these studies are due to discrepancies in the media content of the environments. In the studies conducted by Slater, the virtual environments offered primarily visual stimulation. In our study the auditory stimuli are important (there are virtual students maintaining conversations), and also the students can read their notes in each environment. The individuals whose main representational system is the auditory prefer to process the information through words, and they tend to like listening and reading. For these reason, the results of these studies aren't contradictory, because they offer evidence that the sense of presence is influenced by both user's traits and media content. The differences between studies can also been motivated by the measure utilized to evaluate cognitive styles.

Finally, no differences were found in the sense of presence in function of the experience with 3D games and the use of computers. These results are similar to those found by Schuemie (39), who didn't found relationship between the computer experience and the usability of the virtual environments. Probably spatial intelligence is more important than computer experience in order to construct a mental representation of the virtual world, and facilitate the navigation and interaction through the them.

5. Conclusions

The aim of this study was to investigate the human factors that can lead to experience a high degree of presence. This study is the first one to investigate the relationship between spatial intelligence and sense of presence, it's also pioneer using an immersive virtual reality system to study the influence of the personality characteristics upon the sense of presence. It was also investigated the influence of the degree of test anxiety, cognitive style and the computer experience of the subjects on the sense of presence. Results suggest that the individuals with high spatial intelligence, introversion and test anxiety tend to experience a higher degree of presence. It's interesting to point that higher levels of spatial intelligence are necessary to experience presence in the environments with higher degrees of navigation and/or interaction. Probably the influence of the degree of test anxiety on the sense of presence is mediated by the meaningfulness and emotional activation that the environments are able to produce; to check this hypothesis in future studies it will be interesting to incorporate state-anxiety measurements. It should be noted that the sample used in this study is relatively small and most of the participants were female. In future works it would be interesting to balance the gender of the participants and increase the sample size.

This line of research will contribute to understand the mechanisms that lead to the efficacy of virtual reality exposure psychological treatments, and can help to explain why different individuals can experience different levels of presence when being confronted with the same virtual environments (1).

6. References

[1] J. Steuer. Defining virtual reality. *Journal of Communication, 4(2)*, 73-93. 1992.

[2] M. Lombard, T. Ditton. At the heart of it all: the concept of presence. *Journal of Computer Communication*, 3(2). 1997.

[3] J.D. Prothero, H. G. Hoffman. Widening the field of view increases the sense of presence in inmersive virtual environments. *Technical Report* TR-95-2, Human Interface Technology Lab. 1995.

[4] A.F. Seay, D.M. Krum, L. Hodges, W. Ribarsky. Simulator sickness and presence in a high FOV virtual environment. *Proceedings of the IEEE Virtual Reality 2001 Conference*, March 13-17, 2001, Yokohama, Japan: IEEE Computer Society, pp. 299-300. 2001.

[5] W. Ijsselsteijn, H. de Ridder, J. Freeman, S.E. Avons, D. Bouwhuis. Effects of stereoscopic presentation, image motion, and screen size on subjective and objective corroborative measures of presence. *Presence, 10(3)*, 298-311. 2001.

[6] J.J-W. Lin, H.B.L. Duh, D.E. Parker, H. Abi-Rached, T.A. Furness. Effects of field of view on presence, enjoyment, memory, and simulator sickness in a virtual environment. *Proceedings of the IEEE Virtual Reality 2002 (VR'02)*.2002.

[7] J.D. Prothero, H.G. Hofmann, D.E. Parker, T.A. Furness, M.J. Wells. Foreground/background manipulations affect

presence. *Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting*. 1995

[8] W. Barfield, K.M. Baird, O.J. Bjornneseth. Presence in virtual environments as a function of type of input device and display update rate. *Displays, 19*, 91-98. 1998

[9] C. Hendrix, W. Barfield. Presence within virtual environments as a function of visual display parameters. *Presence, 4*, 274-289. 1996.

[10] J. Freeman, S.E. Avons, R. Medis, D. Pearson, W. Ijsselstein. Using behavioral realism to estimate presence: A study of the utility of postural responses to motion stimuli. *Presence, 9*, 149-164. 2000.

[11] R.B. Welch, T.T. Blackmon, A. Liu, B. Mellers, L.W. Stark. The effects of pictorial realism, delay of visual feedback, and observer interactivity on the subjective sense of presence. *Presence, 5*, 263-273. 1996.

[12] J. Hoffman, H. Bubb. Presence in industrial virtual environment applications- susceptibility and measurement reliability. In G. Riva, F. Davide, W.A. Ijsselsteijn (eds.): *Being There: Concepts, effects and measurement of user presence in synthetic environments*. IOS Press, Amsterdam, The Netherlands. 2003

[13] A. Axelsson, A. Abelin, I. Heldal, R. Schroeder, J. Wideström. Cubes in the cube: A comparision of collaboration in virtual and real environments. *Cyberpsychology and Behavior, 4(2)*, 279-286. 2001.

[14] M. Krijn, P.M.G. Emmelkamp, R. Biemond, C. De Wilde de Ligny, M.J. Schuemie, C.A.P.G. van der Mast. Treatment of acrophobia in virtual reality: The role of inmersion and presence. *Behaviour Research and Therapy, 42*, 229-239. 2004.

[15] C. Hendrix, W. Barfield. The sense of presence within auditory virtual environments. *Presence, 5*, 290-301. 1996.

[16] D. Västfjäll. The subjective sense of presence, emotion recognition, and experienced emotions in auditory virtual environments. *Cyberpsychology and Behavior, 6(2)*, 181-188. 2003.

[17] H. Hoffman, J. Groen, S. Rousseau, A. Hollander, W. Winn, M. Wells, T. Furness. Tactile augmentation: Enhancing presence in virtual reality with tactile feedback from real objects. Paper presented at the *meeting of the American Psychological Society*, San Francisco, CA. 1996.

[18] H. Q. Dinh, N. Walker, L. Hodges. Evaluating the importance of multi-sensory input on memory and the sense of presence in virtual environments. *Proceedings of the IEEE Virtual Reality 1999 Conference*, Houston, 222-228. March 1999.

[19] W. Barfield, C. Hendrix, K. Bystrom. Visualizing the structure of virtual objects using head tracked stereoscopic displays. *Virtual reality annual international symposium (VRAIS'97)*, p.194. 1997.

[20] P. Larsson, D. Västfjäll, M. Kleiner. The actor-observer effect in virtual reality presentations. *Cyberpsychology and Behavior, 4(2)*, 239-246. 2001.

[21] M. Slater, A. Steed, J. McCarthy, F. Maringelli. The influence of body movement on subjective presence in virtual environments. *Human Factors, 40*, 469-477. 1998.

[22] H. Hoffman, J. Prothero, M. Wells, J. Groen. Virtual chess: Meaning enhances user' sense of presence in virtual environments. *International Journal of Human-Computer Interaction, 10(3)*, 251-263. 1998.

[23] M. Alcañiz, R. Baños, C. Botella, B. Rey. The EMMA Project: Emotions as a determinant of presence. *Psychnology, 1(2)*, 141-150. 2003.

[24] R. Baños, C. Botella, V. Liaño, M. , B. Guerrero, B. Rey, M. Alcañiz. Sense of presence in emocional virtual environments. *Proceedings of the Seventh Annual Internacional Workshop Presence*, Valencia, October 2004.

[25] R.M. Baños, C. Botella, M. Alcañiz, V. Liaño, B. Guerrero, B. Rey. Immersion and Emotion: Their impact on the sense of presence. *Cyberpsychology and Behavior, 7(6)*, 2004.

[26] M. Slater, S. Wilbur. A framework for inmersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. *Presence, 6(6)*, 603-616. 1997.

[27] W. Barfield, D. Zelter, T.B. Sheridan et al. Presence and performance within virtual environments. In: W. Barfield, T.A. Furness (eds.), *Virtual environments and advanced interface design*. Oxford: Oxford University Press. 1995.

[28] M.M. North, S.M. North, J.R. Coble. *Virtual reality therapy. An innovative paradigm*. Colorado Springs: IPI Press. 1996.

[29] T. Schubert, F. Friedmann, H. Regenbrecht. The experience of presence: Factor analytic insights. *Presence: Teleoperators and Virtual Environments, 10(3)*, 266-282. 2001.

[30] R.S. Kalawsky. The validity of presence as a reliable human performance metric in inmersive environments. Presented at *Presence 2000: International Workshop on Presence*, Delft, Netherlands. 2000.

[31] M. Slater, M. Usoh. Representation systems, perceptual positions, and presence in inmersive virtual environments, *Presence,2*, 221-233. 1993.

[32] M. Slater, M. Usoh, A. Steed. Depth of presence in virtual environments, *Presence,3*, 130-144. 1994.

[33] M. Slater, M. Usoh, Y. Chrysanthou. The influence of dynamic shadows on presence in inmersive virtual environments. *Virtual Environments, 95*, 8-21. 1995.

[34] B.G. Witmer, M.J. Singer. Measuring presence in virtual environments: A presence questionnaire. *Presence,7*, 225-240. 1998.

[35] S. Bangay, L. Preston. An investigation into factors influencing inmersion in interactive virtual environments. In G. Riva, B.K. Wiederhold, E. Molinari (eds.), *Virtual environments in clinical psychology and neuroscience*. Amsterdam: IOS Press. 1998.

[36] R.M. Baños, C. Botella, A. García-Palacios, H. Villa, C. Perpiñá, M. Gallardo. Psychological variables and reality judgement in virtual environments: The roles of absorption and dissociation. *Cyberpsychology and Behavior, 2(2)*, 143-148. 1999.

[37] C. Sas, G. O'Hare. The presence equation: An investigation into cognitive factors underlying presence. *Presence: Teleoperators and Virtual Environments, 12(5)*, 523-537. 2003.

[38] C. Sas, G.M.P. O'Hare, R. Reilly. Presence and task performance: an approach in the light of cognitive style. *Cognition, Technology and Work, 6(1)*, 53-56. 2004.

[39] M.J. Schuemie, B. Abel, C.A.P.G. van der Mast, M. Krijn. The effect of locomotion technique on presence, fear and usability in a virtual environment. *Proceedings of Euromedia 2005,* Toulouse, April 2005.

[40] M. Althaus, H.K. Gomarus, A.A. Wijers, L.J.M. Mulder, J.L. van Velzen, R.B. Minderaa. Cortical and autonomic correlates of visual selective attention in introverted and extraverted children. *Journal of Psychophysiology, 19(1), 35-49*. 2005.

[41] J. Lessiter, J. Freeman, E. Keogh, J. Davidoff. A cross-media presence questionnaire: The ITC-Sense of Presence Inventory. *Presence, 10(3)*, 282-297. 2001.

[42] R. Riding, S. Rayner. *Cognitive styles and learning strategies: Understanding style differences in learning and behavior*. London: David Fulton. 1998.

[43] M.P. Huang, N.E. Alessi. Mental health implications for presence. *Cyberpsychology and Behavior, 2*, 15-18. 1999.

[44] R.M. Baños, C. Botella, A. García-Palacios, H. Villa, C. Perpiñá, M. Alcañiz. Presence and reality judgement in virtual environments: A unitary construct?. *Cyberpsychology and Behavior, 3(3)*, 327-335. 2000.

[45] L. Hodges, B.O. Rothbaum, R. Cooper, D. Opdyke, T. Meyer, J.J. de Graph, J.S. Willford. Presence as the defining factor in a VR application. *Technical Reports* GIT-GVU. 94-5. Georgia Institute of Technology. 1998.

[46] H.T. Regenbrecht, T.W. Schubert, F. Friedmann. Measuring the sense of presence and its relation to fear of heights in virtual environments. *International Journal of Human-Computer Interaction, 10(3)*, 233-249. 1998.

[47] M. Slater, D. Pertaub, A. Steed. Public speaking in Virtual Reality: Facing an audience of avatars. *IEEE Computer Graphics and Applications, 19(2)*. 1999.

[48] C. Carvallo, I. Alsina, J. Gutiérrez Maldonado. Tratamiento de la ansiedad ante los exámenes mediante exposición a entornos de realidad virtual. Paper presented at the *V Congreso Internacional de la Sociedad Española para el estudio de la ansiedad y el estrés*, Benidorm, Spain, September 2004.

[49] C.D. Spielberger. *Test Anxiety Inventory*. Spanish translation by A. Cano-Vindel.Unpublished manuscript. 1980.

[50] H.J. Eysenck, S.B.G. Eysenck. Eysenck. *Personality Questionnaire-Revised Short Scale*. Spanish adaptation by TEA Ediciones. Madrid. 2001

[51] M. Yela. *Solid Figures Rotaion*. Spanish adaptation by TEA Ediciones. Madrid. 1968.

[52] A. Richardson. *Verbalizer-Visualizer: a cognitive style dimension*. Journal of Mental Imagery, 1, 109-126. 1977.

[53] M.J. Schuemie. *Human-Computer Interaction and presence in virtual reality exposure therapy*. Ph.D dissertation, Delft University of Technology. 2003.

[54] D.G. Walshe, E.J. Lewis, S.I. Kim, K. O’Sullivan, B.K. Wiederhold. Exploring the use of computer games and virtual reality in exposure therapy for fear of driving following a motor vehicle accident. *Cyberpsychology and Behavior*, 6(3), 329-334. 2003.

[55] J. Gutiérrez Maldonado, M. Mora., S. García, P. Edipo. Personalidad, sexo y comunicación mediada por ordenador. *Anuario de Psicología, 32 (2)*, 51-62. 2001.

[56] J.F. Hoorn, E. Konijn, G.C. Van der Veer. Virtual reality: Do not augment realism, augment relevante. *UPGRADE-The European Online Magazine for the IT Professional, http://www.upgrade-cepis.org, IV(1)*, ISSN 1684-5285, pp 18-26. 2003.

The Impact of Personality Factors on the Experience of Spatial Presence

Ana Sacau, Jari Laarni, Niklas Ravaja, Tilo Hartmann
University Fernando Pessoa (Portugal), Helsinki School of Economics (Finland), Helsinki School of Economics (Finland), Hannover University of Music and Drama (Germany)
pssacau@ufp.pt, jari.laarni@hkkk.fi, niklas.ravaja@hkkk.fi, Tilo.Hartmann@ijk.hmt-hannover.de

Abstract

The aim of this paper is to present an experimental investigation conducted to assess the impact of personality factors on the formation of Spatial Presence. Four different media -linear text, hypertext, film and virtual environment- and eight different experimental manipulations were applied. Spatial Presence was measured using the MEC-SPQ questionnaire. Personality was measured using the NEO-FFI personality questionnaire. Other personality traits such absorption, domain-specific interest and spatial visual imagery were also measured. Our findings suggest that absorption, domain-specific interest and agreeableness are good predictors of Spatial Presence. Experimental manipulations, however, had a quite small effect on Spatial Presence.

Keywords---**Presence, Spatial Presence, personality factors, traits, Big Five**

1. Introduction

Sensations of nonmediation have received a growing attention of researchers in the last decades. Among all phenomena of non-mediation, one of the most prominent concepts probably is the construct of “Spatial Presence” (also called “Telepresence”, [1], or “Virtual Presence”, [2]).

Spatial Presence can be defined as the subjective experience of a user or onlooker to be physically located in a mediated space (although it is just an illusion; for reviews, see [3]; [4]; [5]; [6]; [7]). It has been described as “a sense of being there” that occurs “when part or all of a person’s perception fails to accurately acknowledge the role of technology that makes it appear that s/he is in a physical location and environment different from her/his actual location and environment in the physical world” [8].

As a subjective experience, Spatial Presence is supposed to be heavily influenced by individual factors [9] [1], either situation-specific states or more enduring stable dispositions (traits). However, research on the influence of personality factors on the formation of Spatial Presence is rare [10]. As Sas and O’Hare ([11], p. 527) put it “[a] large amount of work has been carried out in the area of technological factors affecting presence. […] Comparatively, the amount of studies trying to delineate the associated human factors determinant on presence is significantly less”. Therefore, in this paper, we elaborate links between stable individual factors and Spatial Presence and report the results of an empirical study in order to illuminate Spatial Presence as a subjective phenomenon.

2. User factors and the formation of Spatial Presence

According to the two-level-model of Spatial Presence [6] [12], the sensation of Spatial Presence can be construed as a two-step process (see figure 1) that emerges from the interplay of media factors on the one side and user factors on the other side. In a first step, the model regards the formation of a spatial situation model (SSM), which is a subjective mental model of the perceived spatial (media) environment (e.g. the user mentally reconstructs the size, shape, depth, etc. of a depicted environment). The second level, in turn, regards the transition from an SSM to Spatial Presence (i.e. the user’s feeling to be situated in the mentally constructed spatial scenery).

The model argues that traits influence the formation process at both stages. On the first level, the construction of a SSM is supposed to be influenced by the user’s domain specific interest and spatial ability. A high *domain-specific interest* should lead to a controlled continuously persistent attention allocation onto the media stimulus. Thus, the perception and mental reconstruction of spatial cues provided by the media environment is triggered that in turn should ease the formation of a SSM. The model argues that missing ‘building blocks’ of spatial information can be derived in a top-down process from memory, and inserted into the construction of a SSM. The more ‘talented’ the user is in terms of adding and integrating spatial information into the mental model (‘*spatial visual imagery*’), the greater the probability of a well-defined SSM.

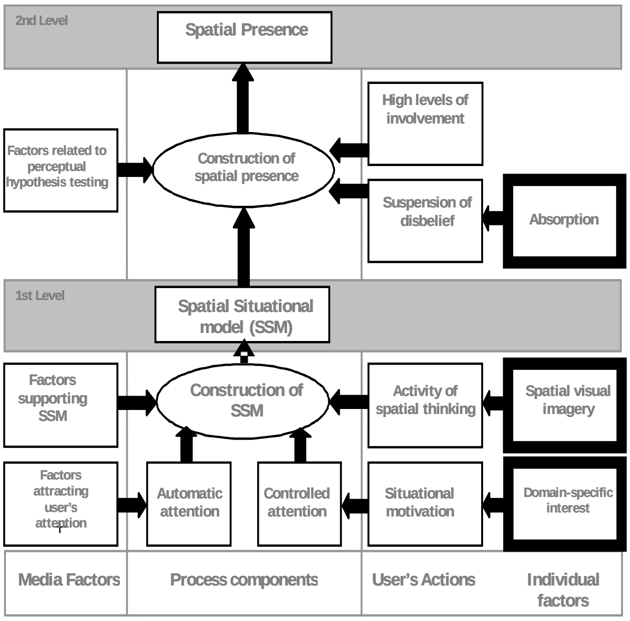


Figure 1. The two-level-model of Spatial Presence [traits are marked by thick boxes]

On the second level, the most important user trait influencing the formation of Spatial Presence is *absorption*. Trait absorption refers to an individual's motivation and skill in dealing with an object in an elaborate manner [13] than, in turn, cognitively 'detracts' the individual from other aspects of the environment. High-absorption individuals display tendencies to become intensely involved with objects (such as media products), and enter the condition of being 'fascinated' without much effort. Therefore it is argued that high-absorption individuals tend to pay more attention to the media stimulus, and also are prone to avoid critical elaborations or evaluations of the stimulus (i.e. to show disbelief).

In sum, the two-level model assumes that a high-domain-specific interest, a strong capability of spatial visual imagery, and a high trait absorption foster Spatial Presence experiences.

3. The Big Five personality traits

The five-factor or Big Five model of personality has emerged as the predominant model for specifying personality structure [14, 15]. This model posits that five broad dimensions—Neuroticism, Extraversion, Openness to Experience, Agreeableness, and Conscientiousness—adequately encompass, at a high level of generality, the full range of personality traits. The model derived foremost from studies of natural language of personality, and the most consistent support for the model comes from studies of rated trait adjectives [14]. Studies of self-report inventories have, however, been less consistent in identifying all five dimensions.

Individuals high on Neuroticism are characterized by a tendency to experience negative affect, such as anxiety, depression or sadness, hostility, and self-consciousness, as well as a tendency to be impulsive [for a review, see 14]. Those high on Extraversion tend to experience positive emotions and to be warm, gregarious, fun-loving, and assertive [14]. Those high on Openness to Experience are inclined to be curious, imaginative, creative, original, artistic, aesthetically sensitive, psychologically minded, and flexible [14]. Agreeableness refers to the tendency to be forgiving, kind, generous, trusting, sympathetic, compliant, altruistic, and trustworthy [14]. Finally, Conscientiousness refers to a tendency to be organized, efficient, reliable, self-disciplined, achievement-oriented, rational, and deliberate [14]. The Big Five personality traits have been associated with diverse outcomes such as stress vulnerability [16], coping [17], vigilance performance [18], and attention-deficit/hyperactivity disorder [19].

4. The Big Five and the formation of Spatial Presence

Given the unprecedented level of interest in personality research and practice enjoyed by the Big Five, it is warranted to focus on these five factors when examining the relationship of personality with Spatial Presence. Although this issue has not been examined before, there may be several potential links between the Big Five traits and the formation of Spatial Presence. Conceptual considerations and empirical evidence indicate that Openness to Experience is closely related to absorption [e.g., 15]. Thus, given the aforementioned important role played by absorption in the formation of Spatial Presence, it would be expected that Openness to Experience may contribute to suspension of disbelief and Spatial Presence experiences. Given the characteristics associated with Openness to Experience, such as being imaginative, creative, and artistic, it would be expected as being related to high visual spatial imagery, i.e., a user trait putatively contributing to Spatial Presence. Characteristics associated with Agreeableness, such as being trusting and compliant [14], might also contribute to Spatial Presence (perhaps through the mediating influence on suspension of disbelief).

Prior research has also shown that a depressed mood elicits self-focused attention and results in decreased involvement with media messages [20]. Given the tendency of high Neuroticism individuals to experience depression or sadness [14], they may show diminished involvement with (external) media stimuli, which may contribute to low Spatial Presence. Self-focused attention may also interfere with the construction of a SSM. Conscientiousness has, in turn, been associated with high perceptual sensitivity (i.e. an ability to detect changes in stimuli) [18]. That being so, it may potentially (a) be related to a critical evaluation of the stimulus (resulting in low Spatial Presence) or (b) contribute to the construction of a SSM through the mediating influence on controlled attention (resulting in high Spatial Presence).

5. The present study

Given the aforementioned considerations, the aim of the present study was to examine the relationship of the Big Five personality factors and the three personality traits included in the two-level-model of Spatial Presence (i.e., absorption, spatial visual imagery, and domain-specific interest) with Spatial Presence experiences during media processing. It was hypothesized all these personality-related factors would be associated with Spatial Presence.

It was also examined how different manipulations (e.g., large visual field vs. small visual field) influence Spatial Presence when processing different types of media (i.e., linear text, hypertext, film, virtual environment).

6. Method

6.1 Participants

The sample was made up of 240 undergraduate or graduate level students (138 females, 102 males) in four countries (Finland, Germany, Portugal and Switzerland). The mean age of the participants was 24.25 with a range between 18 and 41. They were ignorant of the purpose of the study before participating. Participants were paid for their participation (total value 10-13€).

6.2 Stimuli

Each participant was exposed to either linear text, virtual environment with hypertext interaction characteristics (from now Hypertext), film or virtual environment with 3D graphics (from now VE) media stimulus. Table 1 presents the distribution of participants according to stimulus type, location and gender.

In the linear-text experiment, participants read an extract from Ken Follett's book "The Pillars of the Earth" ("Die Säulen der Erde" in German). The 12-page episode portrays how one of the main characters intrudes a cathedral, sets fire and tries to escape from the flames.

		Gender of participant		
		Female	**Male**	**Total**
Type of stimulus	Linear Text (Hannover)	19	21	40
	Hypertext (Helsinki)	36	24	60
	Film (Porto)	30	30	60
	VE (Zürich)	53	27	80
Total		138	102	240

Table 1. Gender of participant by type of stimulus

In the hypertext condition, 'The Art of Singing' CD-ROM (Nothing Hill Publishing Limited 1996) based multimedia stimulus was applied (see Figure 2). It is a commercial 2-D virtual environment in which the user tours around a virtual academy of song. The academy consisted of three floors; on each floor there were several rooms in which different activities took place. The participants had no time to check all the possibilities of the academy ('navigation paths' were thus quite different), but they typically visited all the floors of the house. The field size (FoV) was either 20° or 60° in diameter.

Figure 2. A view to the first floor of the Academy of Song

In the film condition, the participant watched a film that was projected on a screen. The film stimulus was an extract from a Harry Potter and the Philosopher Stone (Warner Bros. Entertainment Inc. 2001). Sequence shows Harry Potter visiting the restricted area of the library at night, under an invisible cloak. The extract is 8 minutes and 39 seconds in length, remaining from 1:21:43 to 1:30:21. The extract was used in its Portuguese version without captions. The FoV was either 20° or 60° in diameter.

In the VE condition, a computer game was applied that was based on Doom 3's (Id Software Inc. 2004) 3D engine (see Figure 3). The user visited a Mayan temple consisting of 17 different rooms on three floors. In this condition the FoV was either 30° or 60°. In the film, hypertext and VE stimulus conditions, the stimuli were generated with a PC computer, and the image was projected on the screen by a beamer. Sounds were presented through high-quality headphones. A standard computer mouse was used for input in the hypertext and VE condition.

Figure 3. A view to the Mayan temple

6.3 Measures

After stimulus presentation, all the participants completed several questionnaires. The MEC Spatial Presence Questionnaire (MEC-SPQ) consists of several scales that measure different dimensions of Spatial Presence [21]. It includes four process factors [Attention Allocation, Spatial Situation Model (SSM), Self Location (SPSL) and Possible Actions (SPPA)], two variables relating to states and actions [High Cognitive Involvement and Suspension of Disbelief (SoD)] and three trait variables [Domain Specific Interest, Spatial Visual Imagery and Absorption]. Since our interest was to investigate the relationship between personality characteristics and Spatial Presence, only Self Location (SPSL) and Possible Actions (SPPA) and trait variables were considered here. Spatial Presence was the mean score of SPSL and SPPA scores.

NEO-FFI is a 60-element version of the NEO-PI-R. It generates scores on five personality factors, neuroticism, extraversion, openness, agreeableness and conscientiousness [22]. The format of the questionnaire is a five-point Likert scale, ranging from “Strongly disagree” to “Strongly agree”. The Finnish, German and Portuguese versions of the NEO-FFI were used. German NEO-FFI is validated. Finnish and Portuguese NEO-FFI are based on NEO-PI-R validated versions for Finland and Portugal, respectively. Table 2 shows descriptives and reliability indices for three trait scales from MEC-SPQ and five personality dimensions from NEO-FFI.

Two internal consistency estimates of reliability were computed for each trait MEC-SPQ scales and five personality NEO-FFI scales: coefficient alpha and split-half coefficient expressed as a Spearman-Brown corrected correlation. For the split-half coefficient, each scale was split into two equal length halves. In the splitting the items were selected by the SPSS. Spearman-Brown reliability indices indicate reasonable reliability with values between .66 and .88.

NEO-FFI scales	Mean	Std. deviation	Skewness	Alpha	Spearman-Brown
Neuroticism	2.833	.683	.079	0.696	0.868
Extraversion	3.569	.489	-.317	0.824	0.680
Openness	3.658	.552	-.186	0.689	0.763
Agreeableness	3.738	.438	-.083	0.619	0.663
Conscientiousness	3.538	.615	-.424	0.824	0.847
MEC-SPQ trait scales					
Domain-specific interest	2.427	.994	.627	0.921	0.876
Spatial visual imagery	3.763	.690	-.242	0.845	0.852
Absorption	3.611	.612	-.204	0.770	0.737

Table 2: Descriptives and reliability indices

6.4 Procedure

For all stimulus conditions, a 1 x 2 between-subjects design was used. In the linear text experiment, the experimental manipulation was the level of suspension of disbelief. Participants were sitting at the table and reading a text. Half of the participants were asked to read the text carefully and to look for mistakes. They were also told that, after reading, they have to report the mistakes. The other half was asked to read the text just as if they were reading a novel at home. They were told that, after reading, they would be asked how much they like the text. In the hypertext condition the participants were told to freely navigate through the environment. For the VE condition the participants’ task was to search for gold bars. For the film condition participants were told to see a film sequence.

The total duration of the presentation of each stimulus was 7-10 minutes. After the stimulus presentation, the participants were asked to fill out the above-mentioned questionnaires.

For all conditions, stimuli were presented in single experimental sessions.

Regression analysis was conducted to assess which personality factors were better predictors of Spatial Presence. Then, the General Linear Model Univariate procedure was applied in order to determine the interaction effects of personality predictors and experimental manipulations on Spatial Presence. The data of the four type of stimulus were analyzed separately. SPSS was used for all data analysis.

7. Result

7.1. Correlation between trait MEC-SPQ scales and NEO-FFI

Correlation coefficients were computed between three trait MEC-SPQ and five NEO-FFI scales. The results of the correlation analysis are presented in table 3. Results suggest that more extroverted participants scored higher on spatial visual imagery. Participants who scored higher on neuroticism and openness also showed higher levels of absorption. Finally, those individuals who scored higher on conscientiousness showed higher domain-specific interest.

	Domain-specific interest	Spatial visual imagery	Absorption
Extraversion	.113	.239**	-.032
Neuroticism	.073	.023	.238**
Openness	-.026	.075	.365**
Agreeableness	.100	.048	.052
Conscientiousness	.161*	.051	.004

** Correlation is significant at .01; * Correlation is significant at .05

Table 3. Correlations between trait MEC-SPQ scales and NEO-FFI scales

7.2. Personality traits and Spatial Presence

Spatial Presence is represented by the mean value of “Spatial location” and “Possible actions” scales of MEC-SPQ. Spatial Presence mean is 2.681 and standard deviation equal to .852.

Correlation coefficients were computed between Spatial Presence and each trait MEC-SPQ scales and five personality factors from NEO-FFI.. The results of the correlation analysis are shown in table 4.

	Spatial Presence
Domain-specific interest	.310**
Spatial visual imagery	.116
Absorption	.190**
Extraversion	.060
Neuroticism	.052
Openness	.011
Agreeableness	.169**
Conscientiousness	-.048

**. Correlation is significant at the 0.01

Table 4. Correlation between Spatial Presence and traits

Two multiple linear regression analysis using two unordered sets of predictors were conducted to predict Spatial Presence. One analysis included the five personality dimensions from NEO-FFI as predictors (Extraversion, Neuroticism, Agreeableness, Openness and Conscientiousness), while the second analysis included three trait factors of MEC-SPQ (Domain-specific interest, Absorption and Spatial visual imagery). The regression equation with trait factors of MEC-SPQ was significant $R^2 = .12$, adjusted $R^2 = .10$, F (3, 236) = 10.68, p = .000. However, the regression equation with five personality dimensions was not significant, $R^2 = .04$, adjusted $R^2 = .02$, F (5, 234) = 2.22, p = .053. Based on these results, trait factors of MEC-SPQ appear to be better predictors of Spatial Presence.

Next, a multiple regression analysis was conducted with all eight trait scales as predictors. The linear combination of the eight measures was significantly related with Spatial Presence, $R^2 = .15$, adjusted $R^2 = .12$, F (8, 231) = 5.23, p = .000. Trait factors from MEC-SPQ predicted significantly over and above the personality dimensions of NEO-FFI, R^2 change = .11, F (3, 231) = 9.83, p = .000, but the five personality dimensions of NEO-FFI did not predict significantly over and above the trait factors of MEC-SPQ, R^2 change = .03, F (5, 231) = 1.84, p = .105.

Standardized Beta coefficients showed three factors for the predicting equation: domain-specific interest (Beta = .28, t = 4.42, p = .000), agreeableness (Beta = .15, t = 2.34, p = .02) and absorption (Beta = .14, t = 2.04, p = .042). Regression model is graphically represented in figure 4.

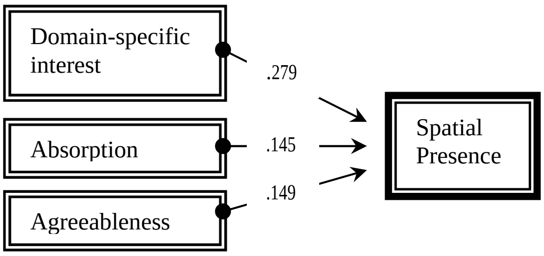


Figure 4: Regression model for Spatial Presence

Diagnoses of collinearity using tolerance coefficients showed that the scales are independent from each other, minimizing standard error (tolerance coefficients from .72 to .92).

7.3. Experimental manipulations and traits interactions on Spatial Presence

Four four-way analyses of variance (ANOVA) were conducted to analyze the effects of each four experimental manipulation on Spatial Presence. In any case, the dependent variable was the Spatial Presence and domain-specific interest, absorption, and agreeableness were recoded into three levels (low, medium and high) and included as factors.

Experimental manipulations for four type of stimulus are represented in table 5.

The first four-way ANOVA was conducted to evaluate the effects of three levels of domain-specific interest, absorption and agreeableness and two levels of text experimental manipulation (with vs. without SoD) on Spatial Presence. The ANOVA indicated no significant interactions between factors but significant main effects for text manipulation, F (1,25) = 4.34, p < .05, partial $\eta^2 = .15$. The text manipulation main effect indicated that participants who were encouraged to simply enjoy the text without checking it for errors (with SoD) reported higher levels of Spatial Presence than those who were encouraged to make a critic reading (without SoD).

Type of stimulus	Experimental condition	Spatial Presence mean	Standard deviation
Text	Without SoD	2.087	.74
	With SoD	2.400	.56
Hypertext	Small FoV	2.654	.65
	Large FoV	2.900	.68
Film	Small FoV	2.534	.91
	Large FoV	2.670	.93
VE	Small FoV	2.865	.95
	Large FoV	2.915	.91

Table 5: Spatial Presence mean values per experimental condition

The second four-way ANOVA was conducted to evaluate the effects of three levels of domain-specific interest, absorption and agreeableness and two levels of hypertext experimental manipulation (large FoV vs. small FoV) on Spatial Presence. The ANOVA indicated no significant interactions between factors but significant main effects for domain-specific interest, F (2,38) = 7.58, p < .005, partial η^2 = .28. Follow-up analysis to the main effect for domain-specific interest examined all pairwise comparisons among three levels of domain-specific interest. The Tukey HSD procedure was used to control for Type I error across pairwise comparisons. The results of the analysis (see table 6) indicate that the group with low domain-specific interest (DSI) reported significantly less sense of Spatial Presence than people who report medium and high levels of domain-specific interest. There was no significant difference between medium and high levels of domain-specific interest.

The third four-way ANOVA was conducted to evaluate the effects of three levels of domain-specific interest, absorption and agreeableness and two levels of film experimental manipulation (large FoV vs. small FoV) on Spatial Presence. The ANOVA indicated three significant main effects for domain-specific interest, F (2,38) = 6.42, p < .005, partial η^2 = .25, absorption, F (2,38) = 17.67, p < .001, partial η^2 = .48 and agreeableness, F (2,38) = 6.34, p < .05, partial η^2 = .14. The ANOVA also indicated two two-level and one three-level significant interactions between factors, domain-specific interest*absorption interaction, F (2,38) = 11.04, p < .001, partial η^2 = .37, domain-specific interest*agreeableness interaction, F (2,38) = 5.23, p < .05, partial η^2 = .22, and domain-specific interest*absorption*agreeableness interaction, F (2,38) = 6.39, p < .05, partial η^2 = .14. Because a three-level interaction was significant, we choose to ignore low-level interactions and main effects. Follow-up tests were conducted to evaluate the eighteen pairwise comparisons (3x3x2). The Tukey HSD procedure was used across pairwise comparisons. The results of the analysis indicate that the group with high domain-specific interest, high level of absorption and medium level of agreeableness reported significantly high sense of Spatial Presence (Mean = 4.75; S. Error = .47).The fourth four-way ANOVA was conducted to evaluate the effects of three levels of domain-specific interest, absorption and agreeableness and two levels of virtual environment experimental manipulation (large FoV vs. small FoV) on Spatial Presence. The ANOVA indicated no significant interactions between factors but significant main effects for domain-specific interest, F (2,59) = 4.69, p < .05, partial η^2 = .14. Follow-up analysis to the main effect for domain-specific interest examined all pairwise comparisons among three levels of domain-specific interest. The Tukey HSD procedure was used to control for Type I error across pairwise comparisons. The results of the analysis indicate that the group with low domain-specific interest reported significantly less sense of Spatial Presence than people who reported medium levels of domain-specific interest (see table 6). There was no significant difference between low and high levels and medium and high levels of domain-specific interest (DSI).

Type of stimulus	Personality factor(s) main effects	Levels of factor	Spatial Presence means	Standard deviation
Hypertext	DSI	Low	2.25	.48
		Médium	2.91	.60
		High	3.17	.69
VE	DSI	Low	2.59	.82
		Médium	3.57	.84

Table 6: Spatial Presence mean values for personality main effects on hypertext and VE conditions

8. Discussion

Our results showed that three personality variables (i.e., domain specific interest, absorption and agreeableness) were associated with Spatial Presence. There was a positive correlation between Spatial Presence and domain specific interest for three of the four conditions. Even though all the differences between groups with low, medium and high level of domain specific interest were not significant in the hypertext, film and VE conditions, there was some indication that subjective sense of Spatial Presence increases with increasing domain specific interest. Results of the film experiment also indicate that higher levels of absorption and agreeableness are associated with higher levels of Spatial Presence.

Two of the three trait factors of the MEC-SPQ (i.e., domain specific interest and absorption) were significantly associated with Spatial Presence. Visuo-spatial imagery was not correlated with presence reflecting perhaps the fact that questionnaires are not a sensitive enough method to detect differences in visualization abilities.

The results suggest that when users are interested in the topic of the media stimulus, they are more motivated to pay attention to the stimulus, and thus as a result, experience higher levels of Spatial Presence. It was reasonable to hypothesize that domain specific interest would play a more prominent role when using low-immersive media stimuli such as linear text. Surprisingly, the effect was significant only for the hypertext, film and VE stimuli, not for the linear text stimulus. The role of domain specific interest might be the larger the more specific the topic of the stimulus. Since the media stimuli we used here were commercial products and of general interest, its role was perhaps smaller than if we had used stimuli that are of less general interest.

Absorption was significantly associated with Spatial Presence, and in the film-stimulus condition the group with high level of absorption experienced higher levels of Spatial Presence. The finding that those people who have a higher tendency to dwell on the experiences and on the media objects themselves experience higher levels of Spatial Presence in the film-stimulus condition is consistent with previous studies. Recently, Sas [23] found a significant association between absorption and presence, and Laarni et al. [24] showed that presence correlated significantly with self-transcendence which, in turn, has shown to be related to absorption [25].

Only one factor of the Big Five (i.e., agreeableness) was associated with Spatial Presence. Those people who get high scores on agreeableness are typically altruistic, helpful, friendly, tenderminded, credulous and empathetic; those people who get low scores are, in turn, typically selfish, distrustful, competitive and antagonistic. It is possible that those users who get high scores on this scale are more eager to suspend of disbelief and adapt to the media world. Empathy is typically characterized as a state in which a person is able to perceive accurately the internal reference frame of another person and be in-tune with him/her. Even though empathy is then normally considered when talking about interpersonal relations, it is also possible to feel empathetic to non-living objects such as different types of media stimuli. Empathy has shown to be typical to those who get high scores on agreeableness [26], and it is possible that this ability is crucial for experiences of presence. Interestingly, Sas [23] found that empathy was positively associated with presence.

It is also possible that demand characteristics play a differential role for those who get high scores on agreeableness and for those who get low scores. Demand characteristics are a problem caused when participants can predict the response that is expected and they respond either in accord to or against those expectations. It is possible that high scorers on agreeableness are more willing to respond in accord to expectations than those who get low scores. For example, when high scorers discovered the aim of the study, they tried to be helpful and gave higher presence ratings to the stimuli.

Contrary to our hypothesis, there was no association between openness to experience and Spatial Presence. This negative finding may reflect the fact that the Big Five's openness to experience is factorially complex [27]. For example, such aspects of the Openness to experience as intellectual curiosity and willingness to try new things are not necessarily associated with Spatial Presence, and they are not included into the absorption scale.

Overall, the Big Five personality factors seemed to contribute quite little to Spatial Presence. It is possible that personality factors, and user-related factors in general, play a minor role in Spatial Presence. Formal characteristics of the media and situational factors are much more important and can explain most of the variance in ratings. Another possibility is that our measures are not sensitive enough to show the effects of user-related factors on Spatial Presence. Especially, if personality factors exert their influence on the first stage of the construction of Spatial Presence, their effect is not easily seen if we measure the products of the second stage (i.e., Spatial Presence itself). A better approach would then be to study their effect on the first stage constructs (ie., on attentional engagement and construction of spatial situation model).

We have recently suggested that the effect of user-related factors may interact in a complex way with the properties of media stimulus. For example, in case of low-immersive media such as radio and TV, the role of user-related factors may be larger than in case of complex interactive virtual environments [10]. The present results, however, do not provide much support to the claim that different aspects of personality would be important for different types of media stimuli. For example, it might be assumed that the effect of personality would be more prominent when using low-immersive media stimuli such as linear text, but this hypothesis was not supported. Different media properties such as FoV had not any effect on Spatial Presence. It may be possible that the difference between small and large FoVs (20º vs. 60º) was not large enough in the present study.

If state of presence is something that is worth to be gained, it might be a good idea to try to identify those people that are better able to experience presence and those who are not able to experience it at all [10]. For example, training in simulators or virtual therapy services could be directed to those people who are better able to get absorpted to experiences and who are more friendly and empathetic. However, much research has to be done before people can be chosen for these services on the basis of personality test results.

9. Conclusion

We have shown that some personality factors are important determinants of Spatial Presence. Especially, it was found that domain specific interest, absorption and agreeableness are positively associated with Spatial Presence. Users who are more interested in the topic of the stimulus, who are more prone to dwell on the experiences and who are more empathetic gave higher Spatial Presence ratings than those who had lower scores on these scales. The effect of domain specific interest was prominent for the hypertext, film and VE stimulus; the effect of absorption and agreeableness was the clearest for the film stimulus.

Overall, the present results provide some support for the underlying theoretical model of Spatial Presence which argues that domain specific interest and absorption have an impact on the formation of presence experiences

However, better empirical evidence is needed to show whether cognitive abilities (i.e. spatial visual imagery) play any role in the formation of presence.

Acknowledgements

This work has been supported by the European Community, Information, Society and Technology Program, within the research project “Presence: Measurement, Effects, Conditions (MEC)”, IST-2001-37661 (http://www.ijk.hmt-hannover.de/presence). We thankfully acknowledge the European Community’s support.

References

[1] Steuer, J. (1992). Defining Virtual Reality: Dimension determining Telepresence. *Journal of Communication, 42*(4), 73-93.

[2] Sheridan, T.B. (1992). Musings on telepresence and virtual presence. Presence: *Teleoperators and Virtual Environments, 1*(1), 120-125.

[3] Draper, J.V., Kaber, D.B., & Usher, J.M. (1998). Telepresence. *Human Factors, 40*(3), 354-375.

[4] Barfield, W., Zeltzer, D., Sheridan, T.B., & Slater, M., (1995). Presence and performance within virtual environments. In W. Barfield & T.A. Furness III (Eds.). *Virtual environments and advanced interface design* (pp. 473-513). Oxford: Oxford University Press.

[5] Riva, G., Davide, F., & Ijsselsteijn, W.A. (2003). (Eds.), *Being there: concepts, effects and measurement of user presence in synthetic environments*. Amsterdam: Ios Press.

[6] Wirth, W., Hartmann, T., Böcking, S., Vorderer, P., Klimmt, C., Schramm, H., Saari, T., Laarni, J., Ravaja, N., Gouveia, F. R., Biocca, F., Sacau, A., Jäncke, L., Baumgartner, T., & Jäncke, P. (submitted for publication). *A Process Model of the Formation of Spatial Presence Experiences.*

[7] Lee, K.M. (2004). Presence, explicated. *Communication Theory, 14*, 27-50

[8] ISPR (2001). *What is Presence?* [Online]. Available: http://www.temple.edu/mmc/explicat.htm [9.10.2002].

[9] Heeter, C. (1992). Being there: The subjective experience of presence. *Presence, 1*(2), 262-271.

[10] Sacau, A., Laarni, J., & Hartmann, T. (submitted for publication). *Influence of individual factors on Presence.*

[11] Sas, C., & O’Hare, G.M.P. (2003). Presence equation: An investigation into cognitive factors underlying presence. Presence: *Teleoperators and Virtual Environments, 12*(5), 523-537.

[12] Hartmann, T., Böcking, S., Schramm, H., Wirth, W., Klimmt, C., & Vorderer, P. (2005). Räumliche Präsenz als Rezeptionsmodalität: Ein theoretisches Modell zur Entstehung von Präsenzerleben [Spatial presence as modus of reception: A theoretical model]. In V. Gehrau, H. Bilandzic & J. Woelke (Eds.), *Rezeptionsstrategien und Rezeptionsmodalitäten* (pp. 21-37). München: R. Fischer.

[13] Wild, T. C., Kuiken, D., & Schopflocher, D. (1995). The role of absorption in experiential involvement. *Journal of Personality and Social Psychology, 69(*3), 569-579.

[14] McCrae, R.R. (Ed.). (1992). The five-factor model: Issues and applications [Special issue]. *Journal of Personality, 60* (2).

[15] Church, A. T. (1994). Relating the Tellegen and Five-Factor Models of personality structure. *Journal of Personality and Social Psychology, 67*, 898-909.

[16] Bolger, N. & Zuckerman, A. (1995). A framework for studying personality in the stress process. *Journal of Personality and Social Psychology, 69*, 890-902.

[17] Penley, J.A., & Tomaka, J. (2002). Associations among the Big Five, emotional responses, and coping with acute stress. *Personality and Individual Differences, 32*, 1215–1228.

[18] Rose, C. L., Murphy, L. B., Byard, L., & Nikzad, K. (2002). The role of the Big Five personality factors in vigilance performance and workload. *European Journal of Personality, 16*, 185-200.

[19] Nigg, J. T., John, O. P., Blaskey, L. G., Huang-Pollock, C.L., Willicut, E, G. Henshaw, S. P., & Pennington, B. (2002). Big Five dimensions and ADHD symptoms: Links between personality traits and clinical symptoms. *Journal of Personality & Social Psychology, 83*, 451-469.

[20] Ravaja, N., Saari, T., Kallinen, K., & Laarni, J. (in press). The role of mood in the processing of media messages from a small screen: Effects on subjective and physiological responses. *Media Psychology*.

[21] Vorderer, P., Wirth, W., Saari, T., Gouveia, F. R., Biocca, F., Jäncke, L., Böcking, S., Schramm, H., Gysbers, A., Hartmann, T., Klimmt, C., Laarni, J., Ravaja, N., Sacau, A., Baumgartner, T., and Jäncke, P. (2004). *Development of the MEC Spatial Presence Questionnaire (MEC-SPQ).* Unpublished report to the European Community, Project Presence: MEC (IST-2001-37661).

[22] Costa, P. T. Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Odessa, FL: Psychological Assessment Resources.

[23] Sas, C. (2004) Individual differences in virtual environments. In M. Bubak, G. Dick van Albada, P. Sloot

& J. Dongarra (Eds.). *Computational Science - ICCS 2004, 4th International Conference, Kraków, Poland, June 6-9, 2004, Proceedings, Part III. Lecture Notes in Computer Science 3038* (pp. 1017-1024). Springer-Verlag.

[24] Laarni, J., Ravaja, N., Saari, T., & Hartmann, T. (2004). Personality-related differences in subjective presence. In M. Alcañiz & B. Rey (Eds.). *Proceedings of the Seventh Annual International Workshop Presence 2004* (pp.88-95). Valencia: Ed.UPV.

[25] Laidlaw, T.M., Dwivedi, P., Naito, A., & Gruzelier, J.H. (2005). Low self-directedness (TCI), mood, schizotypy and hypnotic susceptibility. *Personality and Individual Differences,* in press.

[26] Ashton, M.C., Paunonen, S.V., Helmes, E., & Jackson, D.N. (1998). Kin altruism, reciprocal altruism, and the Big Five personality factors. *Evolution and Human Behavior, 19*, 243–255.

[27] Glisky, M.L., Tataryn, D.J., Tobias, B.A., Kihlstrom, J.F., & McConkey, K.M. (1991) Absorption, openness to experience, and hypnotizability. *Journal of Personality & Social Psychology, 60,* 263-272.

Culture Matters - A Study on Presence in an Interactive Movie

Jun Hu, Christoph Bartneck
Department of Industrial Design, Eindhoven University of Technology
{j.hu@tue.nl, christoph@bartneck.de}

Abstract

We conducted a cross cultural study to test the influences of different cultural backgrounds on the user's presence experience in interacting with a distributed interactive movie. In addition we were interested in the effects of embodied interaction on presence. The influence of culture background is clear - Chinese participants perceived more presence than Dutch participants in all conditions. The results also show that interaction methods (direct touch against remote control) had no influence, while embodiment (robot against screen agent) had mixed effects on presence.

Keywords--- **presence, culture, interactive movie**.

1. Introduction

The user's character is believed to influence the user's feeling of presence. The user's cultural background is often mentioned as such a characteristic [1, 2]. A few cross-cultural presence studies are available [3], but none investigated the relationship between the user's cultural background and presence directly. This influence is, at this point in time, more of a conjecture than a proven fact, and therefore we conducted an empirical study to investigate the relationship.

In absence of a clear definition of what cultural factors may influence presence, a good approach is to include participants from clearly different cultures. Using Dutch and Chinese participants in our study optimized cultural diversion. Hofstede [4] provides an empirical framework of culture by defining several dimensions of culture, such as power distance, individualism/collectivism, masculinity/femininity, uncertainty avoidance and long/short-term orientation. China and the Netherlands differ substantially on all dimensions except uncertainty avoidance (see table.1). Power distance, for example, refers to the extend to which less powerful members expect and accept unequal power distribution within a culture. The Netherlands rank very low on this dimension, while China ranks very high.

From an application point of view, China is currently one of the most promising economic opportunities. Its vast populace and large physical size alone mark it as a powerful global player. China's gross domestic product (GDP) growth of over seven percent indicates its steaming economic situation. Most Chinese already have access to a TV and the local TV manufacturers satisfy the domestic market. However, technology utilizing presence is not yet produced or consumed. Awareness of cultural differences in presence may help companies to create better products for the different markets.

Table 1:Hofstede's [1993] Culture Dimension Scores for Dutch and Chinese

	Dutch	Chinese
Power Distance	38L	80H
Individualism	80H	20L
Masculinity	14L	50M
Uncertainty Avoidance	53M	60M
Long Term Orientation	44M	118H

H = top third, M = medium third, L = bottom third (among 53 countries and regions for the first four dimensions; among 23 countries for the fifth.)

At the same time, we were interested in distributed interactive media and their influences on the user's feeling of presence. We have entered a new media era: passive television programs become interactive with the red button on your remote control [5]. Video games come with many different controlling interfaces such as dancing mats, EyeToy cameras, driving wheels and boxing Gametraks [6]. The D-BOX Odyssee motion simulation system even introduces realistic motion experiences, which were originally designed for theme parks, into our living rooms [7]. In the vision of Ambient Intelligence [8], the next generation of people's interactive media experience will not unfold only on a computer or television, or in a head set, but in the whole physical environment. The environments involve multiple devices that enable natural interactions and adapt to the users and their needs.

Formerly, distribution only revered to the distribution of data or computational processes in a network. Ambient intelligent environments build on this technology and extend it by distributing synchronized and interactive multimedia content on to multiple devices. Previous systems already employed distributed presentation to enhance the entertainment experience and thereby increasing the immersiveness of the content. Multichannel surround sound systems, for example, distribute sound all around the audience and hence provide a more realistic and natural sound experience. The ambient intelligence [8] concept goes beyond such sound distributions by distributing content through other channels in the user's environment. Displays in the room may show video clips, lamps may change its color and brightness, robots may dance and sing, and couches may vibrate. The virtual space or the content, then, is no longer yielded in traditional audio and video materials by one TV set, but now expanded into the user's surroundings covering more sensory modalities. The light color, robotic behavior and the couch vibration are parts of delivered content, conveying a virtual experience but with a direct physical embodiment. Ambient intelligence is therefore a distinct extension to classical virtual environments.

However, distributing interactive content to multiple devices would also increase the complexity of interaction. The environment together may become difficult to understand and to control. To ease the situation, embodied characters, such as eMuu [9] or Tony [10], may be used to give such an environment a concrete face. These characters have a physical embodiment and may present content through their behavior and interact with the user through speech and body language. They can even be used as input devices.

Moreover, the influence of embodiment on the user's presence experience seems unclear. On the one hand, embodiment extends the distributed content from an on-screen virtual environment to a physical environment. The physical embodiment improves the content's aliveness and fidelity by stimulating more sensors of the user. This might result in an increased feeling of presence [11]. On the other hand, the physical embodiment may transfer more attention from the virtual environment to the physical environment. The physical embodiment may remind the user of its existence in this world and may break down the illusion of being there and hence would result in less feeling of presence [2]. The division of attention itself might also have such an effect.

To control interactive content, the user requires interaction devices. A physical embodiment would invite direct manipulation. A robot could, for example, ask the use to touch its shoulder to select an option. Interaction with a virtual on-screen character may favor the use of a remote control. Embodiment in interactive media can therefore not be studied without considering the interaction method. We therefore included two interaction methods in our study.

In this framework of interactive distributed media we defined the following three research questions:

1. What influence has the user's cultural background on the users' presence experience?

2. What influence does the embodiment of a virtual character have on the users' presence experience?

3. Would direct touching the presented content objects bring more presence than pressing buttons on remote controls?

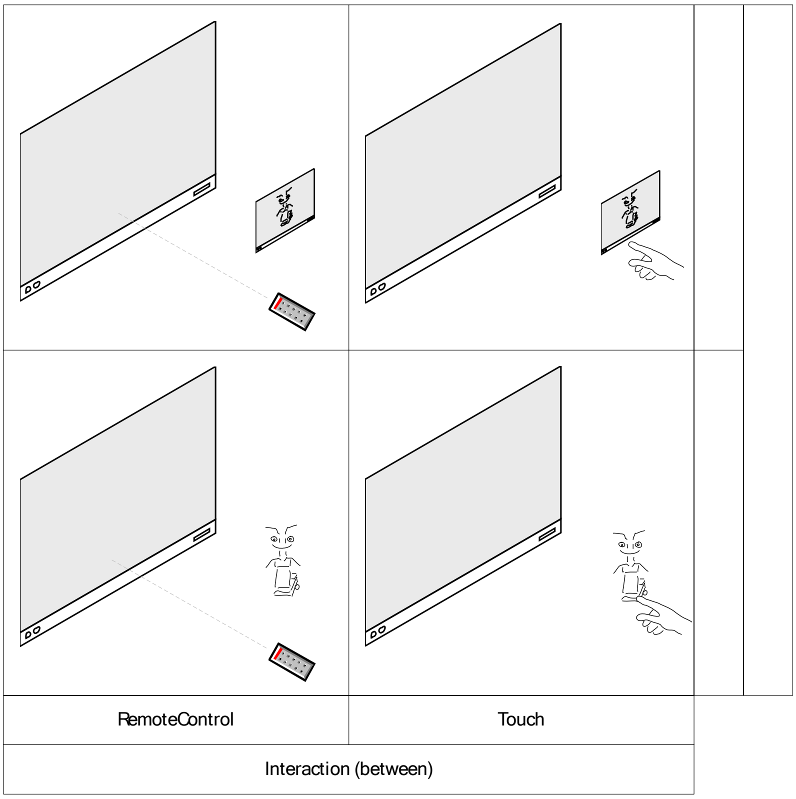


Figure 1: Conditions of the experiment with Chinese and Dutch participants

2. Experiment

We conducted a 2 (Interaction) × 2 (Embodiment) × 2 (Culture) mixed between/within experiment (see figure 1). Interaction and culture were the between participant factors. Interaction had the conditions RemoteControl and DirectTouch, and culture had the conditions Dutch and Chinese. Embodiment was a within participant factor. Embodiment had the conditions ScreenAgent and Robot.

2.1 Measurements

The original ITC-SOPI [12] questionnaire was used and only the definition of the *Displayed Environment* in the introduction was adjusted to include the robot/screen character. The Chinese participants had a good understanding of the English language and therefore no validated translation was necessary for them. The questions remained unchanged and are clustered into four groups: 1. *Spatial Presence*, a feeling of being located in the virtual space; 2. *Engagement*, a sense of involvement with narrative unfolding within virtual space; 3. Ecological validity, a sense of the *naturalness* of the mediated content; 4. *Negative effects*, a measure of the adverse effects of prolonged exposure to the immersive content.

2.2 Participants

19 Chinese and 24 Dutch between the age of 16 and 48 (14 female, 29 male) participated in the experiment. Most of them were students and teachers from Eindhoven University of Technology, with various backgrounds in computer science, industrial design, electronic engineering, chemistry, mathematics and technology management. The Chinese participants were no longer than two years in the Netherlands. All participants had good command of the English language and were frequently exposed to English speaking media, such as movies, web pages, news papers and TV shows.

2.3 Setup

The experiment took place in a living room laboratory (see figure 2). The participants were seated on a couch in front of a table. The coach was 3.5*m* away from the main screen, which was projected onto a wall in front. The projection had a size of 2.5*m* × 1.88*m* with 1400 × 1050 pixels. The second screen was located 0.5*m* from the coach, standing on the table. The secondary screen was 30*cm* × 23*cm* with 1280 × 1024 pixels LCD touch screen (Philips DesXcape Smart Display).

In the Robot conditions, the secondary touch screen was replaced with a Lego robot that had about the same height. In the ScreenAgent conditions, the secondary screen displayed a full screen agent of the robot.

(a) ScreenRemote/ScreenTouch

(b) RobotRemote/RobotTouch

Figure 2: Experiment setup

The behavior of the screen based agent and the Lego robot were identical. They played the role of a TV companion by looking randomly at the user and the screen, but always looking at the user while speaking. Speakers were hidden under the table and were used to produce the speech, which was based on the standard Apple Speech Synthesis software. At the start of every movie, the character introduced himself and its role.

Since a media content that is acceptable in one culture can be perceived inappropriate, rude or offensive in another [13], the movie was designed to be culturally neutral. The movie had an international cast: the applicant was played by a Moroccan, the employer by a Dutch, the secretary by an American, and the passer-by by a Chilean. The actors spoke English. This study does not investigate the influence of media content on presence and therefore the story and movie cuts were neither too exciting nor too boring for both Dutch and Chinese participants. Otherwise they might have masked the effects of embodiment and culture.

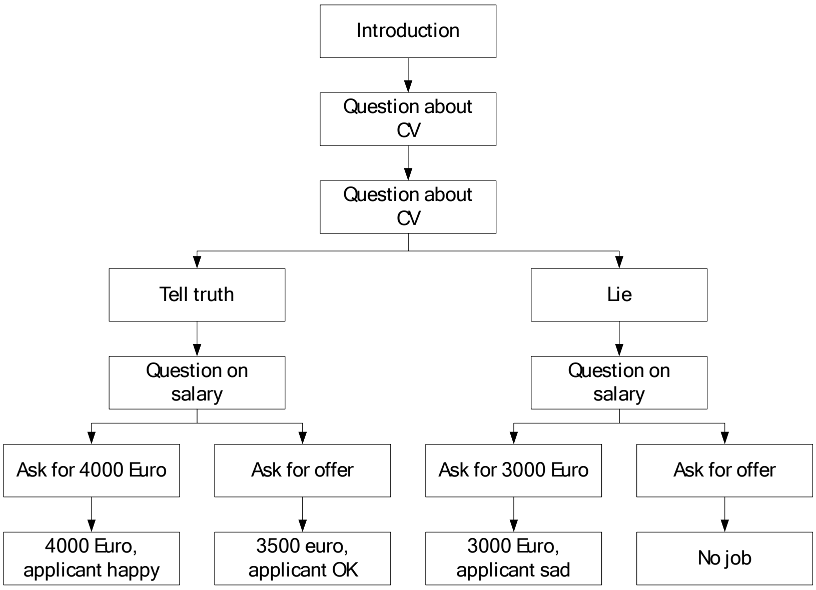


Figure 3: Storyline

The interactive movie, about approximately 10 minutes, was about a job interview in which the participants had to make decisions for the applicant. The storyline was discussed with several Chinese and Dutch to assure that the actions of the characters would be plausible in both cultures. The movie had two decision points, which resulted in four possible movie endings (see figure 3). The participants chose different options for decisions almost all the time. At every decision point camera would zoom in on the applicant's forehead (see figure 4). The actor then cycled through two options in his mind. He looked first to the left and thought aloud about one option, before he looked right and thought aloud about the second option. In the remote condition the screen would show one icon on the left and a different icon on the right. The icons were identical to two icons on the remote control. In the robot condition, the participant had to touch the left or the right shoulder of the robot to make the decision.

Figure 4. A decision point

2.4 Procedure

After reading an introduction that explained the structure of the experiment the participants started with a training session. In this session, the participants watched an unrelated interactive movie that had only one decision point, during which the participants could make the decision using the remote control. Afterwards, they had the opportunity to ask questions about the process of the experiment. Next, the participant were randomly assigned to one of the between-participant conditions, which each consisted of two movies and a questionnaire after each movie. The participant received five Euros for their efforts.

3 Results

The mean scores for all measurements, including their standard deviations are presented in table 2 and graphically in figure 5.

A 2 (embodiment) × 2 (interaction) × 2 (culture) repeated measures ANOVA was conducted. Interaction had no significant influence on any of the measurements. Embodiment and culture both had significant influence on almost all measurements (see table 3).

Interaction was removed as a factor from the further analyses since it had no effect on the measurements. The means for all remaining conditions are summarized in figure 6 and were used as the basis for the further analyses.

Paired Sample t-Tests were performed across both culture conditions to test the influence of embodiment. The measurements for Spatial Presence were significantly ($t(42) = 2.235$, $p = 0.031$) higher in the ScreenAgent condition than in the Robot condition. Negative Effects were significantly (*t*(42) = 2.38, *p* = 0.022) higher in the Robot condition than in the ScreenAgent condition.

Independent Samples t-Tests were performed to test the influence of culture. All measurements between the Dutch and the Chinese participants differed significantly, except for engage in the screen condition, which just missed the significance level (*t*(41) = 2.007, *p* = 0.051).

Table 2: Mean and standard deviation for all measurements

Embodiment	Culture	Interaction	Measurement	Mean	Std.Dev.
ScreenAgent	Chinese	RemoteControl	Spatial Presence	3.08	0.18
			Engagement	3.35	0.37
			Naturalness	3.17	0.32
			Negative Effects	1.96	0.55
		DirectTouch	Spatial Presence	2.79	0.37
			Engagement	3.28	0.41
			Naturalness	2.92	0.61
			Negative Effects	1.83	0.52
	Dutch	RemoteControl	Spatial Presence	2.56	0.29
			Engagement	3.17	0.51
			Naturalness	2.75	0.50
			Negative Effects	1.46	0.43
		DirectTouch	Spatial Presence	2.44	0.45
			Engagement	2.84	0.58
			Naturalness	2.58	0.74
			Negative Effects	1.5	0.36
Robot	Chinese	RemoteControl	Spatial Presence	2.99	0.2
			Engagement	3.33	0.24
			Naturalness	2.73	0.17
			Negative Effects	3.28	0.42
		DirectTouch	Spatial Presence	2.72	0.59
			Engagement	3.22	0.43
			Naturalness	3.08	0.32
			Negative Effects	3.35	0.52
	Dutch	RemoteControl	Spatial Presence	2.51	0.25
			Engagement	3.07	0.61
			Naturalness	2.55	0.66
			Negative Effects	2.9	0.4
		DirectTouch	Spatial Presence	2.26	0.41
			Engagement	2.86	0.56
			Naturalness	2.42	0.59
			Negative Effects	2.52	0.82

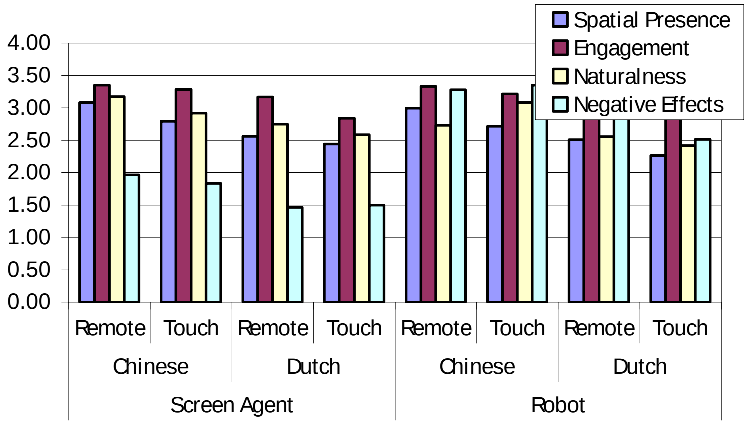


Figure 5. Means of all measurements in all conditions

Table 3. F and p values for culture and embodiment

Factor	Measurement	F (1,39)	p
Embodiment	Spatial Presence	4.789	**0.035**
	Engagement	0.515	0.477
	Naturalness	4.335	**0.044**
	Negative Effects	119.973	**0.001**
Culture	Spatial Presence	19.49	**0.001**
	Engagement	4.962	**0.032**
	Naturalness	7.494	**0.009**
	Negative Effects	24.491	**0.001**

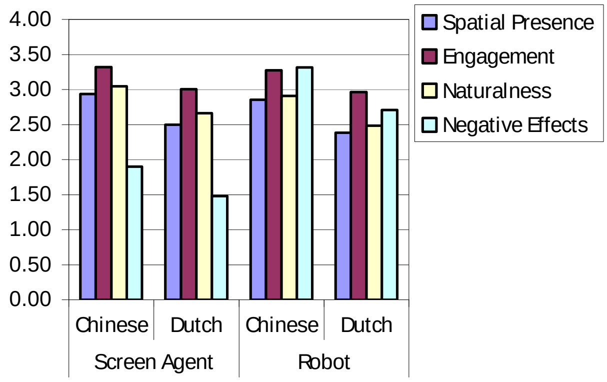


Figure 6: Means in the culture and embodiment conditions

4 Discussion

4.1 Culture effects

The participants' cultural background clearly influenced the measurements. Chinese participants perceived more presence than Dutch participants in all conditions.

The next question will be what aspects of the cultural background have the greatest influence on presence. Hofstede [15, 4] suggested several dimensions through which cultures and organizations may be characterized. Peppas [14] points out that Chinese culture is much influenced by the 2000-year history of Confucianism, based on an intensive overview on the studies on Chinese culture. Confucianism, as a morel system, defines three cardinal guides (ruler guides subject, father guides son and husband guides wife) and five constant virtues (humanity, righteousness, propriety, wisdom and fidelity). It roots the Chinese culture has high power distance, low individualism, medium masculinity and uncertainty avoidance, and very much of long term orientation. Living in harmony, maintaining a good social network, and protecting one's "face" are important in Chinese society.

One might suspect that because of the importance of politeness and maintaining one's or the other's "face", the Chinese participants might simply be more polite in answering questions. Our measurements show that they also gave higher scores to Negative Effects and therefore did not simply respond politely.

Comparing to Chinese, Dutch culture has much less power distance (means more equality and modest leadership), much higher individualism, more femininity, similar uncertainty avoidance and it is short term oriented. Maintaining the consensus (although it is often time consuming), and being (even pretending to be) modest are considered important in Dutch culture.

However none of Hofstede's [15, 4] culture dimensions appear relevant to presence at first sight. One might speculate that the long-term orientation in Chinese culture would result in more patience towards imperfections. They might have more easily tolerated the noise emitted by the robot and the occasional visibility of a microphone in the movie. Further studies are necessary to investigate this issue.

4.2 Embodiment effects

The influence of embodiment on all measurements does not conform to the expected results defined in the construct of presence. According to Lessiter et al. [12],

> *Whilst in the current study Negative Effects was not strongly correlated (positively or negatively) with Engagement or Ecological Validity, it was significantly but modestly (and positively) related to Sense of Physical Space.*

However, in our results Spatial Presence and Naturalness are higher in the ScreenAgent condition, while Negative Effects was higher in the Robot condition. Negative Effects appear to have been affected by something else than presence.

During the experiment, the robot's motor emitted noise, which caused the participants to look at it. A moving physical object is potentially dangerous and hence attracts attention. Clearly, the robot emphasized the participants feeling of being in the room and not in the movie and thereby reducing the presence experience. The screen character did not emit noise and is unable to pose physical danger to the user. It therefore did not attract as much attention as the robot.

The participants were frequently switching between looking at the movie and the robot and hence divide their attention. This switching made it hard for the users to stay focused and might cause the high negative experience. Eggen, Feijs, Graaf, and Peters [16] showed that a divided attention space reduced the user's immersion. Further research is necessary to determine if divided attention increases the negative effects of multiple displays. The extra costs necessary to build and maintain a robot for an interactive movie appear unjustified in relation to its benefit.

The different interaction methods (using a remote control or touching directly) had no influence on the measurements. The participants did not experience more or less presence when they interacted with a remote control or with the screen/robot directly. This is to some degree surprising, since the participants had to lean forward to touch the screen/robot directly, while they could remain leaned back using the remote. The necessity to make a choice might have overshadowed the difference in physical movement. To create compelling sense of presence it might be useful to pay more attention to the physical output than to the input.

4.3 Future Research

In this study, several factors were investigated besides the cultural background of the participants. It might be beneficial to conduct a dedicated study on the influence of culture on presence. Such a study could then also cover more than the two cultures investigated in this study. In addition, it appears necessary to further connect the results of such a study to existing results in other research areas, such as cross-cultural communication studies. Qualitative interview might help to gain better insights into to social and cultural viewpoints of the participants in relation to presence.

Ambient Intelligence is currently a major research theme in the European academic and commercial world, but the results of this study show that cultural aspects do play a role in the design of future technology. Given China’s rapid grow and potential, it might be valuable to include an “Eastern Perspective” into the European research, in particular since more and more consumer electronics are produced and consumed in Asia.

5 Acknowledgments

The authors would like to thank Loe Feijs, Maddy Jansen and Kees Overbeeke for their help and support. In addition, we would like to thank all participants of the study.

References

[1] W.A. IJsselsteijn, H.d. Ridder, J. Freeman, and S.E. Avons. Presence: concept, determinants, and measurement. *Human Vision and Electronic Imaging V*, 3959(1):520-529, 2000.

[2] J. Freeman, J. Lessiter, and W. IJsselsteijn. An introduction to presence: A sense of being there in a mediated environment. *The Psychologist*, 14:190-194, 2001.

[3] T.-T. Chang, X. Wang, and Lim. Cross-cultural communication, media and learning processes in asynchronous learning networks. In *HICSS (the 35th Annual Hawaii International Conference on System Sciences)*, pages 113-122, 2002.

[4] Geert Hofstede. Cultural constraints in management theories. *Academy of management executive*, 7(1):81-94, 1993.

[5] J. Bennett. *Red Button Revolution - Power to the People*. 2004. URL http://www.bbc.co.uk/pressoffice/speeches/stories/bennett_mip.shtml.

[6] In2Games. *Gametrak :: Come in and Play*. 2005. In2Games Corporate and Press Site: http://www.in2games.uk.com.

[7] D-BOX. *D-BOX Motion Simulators*. 2005. D-BOX Technologies Inc. website: http://www.d-box.com.

[8] Emile Aarts and Stefano Marzano. *The New Everyday View on Ambient Intelligence*. Uitgeverij 010 Publishers, 2003.

[9] C. Bartneck. *eMuu - An Embodied Emotional Character for the Ambient Intelligent Home*. Phd thesis, Eindhoven University of Technology, 2002.

[10] C. Bartneck and J. Hu. Rapid prototyping for interactive robots. In F. Groen, N. Amato, A. Bonarini, E. Yoshida, and B. Kröse, editors, *The 8th Conference on Intelligent Autonomous Systems (IAS-8)*, pages 136-145, Amsterdam, 2004. IOS press.

[11] M. Lombart and T. Ditton. At the heart of it all: The concept of presence. *Journal of Computer Mediated Communication*, 3(2), 1997.

[12] J. Lessiter, J. Freeman, E. Keogh, and J. Davidoff. A cross-media presence questionnaire: The itc sense of presence inventory. *Presence: Teleoperators and Virtual Environments*, 10(3):282-297, 2001.

[13] M. Lu and D.F. Walker. Do they look at educational multimedia differently than we do? a study of software evaluation in taiwan and the united states. *International Journal of Instructional Media*, 26(1), 1999.

[14] Spero C Peppas. Us core values: a chinese perspective. *International Journal of Sociology and Social Policy*, 24(6):58-75, 2004.

[15] Geert Hofstede. The cultural relativity of organizational practices and theories. *Journal of International Business Studies*, 14(2, Special Issue on Cross-Cultural Management):75-89, 1983.

[16] B. Eggen, L. Feijs, M.d. Graaf, and P. Peters. Breaking the flow - intervention in computer game play through physical and on-screen interaction. In *the ’Level Up’ Digital Games Research Conference*, 2003.

Session 6

22nd September, 2005

Talking Faces and Social Collaboration

15.00-15.30 *The Influence of Lip Animation on the Perception of Speech in Virtual Environments*
Edwin Blake and Johan Verwey
Department of Computer Science, University of Cape Town, South Africa

15.30-16.00 *Providing High Social Presence for Mobile Systems via an Unobtrusive Face Capture System*
Miguel A. Figueroa-Villanueva[1], Frank A. Biocca[2], Chandan K. Reddy[1], Jannick P. Rolland[3], and George C. Stockman[1]
[1] Computer Science and Engineering Department, Michigan State University East Lansing, USA
[2] Department of Telecommunications, Michigan State University East Lansing, USA
[3] School of Optics University of Central Florida Orlando, USA

16.00-16.30 *A Study of Influential Factors on Effective Closely-Coupled Collaboration based on single user perceptions*
Oliver Otto, David J. Roberts and Robin Wolff
The Centre for Virtual Environments, University of Salford, UK

16.30-17.00 *Social Presence in Two- and Three-dimensional Videoconferencing*
Joerg Hauber[1], Holger Regenbrecht[2], Aimee Hills[2], Andrew Cockburn[1] and Mark Billinghurst[1]
[1] University of Canterbury, Christchurch, New Zealand
[2] University of Otago, Dunedin, New Zealand

The Influence of Lip Animation on the Perception of Speech in Virtual Environments

J.H. Verwey, E.H. Blake
Collaborative Visual Computing Laboratory
Department of Computer Science
University of Cape Town
Rondebosch, South Africa
{jverwey@cs.uct.ac.za, edwin@cs.uct.ac.za}

Abstract

The addition of facial animation to characters greatly contributes to realism and presence in virtual environments. Even simple animations can make a character seem more lifelike and more believable. The purpose of this study was to determine whether the rudimentary lip animations used in most virtual environments could influence the perception of speech. The results show that lip animation can indeed enhance speech perception if done correctly. Lip movement that does not correlate with the presented speech however resulted in worse performance in the presence of masking noise than when no lip animation was used at all.

1. Introduction

The ability to read a speaker's lips has a significant impact on speech perception [26] . In other forms of media like television and cinema, this visual cue is readily available. For virtual environments lip animations have to be created for every character that will be speaking. Since this can be a time consuming process, most applications provide only very rudimentary lip animations, if at all. While some lip movement certainly contributes to realism and a feeling of presence [6] [10] , it is uncertain whether it can contribute to speech perception. Lip reading in real life provides additional visual information that is integrated with the auditory information in the perception of speech. The perceptual system will however rely more on the visual modality when the auditory cues are weak [9] . Virtual environments are interactive by nature. Sounds can be generated at arbitrary times, which could make it more difficult to hear spoken dialog. In contrast to this, sound tracks for film and television are completely linear and sound engineers have exact control over what the listener will hear. Providing correctly animated lips may therefore be more important in virtual environments where a greater reliance is placed on visual information than in an animated film where the sound track can be edited until the dialog is clear.

Most studies involving lip reading make use of video streams of real faces. It has been shown that video streams with frame rates as low as five frames per second can still contribute to speech perception [14] . Some studies have shown that the artificial reconstruction of lip movement using 3D geometric models can also benefit hearing performance [18] . This benefit may extend to simpler lip animations typically used in virtual environments.

In this study subjects were required to identify spoken words that were accompanied either with simple but correctly constructed lip animation, incorrect animation or no lip animation at all. A noise masker was presented together with the spoken sentence in order to make the task more difficult. We show that correct lip-animation enhances speech perception, but incorrect animation degrades speech perception. The effect is most pronounced when hearing is made difficult by the masking sound. Under these conditions the visual modality is favoured and subjects tend to perceive the visually presented rather than the auditory presented words.

1.1. Visemes for Animation

Auditory speech sounds are classified into units called phonemes. The visual counterpart for a phoneme is called a *viseme* [5] . A viseme represents the shape of the lips when articulating an auditory syllable. Many phonemes however have ambiguous visual representations and map to the same viseme. The Preston Blair phoneme series [2] is a popular set of visemes often used for facial animations in cartoons. In this series only 10 visemes are used to map to all possible phonemes (Figure 1).

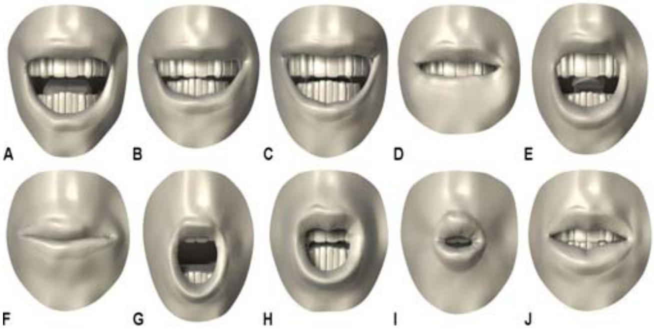


Figure 1. The Preston Blair phoneme series. Each visual representation (viseme) represents one or more auditory phonemes.

Chen *et al* [7] presents an overview of different methods of creating speech-driven facial animations and lip synchronization. Lip animations are constructed by either using a flipbook method, or by using geometry morphing. The flipbook method rapidly displays a list of consecutive visemes together with the auditory speech to create an impression of lip movement. Since there are a limited number of facial expressions, this method can result in jerky animations when no intermediate frames are drawn for the transition between different visemes. The geometry morphing method requires a 3D model of a face to be constructed. The geometry of the face can be smoothly interpolated between different facial expressions resulting in very smooth animation.

Both methods require the different visemes to be synchronized with auditory phonemes as they are spoken. Lip animations can be derived from acoustical speech input by using various computational methods. Lavagetto made use of neural networks for speech-driven facial animation in a multimedia telephone application for the hard of hearing [18] . He showed that the resulting lip animations of a geometric model were useful for enhancing speech perception. Much simpler methods are used for creating animations when using the flipbook method. Software tools like PAMELA [25] extract phonemes from a given text sentences and map them to visemes. The time offset for each viseme can be manually adjusted until the animation looks realistic.

The computational cost involved in creating facial animations directly from the acoustical speech data can be prohibitive for virtual environments that typically spend most processing time on graphics, physics and artificial intelligence computations. The flipbook method is more suitable for these kinds of applications since it uses very few computational resources [7] .

1.2. Accurate Speech Perception

Animations need to be carefully constructed. Phonemes should map to the correct visemes in order for the additional visual information to contribute to speech perception. “The *McGurk* Effect” [21] has shown that different syllables can even be perceived when contradictory visual information is presented together with auditory speech. This effect illustrates how incongruent auditory and visual information can cause a different perception of the auditory stimuli.

For example, when someone hears the auditory syllable /ba/ but sees the visible syllable /ga/ (Viseme (B) followed by viseme (A) in Figure 1) being articulated, it is usually perceived as /da/. The perceived audio-visual syllable has the same visual representation as the presented visual syllable but differs from the presented auditory syllable.

Only some combinations of auditory and visual syllables produce McGurk effects. These studies typically use a limited set of stimuli that usually only consist of single syllables. It has however been shown that the McGurk effect can be obtained for normal words. If the visually and auditory presented words are picked very carefully a completely different English word can be perceived. If for example the auditory word ‘mail’ were presented together with the visual word ‘deal’, the word ‘nail’ would be perceived [9] . It is clear that the visual representation of a spoken sentence can have a significant impact on the perception of the words.

Adverse listening conditions may further aggravate this effect. Interaction in the virtual environment could cause additional sounds to be produced that could drown out spoken dialog. The addition of a masking noise will cause greater reliance on the visual cues. When two sources of information conflict, in this case visual and auditory, the stronger source is usually favoured [9] . Incorrectly constructed lip animations may therefore result in worse hearing performance than when the listener only relies on auditory information.

1.3. Directional Sound

Immersive virtual environments often present a variety of background sounds, music, dialog and effects simultaneously. The human perceptual system has the remarkable ability to pay selective attention to a sound of interest in the midst of other competing sounds. This is often called the “Cocktail-party Effect” [8] [12] . This ability allows listeners to attend to a specific voice while ignoring other voices and background noise. One of the contributing factors in distinguishing sound sources is their physical location [3] . A difference in the location of sound sources greatly enhances the intelligibility of speech in the midst of a masking noise or other competing voices. This is referred to as a *spatial release from masking* or *spatial unmasking* [12] . Directional sound can influence speech perception in real life as well as in virtual environments.

Research in virtual auditory environments has shown that it is possible for sounds to be presented over stereo headphones in such a way that it is perceived as coming from any position in 3D space [1] . Digitized sound data are manipulated to create a stereo sound stream with the separate channels representing the sound that would be perceived at each ear. Slight changes in level, timing and spectrum at each ear will cause virtual sound sources to be perceived at different locations when played over stereo headphones. This is referred to as *sound spatialization* or more commonly, *3D sound*. It has been shown that a release from masking can be obtained in virtual auditory environments where virtual sound sources are spatially separated from one another [11] .

In this study we presented target speech sentences from different locations relative to a masking sound. This allowed us to investigate the influence of the visual lip animation cues at different levels of hearing difficulty.

3. Method

We investigated the problem by performing a large number of trials over an extended period on a few volunteers.

3.1. Subjects

Four paid volunteers were recruited as test subjects for this research. All subjects were between the ages 20 and 30, had self-reported normal hearing and normal or corrected-to-normal vision. Subjects were not informed of the goal of the experiments. The use of four subjects may seem rather few in the Virtual Reality field. The nature of the phenomena being investigated, however, are such that we do not expect much variation between subjects, provided they have normal hearing. The variation is expected to arise within the experimental subjects and should decrease as the task is learned. The strategy in this sort of research is to choose a few volunteers and then to conduct a very large number of trials with each person. The most rigorous approach would have been to first establish that our subjects did have normal hearing and vision but in practice self-report as well as an initial control of the results for outliers is acceptable. This methodology is consistent with other speech perception studies [4] [16] [24] .

3.1. Stimuli

Sentences from the Coordinate Response Measure (CRM) corpus [23] were used as auditory stimuli. This corpus has a limited vocabulary with target words consisting of a call sign, a colour and a number. Sentences have the following format:

“Ready (Call sign) go to (Colour) (Number) now.”

The call sign can be ‘Arrow’, ‘Baron’, ‘Charlie’, ‘Ringo’, ‘Laker’ or ‘Tiger’. The possible colours are ‘Blue’, ‘Red’, ‘White’ or ‘Green’ while the numbers range from one to eight. Subjects were required to identify the correct colour and number combination in a spoken sentence while a masking noise was simultaneously presented. Although the CRM corpus is publicly available for research, all speakers used for the recordings had American accents. Since some subjects might find it difficult to recognize a foreign accent, especially in noisy conditions, it was decided to create a CRM corpus using a native speaker. A native English-speaking female drama student was used as voice talent. Professional sound engineers were employed to record the target stimuli. Each sentence was 2.5 seconds in length on average and was recorded at 48 kHz. The sound files were first edited to make sure every file immediately started with the first word without any delay. The sound files were also trimmed at the end after the last word has been spoken.

Since the call sign was not important for our experiments, only the call sign “Baron” was used. The number 7 was not used in any trials since it is the only two-syllable number and would be easier to recognize. This left four colours and seven numbers in the vocabulary. With 28 possible permutations of colour and number, the chance of a subject guessing both the correct colour and number is 3.6%. In some cases subjects may have been able to recognize only one of the target words, this would clearly be better than recognizing nothing at all. Since this information would be lost when using absolute scoring, it was decided to award a point for answering the correct colour and another point for the correct number. When scored this way the chance of a correct guess is 19.6%.

Since distracting sounds in virtual environments are not limited to speech sources, it was decided not to use speech spectrum or speech shaped noise for these experiments as is common in speech perception studies. White noise of the same length as the longest speech stimulus was generated for the masking stimuli. Ten different masking files were created in this way and were randomly presented during experiments. The root mean square (RMS) energy of a sound file refers to the square root of the mean of the squares of the all the digitized sound sample values. In order to make sure the target-to-noise-ratio was calculated correctly, the RMS energy of the masker and the target sounds were first normalized. All target files were scaled to have data values in the (-1, 1) range. The minimum RMS energy for these files was then calculated and all files were scaled to have the same RMS energy. The masker stimuli were then scaled to have the same RMS as the normalized target stimulus. All stimuli were ramped with a cosine-squared window to remove any clicking at the beginning and end of sentences when presented.

Brungart [5] provides an overview of the creation of spatialized audio over stereo headphones. This involves convolving the impulse responses measured on a KEMAR dummy head with the target stimuli to create a separate set of stereo sound files. KEMAR is a standard audiological research mannequin manufactured by Knowles Electronics. MIT Media Lab measured the impulse responses used in this study [15] . Spatialized sounds were produced by convolving the signals with KEMAR HRTFs for angles 0° and 15° in the horizontal plane. All sounds were created with a zero elevation angle. No further processing was performed on the stereo sound files during presentation.

For visual stimuli 3D models were used to represent the sound producing objects. These models can be seen in Figure 2.

Figure 2. A screen shot of the virtual environment.

A television screen that displayed a snowy picture, as is common when there is bad reception, represented the masker object. A face representing the target was presented on a separate television in the virtual environment. The snowy television was animated by randomly switching between different noisy images at a constant frame rate.

Illustrations of the Preston Blair phoneme series [2] were used for animating the face of the character. Each animation frame in Figure 3 represents a viseme that corresponds to one or more phonemes. An animation file containing the relative timing offsets of different frames was created with the help of a lip synchronization utility called PAMELA [25] . This tool can determine the correct phonemes to use for any given English sentence. These phonemes were then mapped to appropriate visemes in the Preston Blair series as shown in Table 1.

Phoneme	Example	Viseme
AA	Father = F AA DH ER	J
AE	At = AE T	J
AH	Hut = HH AH T	J
AO	Dog = D AO G	J
AW	Cow = C AW	J
AY	Hide = HH AY D	J
B	Be = B IY	G
CH	Cheese = CH IY Z	C
D	Deed = D IY D	C
DH	Thee = DH IY	C
EH	Ed = EH D	D
ER	Hurt = HH ER T	B
EY	Ate = EY T	D
F	Fee = F IY	F
G	Green = G R IY N	C
HH	He = HH IY	H
IH	It = IH T	J
IY	Eat = IY T	D
JH	Gee = JH IY	H
K	Key = K IY	C
L	Lee = L IY	E
M	Me = M IY	G
N	Knee = N IY	C
NG	Ping = P IY NG	H
OW	Oat = OW T	I
OY	Toy = T OY	I
P	Pea = P IY	G
R	Read = R IY D	C
S	Sea= S IY	C
SH	She = SH IY	H
T	Tea = T IY	H
TH	Theta = TH IY T AH	C
UH	Hood = HH UH D	I
UW	Two = T UW	I
V	Vee = V IY	F
W	We = W IY	A
Y	Yield = Y IY L D	C
Z	Zee = Z IY	C
ZH	Seizure = S IY ZH ER	H

Table 1. Phonemes mapped to visemes.

While Pamela cannot create the correct timing offsets for each frame from the speech file, it does allow the user to adjust the timing offsets until the animation looks correct. For the sentence "Ready Baron, go to blue, one now", Pamela would produce the following phonemes for each word: Ready - R, EH, D, IY, Baron - B, AE, R, AH, N, go - G, O, to - T, UW, blue - B, L, UW, one - W, AH, N, now - N, AW. These phonemes were mapped to the following visemes in Figure 3: Ready - C, D, C, C, Baron - G, J, C, J, C, go – C, I, to – H, I, blue – G, E, I, one – A, J, C, now C, J. Animation files were created in this way for every sentence presented during the experimental trials.

Figure 3. Target speech animation frames.

3.2. Procedure

The experimental software was run on a desktop-based system with a 3000 MHz Intel Pentium processor and 512 MB RAM. The system was also equipped with a GeForce FX5900 graphics card with 128 MB onboard RAM and a Creative Labs Sound Blaster Audigy 2 sound card. A pair of Sennheizer HD 580 circum-aural headphones was used as for the auditory display and a Virtual Research V6 Head-mounted display (HMD) for the visual display. This HMD supports a resolution of 640x480 and can display a 60° field-of-view. The virtual environment application was written in C++ using the Microsoft DirectX API [22] .

All subjects participated in five experimental sessions on five consecutive days. During each experimental session an adaptive method was first used to determine the subject's speech reception threshold (SRT). This refers to the minimum target-to-noise ratio (TNR) at which subjects can reliably perform the task. The transformed up-down method [19] was used to determine the SRT. This method targets the 71% correct response threshold. The method starts with equal target and masking noise levels yielding a target-to-noise ratio of 0 dB. For normal hearing subjects it is very easy to achieve a 100% correct score in this condition.

The task is then progressively made more difficult by lowering the volume of the target stimulus whenever the subject scores two correct answers in a row. As soon as the subject gives a single incorrect response the level is adjusted to make it easier again. A reversal happens when the subject either scores two consecutive trials correct after an incorrect response, or if an incorrect response directly follows two or more correct responses. The process is stopped when the number of reversals reaches a predetermined threshold. During this adaptive procedure the target level is adjusted with varying amounts. Initially big step sizes are used in order to reach the SRT more quickly. The step size is progressively made smaller to obtain a more optimal SRT.

The following values were determined during pilot studies.

- Until 1st reversal, adjust the volume by 5dB.
- Until 3rd reversal, adjust the volume by 3dB.
- Until 7th reversal, adjust the volume by 1dB.
- Until 13th reversal, adjust the volume by 0.5dB.

Once the SRT has been determined for the subject, all experimental trials can be presented at the measured TNR for different conditions. The 71% correct response threshold leaves enough room to show an increase or decrease in performance when the experimental condition is changed.

For the first 3 days subjects had to complete 3 adaptive learning blocks to find an adequate TNR for each subject. During the adaptive trials only audio was presented and the target and masker objects were invisible. The auditory masker was always presented at 0° while the target was presented at 15° to the right. Each of these blocks lasted for about 5-6 minutes. An experimental block of up to 20 minutes followed after this. The average TNR measured in the 3 adaptive blocks was used as the TNR for the experimental block. On the last two days no adaptive blocks were conducted, but two experimental blocks, using the average TNR measured on the third day. Three visual conditions were presented. The face could be correctly animated, incorrectly animated or not animated at all. For incorrectly animated conditions, the animations of a different target sentence were randomly selected. For example, if the auditory stimulus was the sentence “Ready Baron go to blue, one now”, the visual animation for the sentence “Ready Baron go to green, five now” could be presented.

Two spatial conditions were presented. The target object was either presented at 0° or 15° to the right. The masker was always presented directly in front of the listener at 0°. When both the masker and target were presented from the same direction, the target object obscured the masker object as seen in Figure 4.

Figure 4. The co-located condition. The masker object is obscured.

	Target	Masker	Animation
1.	0°	0°	Animated
2.	0°	0°	Non-animated
3	0°	0°	Incorrectly animated
4.	15°	0°	Animated
5.	15°	0°	Non-animated
6.	15°	0°	Incorrectly animated

Table 2. Spatial and visual conditions presented.

The six different conditions presented are summarized in Table 2. The reason for using two spatial positions in this experiment was to investigate the influence of lip-animation at different levels of hearing difficulty. Because of spatial unmasking, the target sentence would be easier to hear when presented at 15° than in the co-located condition.

A slightly transparent input console was superimposed on the display area at the end of each trial. This allowed subjects to provide responses without having to remove the HMD. This can be seen in Figure 5.

Some target words in the CRM corpus are easier to recognize than others [4] . If some words presented in one condition were easier to identify than words presented in another condition, this would create a misleading bias towards one condition. To prevent this, all sentences were presented an equal number of times under all experimental conditions. This ensured that an equal number of easy and difficult sentences were presented for all conditions, removing the bias towards any one condition.

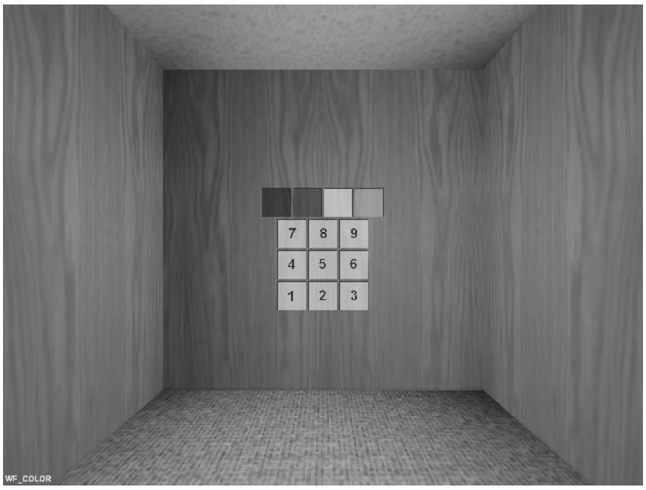


Figure 5. Input console for trial responses.

To minimize the effect of fatigue, all sessions were kept under one hour and subjects were given a short break between blocks of trials. During pilot testing it was observed that subjects tend to perform better towards the end of a block than at the very beginning. To account for any learning effects within a block, a few warm-up trials were first presented. These trials were not considered for data analysis.

During each experimental block, 28 trials were presented for six different conditions. The last two sessions contained two experimental blocks and no adaptive blocks. A total of seven experimental blocks were conducted over the five days. This resulted in 196 trials per condition. This excludes any adaptive trials since the number of trials presented during each of these blocks naturally varies. To account for learning effects, the first two experimental blocks were not considered for data analysis, leaving 140 usable trials per condition for every subject.

5. Results

Figure 6 shows subject performance for different visual conditions. From left to right the conditions were: correctly animated, non-animated and randomly animated. A spatial release from masking was observed for all three visual conditions. Subjects performed best for correct lip animations and worst when incorrect animations were used. An ANOVA between the co-located and separated conditions showed a significant difference between the two spatial conditions [F(1,3) = 664.97, (p<0.001)]. Further analysis showed that the difference is significant for all visual conditions. An ANOVA across the different visual conditions also revealed a statistically significant difference between the three visual conditions [F(2, 6) = 28.2, p < 0.001]. Further comparisons revealed that all visual conditions differ significantly for both spatial conditions.

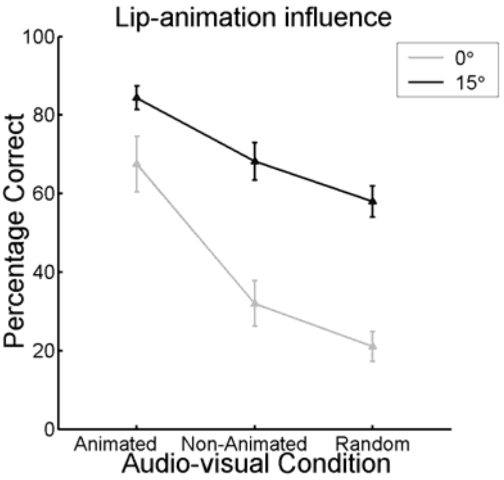


Figure 6. Subject performance under different visual and spatial conditions. 0º means the target and masker were co-located and 15º is the spatially separated condition where the target sound was located to the right. The vertical bars represent the standard deviation.

From Figure 6 we saw that performance for the incorrectly animated condition was worse than the correctly animated and non-animated visual conditions. In this condition animations from different colour and number combinations were used as visual stimuli. The question arises whether this incorrect visual information is merely distracting or whether it created a perceptual bias in favour of the visually presented words.

One could use an alternative scoring to determine how well the subject would have performed if we used the visually presented colour and number as the correct response instead of the auditory. If subjects consistently picked the colours and numbers they saw, one could conclude that subjects relied more strongly on the visual than the auditory cues.

From Figure 7 it is clear that when scoring in this way there is a dramatic difference in the results. On the left the responses are scored according to the auditory presented stimuli. On the right, subject responses are scored against the visually presented stimuli.

For the co-located condition, subjects performed better when using the alternative scoring method. The visual score was significantly higher than the auditory score, which is almost the same as chance (19.6%). This implies that subjects tended to answer according to the visually presented stimuli, that is, the visemes, in the co-located condition. In the separated condition, where spatial unmasking resulted in better hearing conditions, subjects tended to answer according to the auditory presented stimuli, ignoring incongruent visual information. In this condition the visual score was slightly above chance indicating that the incorrect animation still had some impact.

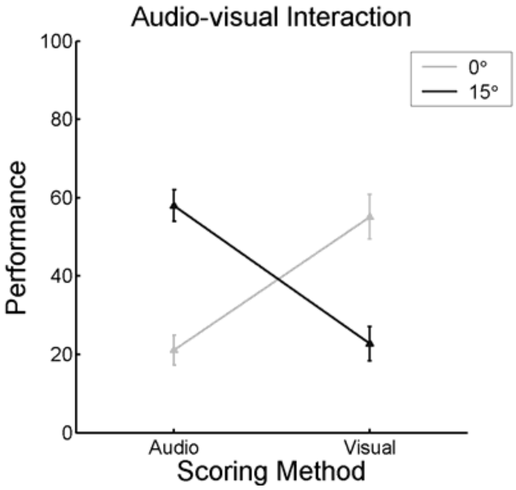


Figure 7. Subject performance when using two different scoring methods for the randomly animated condition. When both the target and the masker were presented at 0° subjects tended to answer according to what they saw rather than what they heard. When the objects were separated by 15°, the auditory cues were stronger and subjects answered according to what they heard while ignoring incorrect visual information.

6. Discussion

Our results show that subjects found it much easier to recognise target words when the noisy television was presented in front and the target television was presented to the right of the masker. Although not the primary objective of this research, this result shows that a spatial release from masking was obtained when the target and masker sounds were presented from different directions. This is consistent with previous findings in the literature [11] .

In this experiment we expected correct lip animation to aid in speech recognition. The results confirmed this and show that correct lip animation significantly contributes to hearing performance. As expected, the incorrect animated condition did result in worse performance than the non-animated condition.

It may be that the interaction between visual and auditory information caused subjects to hear something completely different as is found in experiments involving the McGurk effect. However McGurk experiments are generally very carefully constructed. Only some combinations of strong visual cues with opposing weak auditory cues produce this effect. It is unlikely that the vocabulary of the CRM corpus would result in any McGurk effects when presenting random combinations of auditory and visual stimuli.

In the co-located condition, both the target speech and the masking noise were presented from the same location making it very difficult to hear the target words. The massive increase in performance between the non-animated and correctly animated case in the co-located condition suggests that subjects are able to lip-read very well. It therefore seems likely that the visual cue also had a big influence during the incorrectly animated condition. From Figure 7 we can see that subjects indeed scored higher for the visually presented words than for the auditory ones. This suggests that at least for the co-located condition the visual cue was favoured and subjects answered according to what they saw rather than what they heard.

In the separated condition the target speech and masking noise were spatially removed making it easier to distinguish the target words even though the animation was incorrect. Subjects performed reasonably well and the auditory score was better than the visual score. These results are consistent with findings in the literature that suggests that the stronger cue will usually be favoured when two sources of information conflict [9] .

Overall these results suggest that adding lip animation to characters in virtual environments will significantly increase hearing performance but only if done correctly. What makes these results even more interesting is that the animations used were extremely basic. Other studies suggest that 5 unique frames per second is the bare minimum for visual cues to contribute to speech recognition [14] . Those results were obtained with the use of a video stream and the 5 unique frames did not necessarily include the visemes linked to each phoneme. By constructing the animation in such a way that all visemes are included, a significant increase in hearing performance can still be obtained with minimal effort. Note that these conclusions are only relevant under conditions where it is very difficult to hear. Under normal listening conditions the strong auditory cues will usually be enough to disambiguate any incongruent visual cues. These results do however show that users rely heavily on visual cues under adverse hearing conditions.

7. Conclusion

We expected even rudimentary lip animation to enhance speech perception in virtual environments. A significant improvement in hearing performance for the correctly animated condition over the non-animated condition was demonstrated, confirming the hypothesis. The animations used during this study were extremely basic, consisting of only ten unique frames. We have shown that even such simple animations significantly aid speech perception when correctly synchronized with matching auditory stimuli. The results of subject performance under incorrectly animated conditions show that under adverse hearing conditions, the visual modality is favoured. This implies that when using lip animation, the visemes and phonemes have to match or performance under adverse hearing conditions will be even worse than when no animation was used.

This could have implications for virtual environments with dialog in different languages. Creators of virtual environments do not have exact control over what the user will hear at any given time. Having separate animations for different languages therefore becomes more important for virtual environments than for other forms of media like cartoons or 3D animated films where there is more control over the final audio track.

Although one cannot edit the sound track beforehand, virtual environments do provide additional auditory cues not present in other forms of media. Directional cues can enhance the perception of speech in the midst of competing sounds if the target sound is presented from a different direction than the masking sound. 3D spatialized sound should therefore be used to present speech in virtual environments where the hardware platform supports it.

Acknowledgements

The research was supported by the Innovation Fund of the South African National Research Foundation. The authors would like to thank John Turest-Swartz and his staff from the Contemporary African Music and Arts Archive for providing a recording studio and allowing the use of their offices while conducting our experiments. The assistance in experimental design by Prof. Barbara Shinn-Cunningham, Antje Ihlefeld and Tim Streeter from Boston University is much appreciated.

References

[1] R.D. Begault. *3-D Sound for Virtual Reality and Multimedia. Cambridge, MA: Academic Press Professional. 1994.*

[2] P. Blair. Cartoon Animation. *Available online at* http://www.freetoon.com. 2005.

[3] A. Bregman. Auditory scene analysis. *Cambridge, MA: MIT Press.* 1990.

[4] D. Brungart. Evaluation of speech intelligibility with the coordinate response measure. *Journal of the Acoustical Society of America, 109, 2276-2279*. 2001.

[5] D. Brungart. Near-Field Virtual Audio Displays. *Presence: Teleoperators and Virtual Environments. Vol 11, No. 1, 93-106, February 2002.*

[6] D. Burford, E. Blake. Real-Time Facial Animation for Avatars in Collaborative Virtual Environments. *South African Telecommunications Networks and Applications Conference '99, 18-183. 1999.*

[7] T. Chen, R.R Rao. Audio-visual integration in multimodal communication. *Special Issue on Multimedia Signal Processing, 86, 837-852*. 1998.

[8] E.C. Cherry. Some experiments on the recognition of speech, with one and two ears. *Journal of the Acoustic Society of America, 25, 975–979*. 1953.

[9] D.J. Dekle, C.A. Fowler, M.G. Funnell. Audiovisual integration in the perception of real words. *Perception & Psychophysics 51, 355-362*. 1992.

[10] S. DiPaola, D. Collins. A Social Metaphor-based 3D Virtual Environment. In *Educators Program From the 30th Annual Conference on Computer Graphics and Interactive Techniques. 1-2*. 2003.

[11] R. Drullman, A.W. Bronkhorst. Multichannel speech intelligibility and talker recognition using monaural, binaural and three-dimensional auditory presentation. *Journal of the Acoustic Society of America, 107, 2224-2235*. 2000.

[12] M. Ebata. Spatial unmasking and attention related to the cocktail party problem. *Acoustics, Science and Technology, 24, 208-219*. 2003.

[13] R.L. Freyman, K.S. Helfer, D.D. McCall, R.K. Clifton The role of perceived spatial separation in the unmasking of speech. *Journal of the Acoustic Society of America, 106, 3578-3588*. 1999.

[14] H.W. Frowein, G.F. Smoorenberg, L. Pyters, D. Schinkel Improved Speech Recognition Through Videotelephony: Experiments with the Hard of Hearing. *IEEE Journal on Selected Areas in Communication, 9, 611-616*. 1991.

[15] Gardner B., Martin K. (1994). HRTF Measurements of a KEMAR Dummy-Head Microphone. *Available online at http://sound.media.mit.edu/KEMAR.html*

[16] K.W. Grant, P.F. Seitz. The use of visible speech cues for improving auditory detection of spoken sentences. *Journal of the Acoustic Society of America, 108, 1197-1207*. 2000.

[17] K.P. Green, J.L. Miller. On the role of visual rate information in phonetic perception. *Perception & Psychophysics, 38, 269-276*. 1984.

[18] F Lavagetto. Converting speech into lip movements: A multimedia telephone for hard of hearing people. *IEEE Transactions on Rehabilitation Engineering, 3, 1-14*. 1995.

[19] H. Levitt. Transformed Up-Down Methods in Psychoacoustics. *Journal of the Acoustic Society of America, 49*, 467-476. 1971.

[20] F.J. McGuigan. Experimental Psychology. A Methodological Approach. *Prentice Hall Inc.* 1968.

[21] H. McGurk, J. MacDonald. Hearing lips and seeing voices. *Nature, 264, 746-748*. 1976.

[22] Microsoft Corp. (2004). *MS DirectX Programming Guide. Available at http://www.msdn.microsoft.com/directx/*

[23] T. Moore. Voice communication jamming research. *AGARD Conference Proceedings 331: Aural Communication in Aviation, 2:1–2:6*. 1981.

[24] B.G. Shinn-Cunningham, A. Ihlefeld. Selective and divided attention: Extracting information form simultaneous sound sources. *Proceedings of the 2004 International Conference on Auditory Display, 10, 51-59*. 2004.

[25] M. Strous. PAMELA – Lipsynch utility for Moho. *Available online at* http://www-personal.monash.edu.au/~myless/catnap/pamela/index.html.

[26] W.H Sumby, I. Pollack. Visual Contribution to Speech Intelligibility in Noise. *Journal of the Acoustic Society of America, 26, 212-215*. 1954.

Providing High Social Presence for Mobile Systems via an Unobtrusive Face Capture System

Miguel A. Figueroa-Villanueva[1], Frank A. Biocca[2], Chandan K. Reddy[1],
Jannick P. Rolland[3], George C. Stockman[1]
[1]Computer Science and Engineering Department, Michigan State University, East Lansing, USA,
[2]Department of Telecommunications, Michigan State University, East Lansing, USA,
[3]College of Optics and Photonics, University of Central Florida, Orlando, USA
{[1]miguelf@ieee.org, [2]biocca@msu.edu, [3]jannick@odalab.ucf.edu}

Abstract

During face-to-face collaboration people frequently monitor the other's facial expressions to determine their current state of attention, mood, and comprehension. Capturing a frontal view of the face of mobile users in multi-user collaborative environments has been a challenge for several years. A mobile social presence system is proposed that captures two side views of the face simultaneously and generates a frontal view in real-time. The face is modeled using an active appearance model (AAM) and a mapping of the side model to the frontal model is constructed from training. Frontal views are then generated by applying this mapping to the fitted side model during collaboration. Only a few model coefficients are transmitted for the synthesized facial frames, providing a highly compressed stream. The virtual frontal videos are of good subjective quality and the fitted estimate retains a high fidelity to the true model, with peak signal to noise ratio of about 40DB.

Keywords--- **Mobile face capture, head mounted display (HMD), active appearance model (AAM)**.

1. Introduction

One key motivation for the creation of mobile communication systems and advanced collaborative environments is the increase in real-time communication between mobile and distributed partners. During face-to-face collaboration, users frequently monitor the other's facial expressions to determine their current state of attention, mood, and comprehension. Although we may have face-to-face interactions with workmates or others, many of our social interactions include an increasing number of purely virtual interactions; we rarely or never meet face-to-face. When it comes to communications from remote places, the human face is the most important communicative part of the human body. It has great expressive ability that provides a continuous stream of cues that are used to modulate and tailor interpersonal communication. Mediated communications now include many cases where facial non-verbal information can be critical: negotiation, complex training, emergency communication, stressful or tense interactions, communication of positive affect, and group coordination and motivation. The facial expressions of a remote collaborator or a mobile user can convey a sense of urgency, emotional congruency, lack of understanding or confidence in action, or other nonverbal indicators of communication success or breakdown. With an increase reliance on telecommunication systems for group interaction, there is increased research in advanced social presence technologies. Current advanced teleconferencing and telepresence systems transmit frames of video. These frames are nothing but 2D images from a particular point of view. In order to get additional views, designers use either a panoramic system or interpolate between a set of views.

New and enhanced forms of remote collaboration through sophisticated environments such as those presented in [1-4] provide augmented reality features for a higher degree of *presence* of the remote collaborator in the communication channel and, potentially, free movement and unlimited views of the shared augmented reality environment.

The *Teleportal System* [5] is an augmented reality environment for remote communication and collaboration among multiple users. This effort envisions a Teleportal room such as in [4] that allows single or multiple users to enter a room sized display and use a broadband telecommunication link to engage in face-to-face interaction with other remote users in a 3D, augmented reality environment, hence providing a simultaneous interaction with virtual objects, real objects and models while supporting object interposition. It also allows for unobstructed 3D face-to-face capture and display. This unique feature is designed to support interaction between fully mobile virtual representations of user's faces in 3D space so that their position relative to other participants and objects under discussion is preserved. The goal is to support all the non-verbal and position cues of side-by-side collaboration including attentional cues (e.g., "where is the person looking now"), turn taking and other conversation modulation cues, and situated cues regarding emotional and comprehension states. Figure 1 shows a representation of the kind of interaction that the Teleportal System allows.

Figure 1 Conceptual drawing of the application scenario enabled by the Teleportal System. Two distant users are interacting on a task. One user is instructing how to proceed from a mobile location, while the other is executing the task. Both have visual feedback of the other's environment.

1.1. Main objectives

Capturing a clear, detailed frontal view of the face of mobile users in multi-user collaborative environments has been a challenge for several years. Technologies that occlude the user's field-of-view are not practical and potentially dangerous in full mobile outdoor settings. Other applications of facial capture systems include teleconferencing, wearable computing, and collaborative mixed reality environments. The Mobile Face Capture System (MFCS) is responsible for obtaining and transmitting a quality frontal face video of a remote user involved in the communication. The MFCS proposed here captures the two side views of the face simultaneously and generates the frontal view. This face capture equipment consists of two miniature video cameras and convex mirrors [5]. Figure 2 shows a conceptual drawing that illustrates the face-capture cameras and the mirrors with respect to the user's head. Each of the cameras is pointed towards the respective convex mirror, which is angled to reflect an image of one side of the face. The convex mirrors produce a slight distortion of the side view of the face. The left and right video cameras capture the corresponding side views of the human face in real-time. The goal of the work in this paper is to synthesize a frontal view facial image from the two side views recorded by the head mounted display (HMD) side cameras.

1.2. Advantages

Consider the contrast with conventional capturing techniques, where either the face capture system is static within the environment, for example a single camera mounted on a display, or the capture system is bulky, costly, and computationally expensive, for example a room instrumented with a sea of cameras [1]. The MFCS system is static with respect to the user's head movements, uses only two cameras to produce a wide range of views of a user's head including a possible stereoscopic view, can capture the face regardless of location, and works on any basic processor.

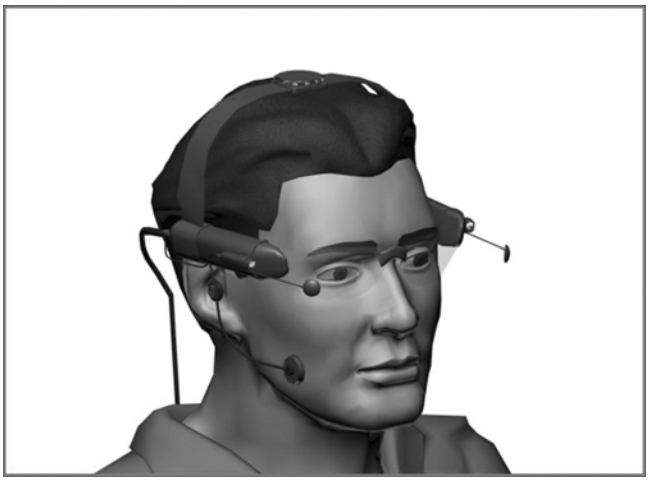


Figure 2 Mobile Face Capture System (MFCS) concept with two convex mirrors and two lipstick cameras.

Most previous systems have been built using a highly instrumented fixed indoor environment, while our work is motivated by a need to be mobile. In its current implementation it relies on the use of a head mounted display (HMD) that includes a projective display and mobile face capture system (MFCS).

The HMD will ultimately allow all participants to: (a) view 3D images of the face of remote collaborators, (b) view unobstructed the real local participants and objects, and (c) view the blending of physical and virtual objects. Although the MFCS system can be used with any display, combining it with an augmented reality (AR) HMD allows the user to see the 2D or 3D faces of collaborators in appropriate locations for interpersonal communication relative to their body or the environment.

Our current MFCS prototype consists of two side cameras and front mirrors as depicted in Figure 2. The basic requirement of the MFCS is that it must produce quality video of the wearer's face without interfering with the ability to perform other required tasks such as object manipulation and the 3D visualization of a remote communicator or of shared data or objects with that communicator. However, around one's office, a participant may reach out to data or files to share with others and any obstruction of any one participant's direct view would prohibit executing those tasks. We also anticipate the use of MFCS with outdoor, fully mobile AR systems or next generation mobile phones. In this demanding setting, the MFCS approach can minimize visual occlusions so as to not interfere with simple walking, driving, object manipulation or non-mediated face-to-face interaction. This current work forms a stepping-stone for the creation of a complete 3D augmented reality based face-to-face communication system that can produce stereoscopic views of the users via a real-time augmented reality display.

1.3. Organization of the paper

The remainder of this paper is organized as follows. Section 2 describes the relevant background for the MFCS. Section 3 describes the hardware system design. The equipment used and the optics issues are discussed. The algorithms and methods used in the MFCS are explained in Section 4. Section 5 illustrates some results of using the prototype MFCS. Section 6 presents conclusions of the work with the MFCS and suggests ideas for future work.

2. Related background

To synthesize a frontal view facial image, a model of the face is created during the training stage. This model is used to characterize the input streams at run time. The model is also used to create or instantiate the desired views (i.e., the frontal view, but potentially a wide range of views).

Face modeling has been used as a tool to aid in a large number of applications such as person identification, face surveillance, face animation, expression cloning, etc. As a result, there are a number of techniques employed to model the face. They can be categorized into 2D and 3D techniques. This paper follows a 2D analysis by synthesis approach: namely, *Active Appearance Models*, for its robustness and computational efficiency.

Active Appearance Models (AAMs) [6-8], which first appeared in [9], are non-linear, generative, and parametric statistical models of a certain deformable object in the 2D image plane. In particular, face modeling has been one of the most popular applications of AAMs [9].

The typical application scenario of AAMs involves a training phase, where the model is built, and a fitting phase, where a search is made to find the optimal model parameters that minimize the distance from the generated model instance and the input image. A detailed and comprehensive survey on the subject of AAMs and the closely related concepts of Active Blobs, Direct Appearance Models, and Morphable Models can be found in [10].

Other approaches to image synthesis have been reported [11] that could be used to synthesize images by interpolating some reference views of a static scene. In [12, 13] some extensions were made to be able to handle dynamic scene interpolation. These techniques rely on estimating the epipolar geometry of the scene and having a set of reliable correspondences. Also, care has to be taken to properly blend or interpolate between images to obtain a visually pleasing result. Typically, AAMs are less sensitive and error-prone than such approaches, at the expense of having to provide a database of training samples.

Reddy et. al. [14] proposed a method for synthesizing a frontal face image using a similar HMD. The proposed approach was to calibrate the system to obtain a set of warping functions to map pixels from the side images to virtual frontal image coordinates. A structured light grid was projected onto the face from the front and the deformation in the side images recorded to be used for warping during operation. Problems included use of structured light in the field, the blending of the two side images at their seam in the frontal image, and image distortion created by facial expressions not modeled well by the static warp. In contrast, the approach of this paper produces consistently smooth images and high compression. The costs are several minutes of training and fitting and the need to store the models at both the sender and receiver [15].

3. Face capture system design

3.1. Current hardware performance

The current HMD prototype cameras, Sony DXC-LS1 with Fujinon YF12B-7 lenses, are tethered by cables to a P4 1.7 GHz PC with 496 MB RAM. A Panasonic GP-KR202 video camera is positioned on the desk to take a real frontal video during training. The subject puts on the MFCS and minor adjustments may be made to the orientation of the mirrors. The subject then faces the Panasonic camera and speaks and gestures using a standard script. Standard office lighting is used. The system records synchronized video from the side MFCS and frontal observing cameras. Current storage resources limit us to recording 70 frames per session. The longest step in the training process is the manual identification of face feature points in the side (46 points) and frontal images (95 points), which may take 15 minutes. Fitting the AAM models to the side and front images and fitting the mapping from the side images to the frontal images is done in real-time. Thus, the entire training time for a single subject is currently about 15 minutes. Future improvements, including more automation in face point identification and sharing of data between subjects should reduce the training to 2 to 10 minutes. During the user task, generation of the virtual frontal video can be done in real time. Matthews and Baker [10] have shown that similar computations can be performed at over 260 frames per second.

Notice that there exist mobile and wireless counterparts for all the equipment in this prototype version, which can be replaced with off-the-shelf and dedicated hardware to obtain a mobile system.

3.2. Optical System Layout

The general layout of the system is shown in Figure 2. The calculations for estimating the variable parameters are simplified by unfolding the overall system. When the system is unfolded, the mirror can be represented as a negative lens (see Figure 3). The main components of this system are the (a) human face, (b) camera, and (c) mirror. The various parameters that are involved in the calculations are as follows.

1. *Human face:* The main parameters of the face that affect the geometry of the system are height and width. Other factors, such as skin color and illumination, affect the performance of the system but have no effect on the geometry. The dimensions of an average face are:
   - H - Height of the head to be captured (~ 250mm).

   - W-Width of the head to be captured (~ 175mm).
2. *Camera:* The main parameters are the size, the weight, the minimum working distance, the field of view, and the depth of field. Based on the approximate values of these parameters, we have obtained the off-the-shelf lipstick camera, Sony DXC-LS1. The two cameras are color balanced using their built-in hardware capabilities. The 12mm focal length lens has the following values:
   - Sensing area: 1/4", or equivalently 3.2mm(y) x 2.4mm(x).
   - Pixel Dimensions: the image sensed has a resolution of 768 x 494.
   - Focal Length (F_c): The focal length of the lens selected is 12 mm (VCL - 12UVM).
   - Field of View (FOV): The field of view of the camera with the above mentioned lens is 15.2^o x 11.4^o.
   - Diameter (D_c): The diameter of the lens and the camera is 12mm.
   - *f*-number (N_c): The *f*-number for this camera lens is 1. Although in practice, we adjust the iris according to illumination, we consider an *f*-number of 1 in the estimation of the other parameters.
   - Minimum Working Distance (MWD): The minimum working distance for the selected lens is 200mm.
   - Depth of Field (DOF): This parameter is dependent critically on the lens *f*-number which will vary with various illuminations. The higher the *f*-number the larger the DOF. This system requires however to consider the DOF of the camera and mirror combined. If the system has large DOF then it will be more portable and can accommodate many users without much change in the position and focus of the cameras. The DOF computation for the camera and mirror combined will be treated elsewhere with an in depth development of the optical layout and design.
3. *Mirror:* This is the most flexible component of the system. Hence, all the parameters of this component are estimated and the component is custom made. The various parameters of the mirror that will affect the geometry of the system are:
   - Diameter (D_m) / *f*-number (N_m)
   - Focal Length (F_m) or Radius of Curvature (R_m)
   - Magnification Factor (M_m)
4. *Distances*: Between these three components, we have the following distances:
   - D_{cm} - Distance between the camera and the mirror.
   - D_{mf} - Distance between the mirror and the face.

3.2.1. Estimation of the Variable Parameters (D_{mf} and D_m). From the theory of pupils and windows, the camera is the limiting aperture from the intermediary image plane located behind the mirror. Hence, the camera acts as the pupil of the system and the mirror is the window.

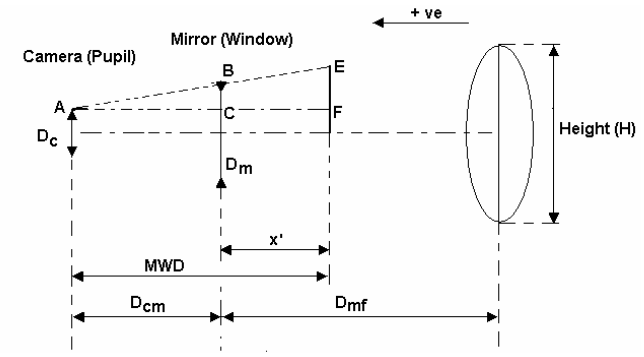


Figure 3 Optical system diagram for the estimation of the variable parameters D_{mf} and D_m.

In the unfolded configuration, the mirror is represented as a negative lens with image focal length f'_m equal in magnitude to that of the mirror with an opposite sign. The imaging equation for the equivalent lens to the mirror yields

$$\frac{1}{x'} = \frac{1}{D_{mf}} + \frac{1}{f'_m} \tag{1}$$

where *x'* is negative because the values D_{mf} and f'_m are negative. Hence, the image in the unfolded case is virtual and thus it is always between the lens and the human face. A study was made of estimated values for D_m as a function of the *f*-number and, based on the practical values for the size of the mirror (D_m) and the distances (D_{mf} and D_{cm}), the mirror was customized. A convex mirror of radius of curvature 155.04 mm was made corresponding to the *f*-number of 2. The convex side of the mirror was coated for the visible light spectrum. Figure 4 shows two sample images obtained from this optical system specification.

Figure 4 Sample images acquired from the current MFCS prototype with the optical specifications above.

4. Virtual view synthesis

4.1. System design

We present a generative and parametric method for face video synthesis. We build an AAM model from training data and use a regularization technique to determine the mapping between the AAM parameters for the side view model and the parameters for the front view model. Figure 5 depicts the training process where the goal is to build the corresponding AAM models and estimate the linear operator that describes the forward mapping.

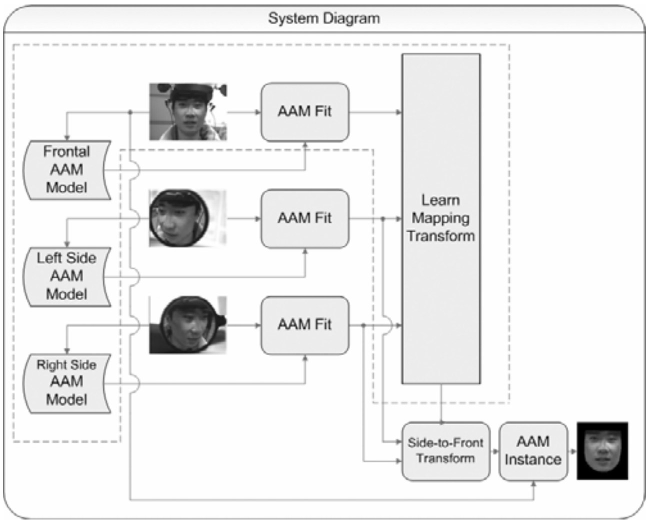


Figure 5 *System diagram.* From training, the AAM model is learned along with the transformation between side view parameters and frontal view parameters. At runtime the learned transformation is used to estimate frontal view parameters that instantiate the frontal view AAM.

After the forward linear operator is estimated, it can then be used to predict the frontal parameters for the respective AAM, as shown in Figure 5.

4.2. AAM modeling

This section describes the basic formulation of the AAM technique that provides the basis of our design. It is divided into the AAM model creation, the model instantiation, and the fitting process and it follows the notation presented in [10].

4.2.1. Model definition. AAMs model the shape, which accounts for the rigid form as well as the possible deformations, and texture (i.e., lighting intensity) of an object.

The shape is defined as a closed triangulated mesh, which can be represented as a vector containing the concatenation of vertex locations:

$$\mathbf{s} = (x_1, y_1, x_2, y_2, \cdots, x_v, y_v)^T \tag{2}$$

where v is the number of vertices of the mesh.

If there are n training shape vectors, then we can assume (provided that enough samples are given) that any new shape can be explained as a linear combination of those given as training:

$$\mathbf{s} = \mathbf{s}_0 + \sum_{i=1}^{n} p_i \mathbf{s}_i \tag{3}$$

where $\mathbf{s}_0$ is the mean shape and the $\mathbf{s}_i$'s are the variations or deformations from the mean. p_i's are the shape parameters.

The texture or appearance can be defined as the pixel intensities relative to the mean shape $\mathbf{s}_0$. Let $\mathbf{x}$ be the pixel locations in $\mathbf{s}_0$, then

$$A(\mathbf{x}) = A_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i A_i(\mathbf{x}) \quad \forall \mathbf{x} \in \mathbf{s_0} \tag{4}$$

is the appearance function.

4.2.2. Model instantiation. Given a set of parameters, $\mathbf{p} = (p_1, p_2, \cdots, p_n)^T$ for the shape and a set of parameters, $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \cdots, \lambda_m)^T$ for the appearance, an image can be synthesized corresponding to an instantiation of the model. The shape and appearance are generated independently by applying the parameters to Equations 3 and 4, respectively. However, the appearance is defined in terms of the mean shape $\mathbf{s}_0$, which requires warping to the generated shape instance. This process can be represented as:

$$I_M(W(\mathbf{x};\mathbf{p})) = A(\mathbf{x}) \tag{5}$$

where $W(\mathbf{x};\mathbf{p})$ is a piecewise affine warp from $\mathbf{s}_0$ to $\mathbf{s}$. $\mathbf{x}$ defines the pixel in $\mathbf{s}_0$ to be warped and $\mathbf{p}$ determines the shape $\mathbf{s}$ to be warped to.

4.2.3. Model fitting. In the fitting phase the goal is to search for the model parameters that minimize the error between the current image and the model instance for those parameters. This error can be defined as:

$$E(\mathbf{x}) = A_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i A_i(\mathbf{x}) - I(W(\mathbf{x};\mathbf{p})) \tag{6}$$

where the first term corresponds to the appearance defined by parameters $\boldsymbol{\lambda}$ of the model at pixel $\mathbf{x}$ in the base mesh, $\mathbf{s}_0$, and the second term corresponds to the pixel in the input image as determined by the warp $W(\mathbf{x};\mathbf{p})$. Hence, the problem has been reduced to an optimization problem with cost function $E(\mathbf{x})$ $\forall \mathbf{x} \in \mathbf{s}_0$ and parameters $\boldsymbol{\lambda}$ and $\mathbf{p}$ to search for. It should be noted that in practice, Principal Component Analysis (PCA) is applied to the shape and texture vectors, which makes the search more manageable.

4.3. Face modeling

Currently, the side view models are created with a mesh of 46 points and the frontal model with a 95 point mesh. In Figure 6 the contours of the base meshes are presented with two sample deviations along the first principal component direction. Note that this corresponds to the opening and closing of the mouth.

4.4. Frontal parameter estimation

The process of training and fitting an AAM has been briefly described in Section 4.2. After obtaining a synchronized stream of M images of the subject from the MFCS side cameras and a frontal camera, one can use this processing to obtain two row vectors $\mathbf{y}_i$ and $\mathbf{x}_i$ containing the frontal and side parameters for the i^{th} image, respectively. This can be written as:

$$\begin{aligned}\mathbf{y}_i &= (\mathbf{p}_{Fi}^T, \lambda_{Fi}^T) \qquad (7)\\ \mathbf{x}_i &= (\mathbf{p}_{Li}^T, \lambda_{Li}^T, \mathbf{p}_{Ri}^T, \lambda_{Ri}^T) \qquad (8)\end{aligned}$$

for $i = 1, \cdots, M$. *F*, *L*, *R* indicate the front, left, and right side parameters, respectively.

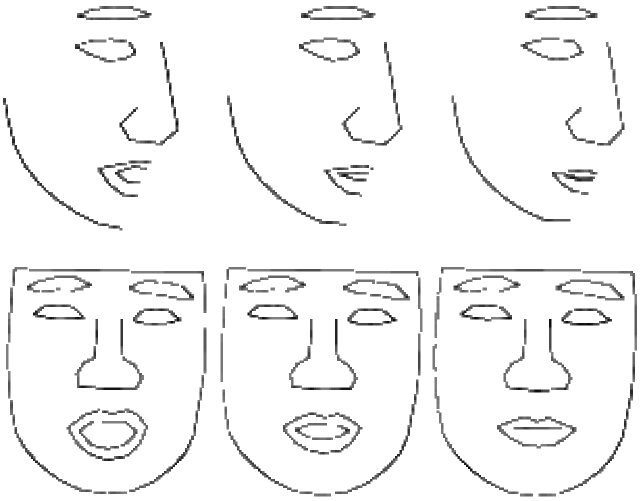


Figure 6 AAM side and front shape mesh contours and two sample variations along the first principle component (mode 1).

To fit a *P*-degree polynomial to the data we can write:

$$y_{ij} = a_{0j} + \mathbf{x}_i\mathbf{a}_{1j} + \mathbf{x}_i^2\mathbf{a}_{2j} + \cdots + \mathbf{x}_i^P\mathbf{a}_{Pj} \qquad (9)$$

where x_i^k denotes *element-by-element* exponentiation by *k*, y_{ij} is the j^{th} element of the i^{th} sample vector $\mathbf{y}_i$, and a_{oj}, $\mathbf{a}_{ij}$ are the coefficients that determine the polynomial.

If we let $\mathbf{y}_{\cdot j}$ denote the j^{th} column of the matrix Y, which has *M* rows equal to the stack of $\mathbf{y}_i$ row vectors, then we can write in matrix notation:

$$\begin{pmatrix} y_{1j} \\ y_{2j} \\ \vdots \\ y_{ij} \\ \vdots \\ y_{Mj} \end{pmatrix} = \begin{pmatrix} 1 & \mathbf{x}_1 & \mathbf{x}_1^2 & \cdots & \mathbf{x}_1^P \\ 1 & \mathbf{x}_2 & \mathbf{x}_2^2 & \cdots & \mathbf{x}_2^P \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \mathbf{x}_i & \mathbf{x}_i^2 & \cdots & \mathbf{x}_i^P \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \mathbf{x}_M & \mathbf{x}_M^2 & \cdots & \mathbf{x}_M^P \end{pmatrix} \begin{pmatrix} a_{0j} \\ \mathbf{a}_{1j} \\ \mathbf{a}_{2j} \\ \vdots \\ \mathbf{a}_{Pj} \end{pmatrix}$$

or alternatively,

$$\mathbf{y}_{\cdot j} = X\mathbf{a}_{\cdot j} \qquad (10)$$

The least squares (LS) solution to the system in Equation 10 minimizes the residual error $\left\| y_{\cdot j} - Xa_{\cdot j} \right\|$ and is given by:

$$\mathbf{a}_{\cdot j} = (X^T X)^{-1} X^T \mathbf{y}_{\cdot j} \qquad (11)$$

for $j = 1, \cdots, B$.

Tikhonov regularization can be employed to reduce the effects of noise in the data and numerical instability related to small singular values of A. In essence, we parameterize the solution to the system in Equation 10 obtaining a balance between trying to fit the data (i.e., reduce the residual error of the solution) and constraining the solution to a minimal norm. The *regularization parameter* μ determines this balance and the solution becomes:

$$\mathbf{a}_{\cdot j}^{(\mu)} = \arg\min\{\| \mathbf{y}_{\cdot j} - X\mathbf{a}_{\cdot j} \| + \mu \| \mathbf{a}_{\cdot j} \|\} \qquad (12)$$

It can be shown that for a given μ the solution that minimizes Equation 12 is:

$$\mathbf{a}_{\cdot j}^{(\mu)} = (X^T X + \mu I)^{-1} X^T \mathbf{y}_{\cdot j} \qquad (13)$$

One common method to choose the value of μ is to select the value that minimizes the *generalized cross-validation* (GCV) defined by:

$$V(\mu) \equiv \frac{\| \mathbf{y}_{\cdot j} - X\mathbf{a}_{\cdot j} \|}{[Tr(I - X(\mu))]^2} \qquad (14)$$

where $X(\mu) = XX^T(XX^T + \mu I)^{-1}$.

5. Experimental results

In this section, we first introduce the results of the AAM modeling on each view of the face and then follow with a discussion of the quantitative results for the parameter estimation.

5.1. AAM Models

An AAM of the frontal face and the side view images was built for each subject as described in Section 4. It was built using 8 frames out of a 71 frame video stream per view. The subsets were spaced at 10 frames apart and the AAM was built to capture 99% of the variation when PCA was applied. This reduced the representation of the face to a model parameterized by only 6-7 coefficients (i.e., we are able to synthesize an image of the face for each of the 71 frames with at most 7 floating point numbers). Figure 9 presents samples of the synthesized faces of two subjects as well as the original frames. The synthesized images are very similar to the original images and they properly convey the facial expressions of the subjects.

5.2. Frontal Parameter Estimates

A *leave-one-out* approach was followed to estimate the residual differences reported in this section. This is done to properly estimate the expected error for unseen images (i.e., images that weren't used to find the solution) and avoid overfitting the data. Two basic measures are reported here: the residual differences of the parameters of the AAM models, $\left\| y_{\cdot j} - X a_{\cdot j} \right\|$, and the *peak signal-to-noise-ratio* (PSNR), which provides a standard measure of similarity between the originally synthesized frontal image and the one synthesized by estimating the parameters using the fitted polynomial.

The PSNR between two $M \times N$ grayscale images, I and $\hat{I}$, is given by:

$$PSNR = 20 \log_{10} \frac{255}{RMSE} \ dB \tag{15}$$

where

$$RMSE = \frac{\sum_{i=1}^{M} \sum_{j=1}^{N} (I(i,j) - \hat{I}(i,j))^2}{NM} \tag{16}$$

In Table 1, it is shown how the residual error for the first coefficient tends to zero as the degree of the polynomial used is increased, while the estimated error using the leave-one-out approach starts increasing after polynomial degree 2. This is an indication that for the limited amount of data that we are currently using (i.e., 71 frames) we can not apply a polynomial fit with degree over 2 and care has to be taken not to be misled by the absolute residual difference.

Error Method	POLYNOMIAL DEGREE			
	1	2	3	4
Absolute Error	0.0019	0.0011	0.0009	0.0005
Leave-one-out	0.0024	0.0023	0.0070	0.0160

Table 1 Absolute residual error vs. leave-one-out estimate for parameter one.

In Table 2, a summary of the residual differences for the first two parameters and the PSNRs is presented. The effect of increasing the polynomial degree without providing enough data is clearly observed for the least squares (LS) solution where the error mean (μ) and the standard deviation (σ) steadily increase. It shows that the regularized least squares (RLS) solution does not blindly rely on the data and therefore is more robust to noise (e.g., outliers) and avoids overfitting the data.

Figure 7 shows the PSNR average +/- the standard deviations for the LS solution and RLS solution as a function of the polynomial degree. It can be observed how the regularized approach has slightly higher PSNR values and partially overcomes the over fitting problem, while the LS approach has a faster decreasing average PSNR and increasing standard deviation.

It should be noted that although the differences in PSNR are not substantial, they are very significant. They should not be disregarded as insignificant, given that as more variation is introduced to the AAM model the number of coefficients necessary to parameterize the face will increase and more ambiguity will be present in the mapping, making these gaps larger. Also, notice how in Figure 8 the image generated by the LS approach is highly distorted, while the one synthesized by RLS is much smoother.

Finally, Figure 9 shows, in the last two rows, the originally synthesized frontal image and the one synthesized by estimating the parameters using a polynomial of degree 2.

SUBJECT I

Method	PSNR (dB)		Coeff 1		Coeff 2	
	μ	σ	μ	σ	μ	σ
LS1	39.55	**2.98**	0.0024	0.0040	**0.0003**	0.0006
LS2	40.26	3.67	0.0022	0.0040	**0.0003**	**0.0005**
LS3	39.70	4.12	0.0069	0.0274	0.0011	0.0060
LS4	39.64	5.02	0.0161	0.0834	0.0021	0.0117
LS5	38.40	6.43	0.2659	1.6369	0.0201	0.1263
RLS1	39.76	3.11	0.0023	0.0037	**0.0003**	0.0006
RLS2	**40.38**	3.39	**0.0019**	**0.0031**	**0.0003**	**0.0005**
RLS3	40.06	3.82	0.0049	0.0214	**0.0003**	**0.0005**
RLS4	40.35	3.99	0.0041	0.0219	**0.0003**	0.0011
RLS5	39.92	4.41	0.0109	0.0539	0.0017	0.0068

Table 2 Results for one subject (other subjects follow similar patterns) of the PSNR and Parameter Residuals for the first two parameters shown for the LS and RLS solutions. The number next to LS and RLS indicates the polynomial degree.

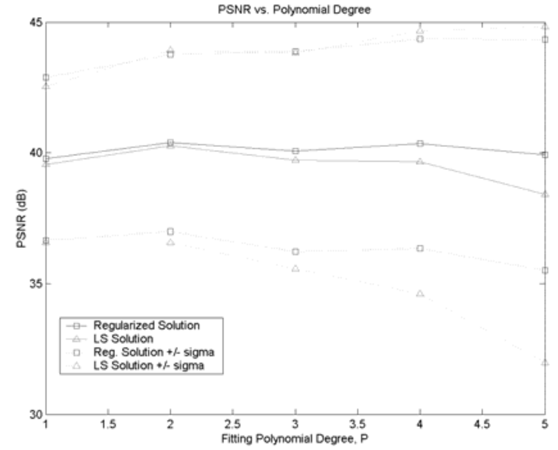


Figure 7 PSNR mean and standard deviation plot.

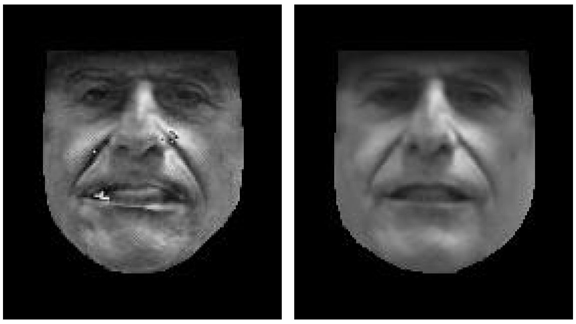


Figure 8 Outlier Effect on LS. It is shown how LS1 (left) is more susceptible to outlier effects than RLS1 (right).

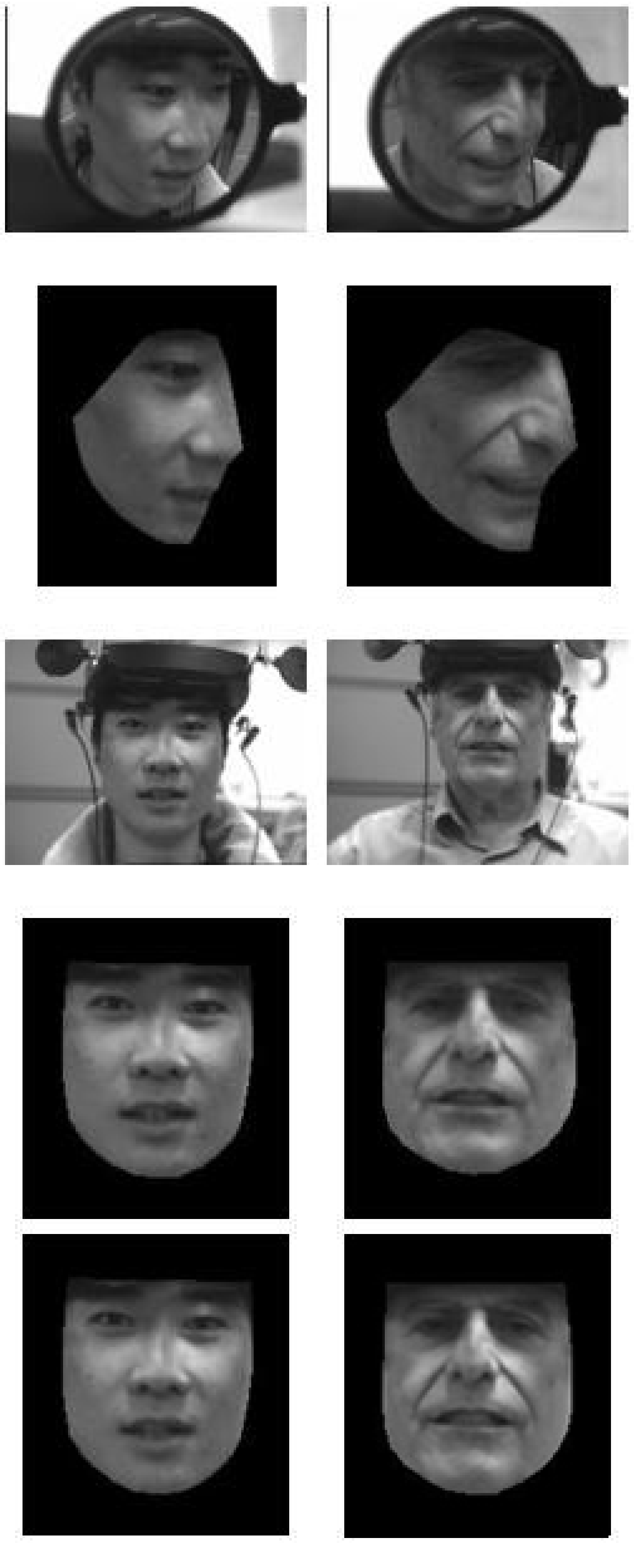


Figure 9 Samples of two side (top) and frontal (center) original views and the AAM model instantiation for each and the respective estimated images (bottom).

6. Concluding discussion

We have designed a system to capture a video stream of two side views of the face using the MFCS and a supplementary view from a third camera used only during the training stage. By modeling the three views of the face using an AAM and finding the regularized solution to the mapping between the two side views and the supplementary view, we can estimate the parameters for this missing view at run time.

By solving this problem using a statistical generative method, we avoid the difficulties associated with blending the two separate images and pose estimation. Our generated videos are of good subjective quality and maintain a high fidelity, about 40 dB PSNR, between the original model and the estimated one. Furthermore, we have a completely automatic system at run time. The AAM’s and the linear regularization techniques employed have proven to be efficient maintaining this application in the real time domain. We conclude that our MFCS and mathematical methods support the intended collaborative distributed applications.

Implemented in a full mobile system, this approach offers the possibility of communicating the full facial expression of a mobile user anywhere and anytime when higher levels of social presence are needed for example emergency, affective, or procedural communication. It is important to note that our frontal videos are generated from video frames taken during training. While this is sufficient for communicating the state of mind of the collaborator, it is not a video or telepresence system. The face is reconstructed from an analysis of changing parameters. For example, it cannot communicate aspects of the current environment; for example, the reflection of a fire on a firefighter’s face. Future research will investigate methods to blend in the environmental lighting, when needed. Future directions include generating full 3D models of the remote user’s face, creating a fully mobile prototype, adding temporal correlation information to the process of estimating the synthetic view parameters, and evaluating the effect of this additional social presence on user behavior during mobile collaboration and communication.

Acknowledgements

The authors would like to thank Timothy F. Cootes for making publicly available his AAM software implementation.

References

[1] H. Fuchs, G. Bishop, K. Arthur, L. McMillan, R. Bajcsy, S. Lee, H. Farid, and T. Kanade. Virtual Space Teleconferencing using a Sea of Cameras. In *Proceedings of the First International Symposium on Medical Robotics and Computer Assisted Surgery (MRCAS)*, Pittsburgh, PA, USA, September 1994, pp. 161-167.

[2] R. Raskar, G. Welch, M. Cutts, A. Lake, L. Stesin, and H. Fuchs. The Office of the Future: A Unified Approach to Image-Based Modeling and Spatially Immersive Displays. In *Proceedings of SIGGRAPH 98*, July 1998, pp. 179-188.

[3] L-Q. Xu, B. Lei, and E. Hendriks. Computer Vision for a 3D Visualization and Telepresence Collaborative Working Environment. In *BT Technology Journal*, 20:64-74, January 2002.

[4] H. Hua, L.D. Brown, C. Gao, and N. Ahuja. A New Collaborative Infrastructure: SCAPE. In *Proceedings of the IEEE Computer Society Conference on Virtual Reality (VR)*, Los Angeles, CA, USA, March 2003, pp. 171-179.

[5] F. Biocca and J.P. Rolland. Teleportal Face-to-Face System. US Patent Number 6,774,869, issued August 10, 2004.

[6] T.F. Cootes, G.J. Edwards, and C.J. Taylor. Active Appearance Models. In *Proceedings of the 5th European Conference on Computer Vision (ECCV)*, Freiburg, Germany, June 1998, pp. 484-498.

[7] T.F. Cootes, G.J. Edwards, and C.J. Taylor. Active Appearance Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 23(6):681-685, June 2001.

[8] T.F. Cootes and P. Kittipanya-ngam. Comparing Variations on the Active Appearance Model Algorithm. In *Proceedings of the British Machine Vision Conference (BMVC)*, Cardiff, UK, September 2002.

[9] G.J. Edwards, C.J. Taylor, and T.F. Cootes. Interpreting Face Images using Active Appearance Models. In *Proceedings of the 3rd International Conference on Automatic Face and Gesture Recognition (FG)*, Nara, Japan, April 1998, pp. 300-305.

[10] I. Matthews and S. Baker. Active Appearance Models Revisited. *International Journal of Computer Vision*, 60(2):135-164, November 2004.

[11] S.M. Seitz and C.R. Dyer. View Morphing. In *Proceedings of SIGGRAPH 96*, 1996, pp. 21-30.

[12] R.A. Manning and C.R. Dyer. Interpolating View and Scene Motion by Dynamic View Morphing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Ft. Collins, CO, USA, June 1999, pp. 1388-1394.

[13] Y. Wexler and A. Shashua. On the Synthesis of Dynamic Scenes from Reference Views. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Hilton Head, SC, USA, June 2000, pp. 1576-1581.

[14] C.K. Reddy, G.C. Stockman, J.P. Rolland, and F.A. Biocca. Mobile Face Capture for Virtual Face Videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshop (CVPRW)*, Washington, DC, USA, June 2004, pp. 77-83.

[15] F.G. hamza Lup and J.P. Rolland. Scene Synchronization for Real-Time Interaction in Distributed Mixed Reality and Virtual Reality Environments. *Special issue Collaborative Virtual Environments, PRESENCE*, 13(3):315-327, 2004.

A Study of Influential Factors on Effective Closely-Coupled Collaboration based on Single User Perceptions

Oliver Otto, Dave Roberts and Robin Wolff
The Centre for Virtual Environments,
University of Salford, Business House, UK
{o.otto, d.j.roberts, r.wolff}@salford.ac.uk

Abstract

The need for collaboration and cooperation across a distance is becoming equal to that between a co-located team. This means that cooperative working must be supported by systems that allow natural social human communication and interaction. A goal of tele-collaboration is to reproduce the effectiveness of co-located teamwork across a distributed team. Although many of today's systems support collaboration, only a subset can be said to support cooperation and no established solution comes close to reproducing the flexibility and efficiency of a face-to-face meeting. Previous work has demonstrated that accessing collaborative virtual environments through CAVE-like display systems provide a natural way of sharing space and objects within it that bring us closer to replicating a face-to-face meeting. Our previous research has demonstrated a link between level of immersion, task performance and feelings of collaboration and cooperation. It was, however, unclear if the advantage came from more natural interaction with objects or more natural interaction with others through objects.

The aim of this paper is to understand the impact of using a CAVE-like display has on user-to-object interaction so that we can isolate this from previous results to find the advantage given to collaboration. Task performance was measured and a questionnaire was used to identify the perceived impact of various display factors .The results from this study indicate that the major impact of immersion is on cooperative tasks. Results showed further a disparity between perceived and actual performance, which is discussed.

1. Introduction

In an increasingly global economy many people are under the pressure to expand collaboration from co-located to geographically distributed groups. Cooperation between people is often centred around common interests. This point of interest may be embodied by some perceivable object. In virtual reality, such objects may represent some physical artefact, information or concept from the real world. It is important for all collaborators to perceive and understand the object in order to work with it. While we cooperate with other people through an object, we use a variety of communicational tools to demonstrate our opinion, desire and intention to others. Be it simply verbally with emotional nuances, with gestures and postures in a non-verbal way or by manipulating the object directly. When interacting remotely, these forms of social human communication (SHC), as well as the representation of the object, need to be mediated through tele-collaboration technology.

If we use a phone or text to communicate, progress can be slow due to possible misunderstandings arising from cues that cannot be communicated through this medium. The use of modern video-conferencing systems gives us more flexibility and support for non-verbal communication, such as pointing towards object attributes. Using video-conferencing systems, however, one only "looks into each other's world", which limits the operating range to move and to be seen. In addition, it is difficult for all participants to interact with a shared object. In particular, it is hard to see how someone is interacting with an object when the operator, observer and object are each in separate windows, as in Access Grid [1].

A collaborative virtual environment (CVE) allows remote people and objects to be situated in a shared synthetic environment, in which one can navigate around and interact with a computer-generated representation of objects and other participants. Thus, whereas tele-conferencing systems allow people to look into each other's space, CVEs allow people and data to be situated in a shared spatial and social context.

In a previous study focusing on closely coupled collaboration in CVEs [2], we found that the exclusive use of spatially immersive (CAVE-like) displays significantly improved task performance and feelings of collaboration & cooperation (Figure 1, Table 1).

Table 1. Performance increase IPT / DT [2]

Sub-task	Description	Predominant activity	Performance increase IPT / DT
ST1	Place foot	Moving	48 %
ST2	Carry beam	Moving	35 %
ST3	Place beam	Positioning	73 %
ST4	drill hole	Use tool	44 %
ST5	Insert screw	Positioning	53 %
ST6	fix beam	Use tool	65 %
ST7	Place T joiner	Positioning	64 %
ST8	drill hole	Use tool	55 %
ST9	Insert screw	Positioning	65 %
ST10	fix T joiner	Use tool	65 %

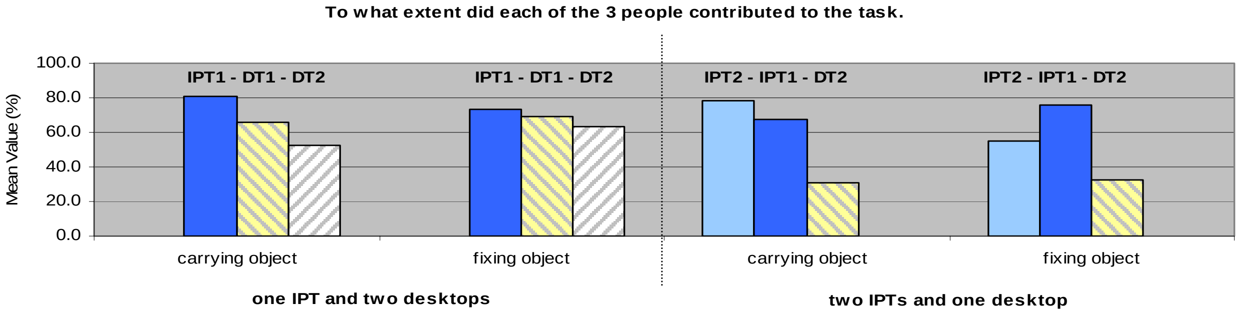


Figure 1. perceived performance results of teamwork study [2] comparing IPT and desktop displays

We were, however, uncertain if the improvements resulted from enhancements through user-to-object interaction or user-to-user collaboration around object interaction. To clarify this question, this study takes a closer look at display relevant factors, such as the field of view (FOV) and user interface, by performing similar tasks to previous trials, however excluding the social and team aspect. A single user trial with a number of volunteers was conducted, which measured the task performance, including time, task order and locomotion within the virtual environment. In addition the users were asked to fill out a questionnaire at the end of the test.

1.1. Related work

Hindmarsh et al. studied collaborative interaction of two users through a set of objects using a desktop based CVE [3], in which the participants were asked to rearrange furniture. The authors found that the limited field of view (FOV) on desktop systems was of great hindrance due to problems with fragmentation of the workflow. It took an unnaturally long time (>20sec) for users to perceive each other's gestures and to reference them to the places and objects in their conversation. The authors concluded that this was caused from a lack of information about other's actions due to their limited window into the world. In addition the study found problems with slow applications and clumsy movements as well as the lack of parallelism for actions. A subsequent study tried to resolve some of the issues with peripheral lenses, which resulted in an enhanced FOV. Although this solution enhanced the awareness, it also showed that peripheral lens distortion can disrupt both a user's own sense, and their notion of the other's sense, of orientation to actions and features within the environment [4].

Large displays are often not placed at a distance due to space constraints. They are typically relatively closer and cast a larger retinal image, thus offer a wider FOV. It is generally agreed that a wider FOVs can increase "immersion" in VEs [5-7]. Large displays in these settings are easy for all users to see and interact with, providing a conduit for social interaction [8], and some researchers have begun to document performance increases for groups working on large displays [9, 10].

Advances in immersive display devices are increasing their acceptance in industry and research [11]. Their support of natural body and head movements may be used to view an object from every angle. An object can be reached for and manipulated with the outstretched hand, usually through holding some input device. The feeling of presence, and particularly the naturalness of interaction with objects, may be improved when the user can see their own body in the context of the virtual environment [12]. Schuemie concludes that little is known about what interaction has to do with presence [13]. It may be argued that even less is known about the relationship between effective interaction on common objects as a focus of interest and co-presence.

Desktop systems use various methods to interact with objects in a virtual environment, such as go-go, ray casting or occlusion techniques [14, 15]. These can be used in CAVE-like displays, but have been primarily developed using head-mounted displays (HMD). Desktop systems use 2D interface controls or virtual spheres or mouse picking, whereas immersive displays normally use one- or two-handed direct manipulation (virtual hand) using a tracking system. Evaluations of interaction techniques for immersive displays found that the virtual-hand is superior to ray casting for the selection and manipulation of objects [15, 16]. The VR community is looking into the use of various displays for various tasks, yet is unable to define which choice to make for specific tasks [11]. Comparisons of usability have been made between immersive and desktop displays [17, 18] and they tend to show an advantage for immersion in certain applications.

Kjeldskov et al. [19] found that non-tracked 3D interaction devices work fine for orientating and moving when using partial immersive displays, but are problematic when using fully immersive displays. In addition they argue that partial and fully immersive displays have different support for close-by interaction (virtual hand) and different affordances for pointing (virtual beam). An experiment by Bowman et al. [20] showed that HMD users are significantly more likely to use natural rotation in a VE than CAVE users. This produces higher levels of spatial orientation, and can make navigation more efficient.

This paper extends a previous study [2] that analysed factors affecting a collaborative task. The study presented in this paper analyses a similar task carried out by a single user, so that factors affecting collaboration can be isolated. The aim is to understand what impact of using a CAVE-like display has on user-to-object interaction, so that we can isolate this from previous results that showed an improvement in multi-user cooperation through shared objects. This will tell us if the advantage comes from more natural interaction with objects or more natural interaction with other participants through objects. Section 2 introduces the task and the setup for the various displays. The results are given in section 3, discussed in 4 in relation to previous studies and summarised in section 5.

2. Experimentation

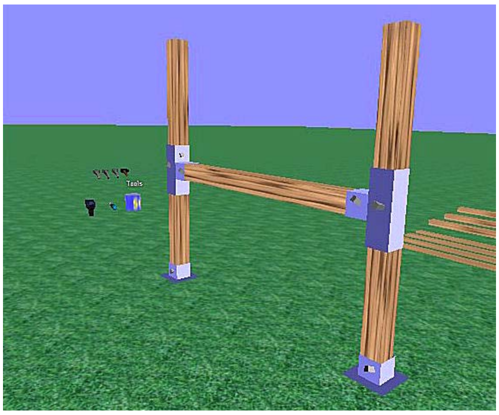


Figure 2. A simple structure to build

In order to understand how different display factors and interaction methods influence a task designed for closely-coupled collaboration, we modified our existing benchmark application, in which remote users are building a given structure (Figure 2) by interacting with simulated wooden beams, metal joiners, screws and a set of tools in a specific order [2, 21]. Objects have to be carried to the construction site and eventually fixed with the appropriate tools and materials. For example, a beam can be inserted into a metal joiner or foot and then fixed in place by drilling a hole and screwing in a screw. The original task required teamwork, as simulated gravity required two people to lift a beam and one person to hold a joiner to a beam while it was being fixed. The need for team work was removed by disabling the simulation of gravity. Our trial was then able to focus on single user interaction with objects. Clearly, interaction would be altered by the lack of gravity, but we considered the effect to be negligible.

2.1. Measurements

For this task we asked 13 student volunteers to participate, each received multiple training sessions to familiarise them with the interface and the task. Earlier trials showed that after three short training sessions the user became familiar with the interface so that their performance reached that of an expert user [2, 21]. The trials needed no longer than 5-10 min per session and display, compared to almost 30-45 min of training and familiarisation per subject.

After evaluation of the results we found a significant difference in measured and perceived performance, which we partially related to the manipulation and navigation on the desktop. To better understand this relationship, we performed a subsequent trial with four people repeating the desktop trial with ray-casting as well as virtual-hand manipulation.

2.2. Display Configuration

We asked all participants to perform this task on a variety of distinct display configurations: a non-immersive desktop system, a partial immersive workbench system and a fully-immersive CAVE-like system (see Table 3). Each trial was first undertaken on a desktop and then repeated on the workbench and in the CAVE-like display. It was assumed that, as participants were practiced in doing the task on them and the trials were short, order was unlikely to impact on results.

Figure 3. the workbench display

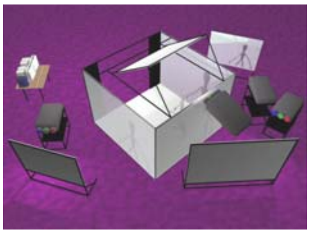


Figure 4. the CAVE-like display

DIVE [22] in version 3.3.5 was chosen as CVE platform for experimentation on all display devices, as it is an established benchmark [17, 23-27]. We extended this DIVE version with an event monitoring plugin that allowed us to monitor the user and object movements for a post-trial analysis.

Table 2. display configurations

Display device	Input device	OS	Stereo	Field of view	Manipulation technique
Desktop	keyboard and mouse	Linux	No	60 degree	ray-casting
Workbench (Figure 3)	tracked wand	Irix	Yes	110 degree	virtual hand
CAVE-like (Figure 4)	tracked wand	Irix	Yes	160 degree	virtual hand

Table 3. summery of some selected questions

Perception of	Desktop		Workbench		CAVE		
	mean	SD	mean	SD	Mean	SD	ANOVA
performance carrying (%)	68.8	12.5	75.3	11.2	81.4	13.6	F(2,29)= 2.70, MS_W=2.04, p=0.084
performance fixing (%)	74.3	13.1	77.9	14.8	91.4	10.0	F(2,28)= 4.96, MS_W=4.02, p=0.014
interface hamper (%)	61.0	23.1	49.4	20.6	37.7	17.2	F(2,30)= 3.59, MS_W=7.36, p=0.040
field of view (%)	53.2	26.4	68.8	14.0	77.9	24.2	F(2,30)= 3.47, MS_W=8.39, p=0.044
missing sense of touch (%)	41.6	28.2	54.5	21.0	59.7	27.7	F(2,30)= 1.44, MS_W=4.73, p=0.252
presence (%)	28.6	18.1	66.2	9.6	85.7	14.3	F(2,30)= 44.67, MS_W=45.48, p=0.000
measured task time (min)	6.1	1.4	7.0	1.8	7.3	2.1	F(2,30)= 1.33, MS_W=4.21, p=0.280

2.3. Questionnaire

Thirteen questions were asked, in which the user compared the different display combinations. Errors arising from a user's misinterpretation of a question were reduced by asking sets of related questions. Answers could be given on a Likert-type scale [28] of 1-7, where 1 represented agreement to a very small extent and 7 to a very large extent. The questionnaire included questions concerning how subjects interacted with the object in the different configurations, as well as how they perceived the interaction with the objects. The questions were similar to those asked in previous studies allowing us to compare our earlier work [2, 21], but were mainly related to performance, field of view and presence.

3. Results

This section documents the results of this study, comparing user performance, manipulation technique, FOV and presence. We first describe the questionnaire results and then the observations and measurements of two selected cases.

3.1. Overall Findings

For the analysis of the questionnaire we used the statistical approach of analysis of variance (ANOVA) to verify the significance of the results. The limit of significant deviance was α=0.05. The results are given with MS_W as the mean square within groups, F(a,b) as the variance between groups/MS_W and p as the actual deviance, with four decimal places. A posthoc Tukey test was applied if a significant difference could be found to verify where those differences appear.

We asked the users *"how well they performed the task of carrying / fixing an object using the different displays"* and an ANOVA showed that there is a significant difference between the desktop and the immersive displays (performance carrying and fixing: F(2,60)= 7.25, MS_W=5.80, p=0.002). On a desktop, performance was perceived to be less effective than it was in the CAVE or workbench. In addition, this contrast was stronger for fixing an object than for carrying it (Figure 5, Table 3).

The question of "*how much did the interface hamper the task*" showed a clear difference between desktop and immersive displays. The keyboard/mouse combination on the desktop system with its, for CVEs typical, complicated combination of shortcuts was clearly perceived to hamper the task much more than the tracking / joystick combination in the CAVE or workbench (Figure 5, Table 3).

Another question was "*how important was the field of view during the interaction*" and again a clear difference can be seen between the desktop and the immersive display (Figure 5, Table 3).

None of our displays had a haptic interface and when asking: "*how much did you miss the feel of touch*" it showed that it was missed more within the immersive displays than at the desktop system. One of the users expressed it as: "The sense of touch was not expected when using the desktop, whereas it was when on the workbench and particularly in the CAVE." (Figure 5, Table 3).

The results above show that the user in the immersive display felt more natural and present in the task. This was confirmed by their answer to our questions regarding presence. The questions "*of their sense of being there*", "*realistic appearance of interaction*" and "*feeling of physical space*" show all a very low perception of presence on the desktop, but a high perception on the immersive displays (Figure 5, Table 3).

Although the questionnaire was used to measure the user's perception of their performance, the time taken by each subject to complete a test-run was taken independently. The performance measured by time appeared to contradict the subject's perception measured from the questionnaire, as shown in Table 3. Average task completion times were 370, 410 and 445 seconds for desktop, workbench and cave respectively. An ANOVA for the measured time showed no significant difference for any of the displays (F(2,30)= 1.33, MS_W=4.21, p=0.280).

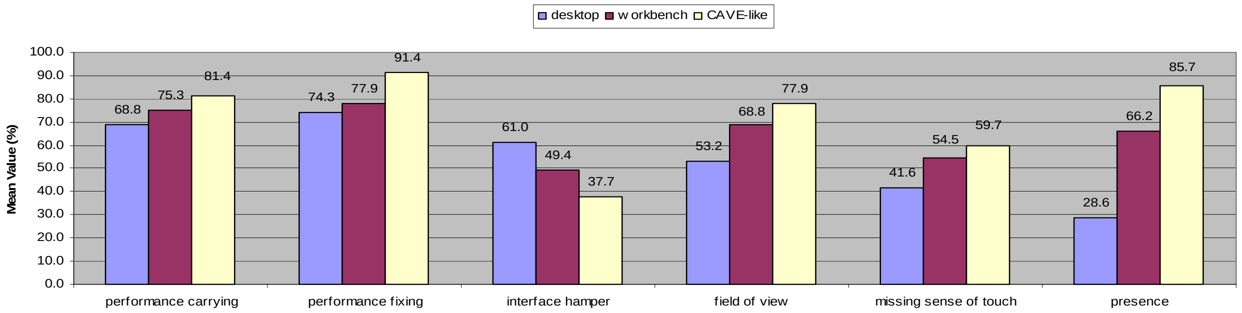


Figure 5. summery questionnaire overview to perception of …

3.2. Comparing two extremes in Detail

The results above show that the users perceived the use of immersive displays as more efficient and suitable than the use of a desktop display. However, these results contradict the task performance measurements. The average time to complete the task was similar for each display, but slightly proportional to the level of immersion. We will discuss this contradiction later in this paper, but first we will look at two opposite cases (Table 4). In the first case (case1) the user had an equally fast time on all displays and in the second case (case2) the desktop time was faster than on the immersive displays. The main difference between the two has been observed in how they used the display interfaces. The former was taking advantage of the display's properties (movability, view frustum, interaction technique), whereas the later used all displays as if he was fixed in his position (Figure 9, Figure 11, Table 5).

The CVE platform used in this trial allows manipulation of objects through ray-casting on the desktop display, whereas a user must physically reach for an object before it can be manipulated through the immersive displays. This has the effect that the desktop user can manipulate objects from a distance, whereas in the immersed setting they must first approach the object. The advantage on the desktop is an apparent increase in the "field of view" when the building site is viewed from a distance. However, this would only work well in an open environment, as it is the case in the experimental setting in this trial. In a normal sized room, surrounded by walls, it would be difficult to see the whole room, and subsequently this would make it necessary to turn around. The effect of a large open environment can be seen in Figure 8a (traces show navigation through environment) where the desktop user moved very little and performed the object manipulation from a distance.

In contrast, in our configuration, the immersive displays required direct manipulation, hence the large amount of movements for both users in Figure 8b and Figure 8c. In addition, a larger amount of movements in a contained space have been recorded for the CAVE-like display compared to the workbench. From observations, we believe this to be due to the difference in modes of interaction across the display types. The CAVE-like display was a 3x3 m room in which the user can freely walk due to the tracking of the body, allowing natural precise and fast movements around an object, if it is close enough (within the 3x3m). This includes the ability to swing the body around, using peripheral vision and eye cascades to control an effective turn to an object of interest, when displayed on another projection wall. The joystick controller is only needed for larger movements. In contrast, on the workbench the user is more restricted (space of 1x1.5m) by the physical space as well as the smaller FOV, making it necessary to use the joystick controller more often for navigation. This can be seen in comparing the fairly straight lines of Figure 8b (using joystick navigation) with curved lines of Figure 8c (user walking within the spatial display). The figures 8a-8c show in addition that the user of case1 is moving less and shorter than the user of case2. This is in harmony with the observations that in case2 the joystick was used more often than in case1, where the user made more use of his physical space to move. The result is an increase of measured completion time of the task for case2.

Table 4. comparison of two opposite cases, using 7-point Likert-type scale

perception of	Desktop	Work-bench	CAVE-like
	case1		
measured task time	6 min	6 min	6 min
main observations	- good use of all walls in the CAVE - "10min ago I was working on the wall, now I am in the middle and that makes a difference"		
	case2		
measured task time	6 min	7 min	9 min
main observations	- a mental picture of the scene seems to be missing - stayed in one place in the immersive display, but lots of joystick movements		

Observations have shown that taking advantage of the natural interface of the immersive display it could increase the feeling of presence and performance (Table 4) as well as reducing the frustration factor, because one may "overshoot" the target when trying to get there with the joystick. Similar observations have been made in previous

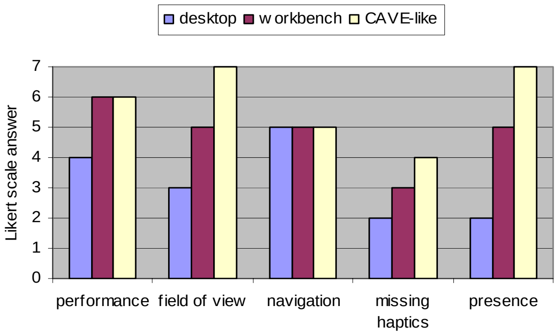


Figure 6. user perception of case1 (with observed flexible interaction pattern)

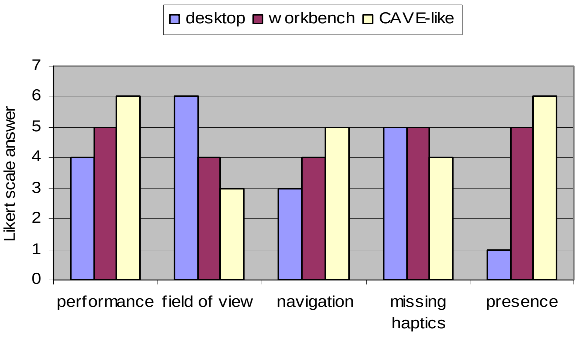
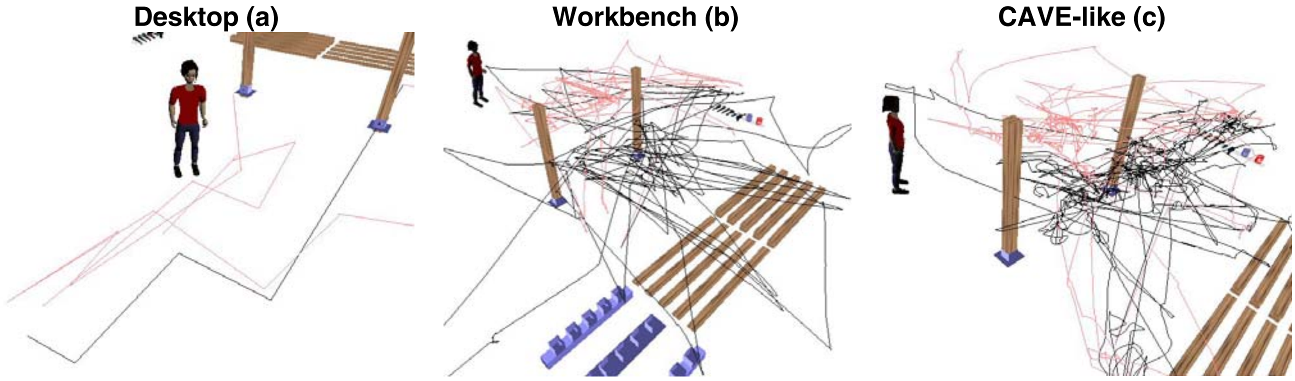


Figure 7. user perception of case2 (with observed ridged interaction pattern)


Desktop (a)

Workbench (b)

CAVE-like (c)

Figure 8. traces of the moving avatar during the task, case1: dark line, case2: bright line

trials during closely-coupled interaction [21], where overshooting led to some observed distress when a user needed more time to adjust their position. Thereby the other user had to wait if one's action was needed to finish a cooperative subtask.

4. Discussion

This section discusses why perceived and measured performance was different, what the FOV has to do with user locomotion & navigation and why the interaction technique influences the user collaboration and performance.

One clear observation was the difference between the perceived performance and the time needed to complete the task (see Figure 5, Table 3). The contradiction of these results may be explained through the relationship of the perception of being there, immersiveness and interaction technique (Figure 9). The results of this study (Figure 5) show significant differences in perceived presence for all displays. The same tendency can be seen for performance, FOV, missing touch and interface problems. Although those tendencies are not as strong as for presence, they show that the more one becomes immersed and engaged the higher is the feeling of being there. Presence is not something that can be clearly measured, but is a feeling created by a number of factors [29]. Those factors, like immersiveness, naturalness of interface and ease of interaction, all appear to contribute to a feeling of being there. Small differences of perception (between displays) for all those factors have a profound influence on the perceived presence. This also explains the difference between perceived and measured performance. If one feels more engaged and present, time will seem to pass quicker and the user's own activity will enhance the feeling of performance. This can also be seen in the reaction of users, who consistently mentioned that the use of the immersive display was much more enjoyable than the desktop.

One objective in this study was to determine how much the FOV would influence task performance. Our hypothesis was that with a wider view frustum the task would become easier and increase performance as the scene is more visibly accessible and therefore objects can be faster spotted. In contrast to the desktop, both immersive displays are similar in the way the user interacts, however the FOV is their main difference. Differences can be seen in the data and observations gathered during this trial. At the workbench, Figure 8b shows clearly longer ways for locomotion in comparison to Figure 8c. In addition, the observation during the trial was that on the workbench the joystick was used more often to attain an object as compared to the CAVE, where physical walking toward an object was easier and only longer distances needed the use of the joystick (Figure 11a and Figure 11b).

From observations, we estimate the relationship between FOV and locomotion as a curve as shown in Figure 11b. An exception is the HMD, which has natural rotation (360°), independent of the FOV. This means that

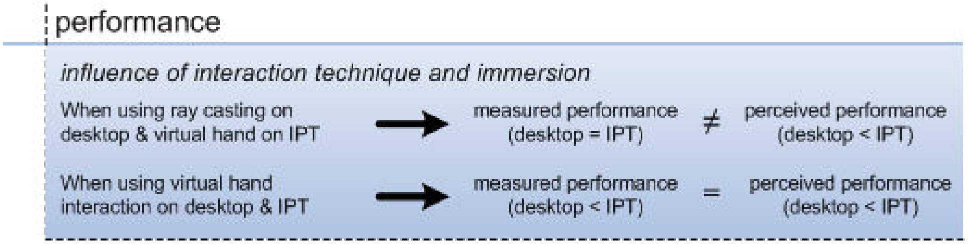


Figure 9. correlation diagram of perceived and measured performance

with an HMD the user may not need to use a joystick to rotate, but rather uses its own body [30]. In contrast, the desktop has the smallest FOV of all the tested displays, yet the locomotion recorded during the trial was very low. The reason for this appears to be based upon the ray-casting manipulation of objects. The user did not need to get close to the object, but could do everything from a remote place, from which the whole scene could be observed. However, in previous trials this behaviour was reason for complaint as other collaborating users could not see the correlation between a user and the object they were interacting with [3, 21]. In addition, working from a remote place is only possible if the given environment supports such behaviour, for example - a world without walls or very large rooms.

Figure 10. correlation diagram of manipulation technique

Therefore, in a subsequent trial to this study we asked users to repeat the task on the desktop, first from a remote location (using ray-casting) and second from a location close to the object (virtual-hand). The result was that the time taken to perform the task doubled for the close-up trial (mean of 9.4min). Therefore we can hypothesise that if we try to improve the collaboration between users by allowing only close-object interaction, time-performance for desktop user will drop due to their limitation in FOV hence resulting in extended locomotion time to orientate (see Figure 9). In addition, a study from Steed et al. [16], that compared ray casting and virtual-hand interaction on HMD and CAVE displays, found that virtual-hand is superior for selection and manipulation of objects (Figure 10).

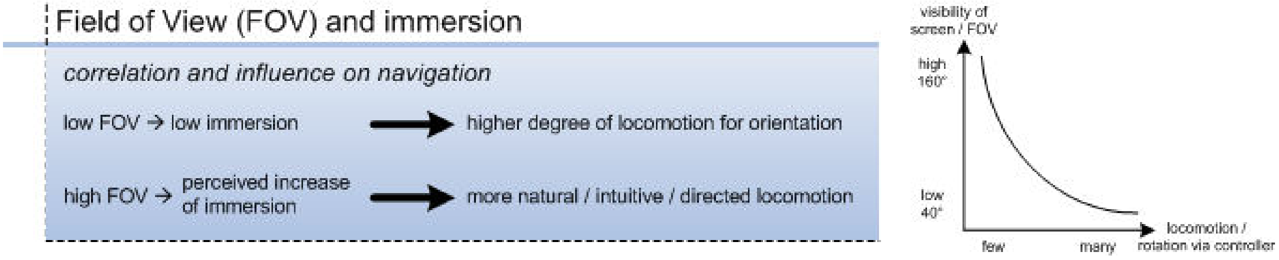


Figure 11 a & 11 b. correlation diagram of field of view and its influence on navigation

This study looked into influences on a single user task. Those influences sustain in a co-presence situation and may even enhance. For example, problems with interface and manipulation of objects can interrupt the workflow in a closely-coupled situation [31]. The previous studies showed that people have a higher perception of the performance of an immersed user, independent of the assessment of themselves or others [2]. They also show a significant difference between two immersed users and a desktop user, which was related to the ease of manipulation and navigation.

Conclusion

The measurement of performance is always difficult to achieve, as it depends on the way we measure and how measurable a task is. This applies as well for performance in a virtual environment. We may be able to measure the time it takes to finish a task, but as this study shows this is not necessarily reconciled with perceived performance.

In previous studies we *measured* an increase of performance in a collaborative task for CAVE-like displays [2], yet no such difference could be measured on a single user task. At the same time both showed an increase of *perceived* performance. Since the display and application properties were identical for the studies, it can be concluded that the *measured performance increase is due to the collaboration*. It seems that CAVE-like displays are better at representing contribution of others, but can trick a single user into thinking they are achieving more than they truly are.

This study has shown that different factors lead to an increasing perception of presence and performance. Factors, such as FOV, manipulation technique and navigation, may as well influence a user's interaction and its effect on other participants in a collaborative task (e.g. no fragmented workflow).

Our studies have focused on a structured task designed for closely coupled collaboration; it remains to be seen how our results affect other task designs.

References

[1] "Access Grid" URL:www.accessgrid.org, 2005.

[2] Roberts, D. J., Wolff, R., and Otto, O., "Constructing a Gazebo: Supporting team work in a tightly coupled, distributed task in virtual reality", *Presence: Teleoperators and Virtual Environments*, vol. 12 (6), pp. 644-668, 2003.

[3] Hindmarsh, J., Fraser, M., Heath, C., Benford, S., and Greenhalgh, C., "Object-Focused Interaction in Collaborative Virtual Environments", *ACM Transactions on Computer-Human Interaction (ToCHI)*, vol. 7 (4), pp. 477-509, 2000.

[4] Fraser, M., Benford, S., Hindmarsh, J., and Heath, C., "Supporting Awareness and Interaction through Collaborative Virtual Interfaces", *In Proceedings of 2th Annual ACM Symposium on User Interface Software and Technology (UIST'99)*, pp. 27-36, 1999.

[5] Lapointe, J.-F. and Vinson, N. G., "Effects of Joystick Mapping and Field-of-View on Human Performance in Virtual Walkthroughs", *In Proceedings of the 1st International Symposium on 3D Data Processing Visualization and Transmission*, Padova, Italy, pp. 490-493, 2002.

[6] Arthur, K., "Effects of field of view on task performance with head-mounted displays", *In Proceedings of Human Factors in Computing Systems*, Vancouver, Canada, pp. 29-30, 1996.

[7] Tan, D. S., Gergle, D., Scupelli, P. G., and Pausch, R., "With Similar Visual Angles, Larger Displays Improve Spatial Performance", *In Proceedings of Conference on Human factors in Computing Systems (CHI)*, Ft. Lauderdale, Florida, USA, pp. 217-224, 2003.

[8] Guimbretière, F., "Fluid interaction for high resolution wall-size displays". Stanford University, Stanford, CA, 2002.

[9] Dudfield, H. J., Macklin, C., Fearnley, R., Simpson, A., and Hall, P., "Big is better? Human factors issues of large screen displays with military command teams", *In Proceedings of People in Control*, pp. 304-309, 2001.

[10] Mortensen, J., Vinagayamoorthy, V., Slater, M., Steed, A., Lok, B., and Whitton, M. C., "Collaboration in Tele-Immersive Environments", *In Proceedings of Eighth Eurographics Workshop on Virtual Environments*, Barcelona, 2002.

[11] Brooks, F. P., "What's Real About Virtual Reality?" *IEEE Computer Graphics & Applications*, vol. 19 (6), pp. 16-27, 1999.

[12] Mine, M. R., Brooks, F. P. J., and Sequin, C. H., "Moving Objects In Space: Exploiting Proprioception In Virtual-Environment Interaction", *In Proceedings of SIGGRAPH 97*, Los Angeles, CA, 1997.

[13] M.J.Schuemie, P. v. d. S., M.Krijn. C.A.P.G.van der Mast, "Research on Presence in VR: a Survey", *Cyberpsychology and Behavior*, vol. 4 (2), pp. 183-202, 2001.

[14] Bowman, D. A., Johnson, D. B., and Hodges, L. F., "Testbed Evaluation of Virtual Environment Interaction Techniques", *Presence: Teleoperators and Virtual Environments*, vol. 10 (1), pp. 75-95, 2001.

[15] Poupyrev, I., Weghorst, S., Billinghurst, M., and Ichikawa, T., "Egocentric object manipulation in virtual environments: empirical evaluation of interaction techniques", *Computer Graphics Forum*, vol. 17 (3), pp. 41-52, 1998.

[16] Steed, A. and Parker, C., "Evaluating Effectiveness of Interaction Techniques across Immersive Virtual Environment Systems", *forthcoming in Presence: Teleoperators and Virtual Environments*, vol. 14 (5), 2005.

[17] Schroeder, R., Steed, A., Axelsson, A. S., Heldal, I., Abelin, A., Widestroem, J., Nilsson, A., and Slater, M., "Collaborating in networked immersive spaces: as good as being there together?" *Computers and Graphics*, vol. 25, pp. 781-788, 2001.

[18] Bowman, D. A., Gabbard, J. L., and Hix, D., "A Survey of Usability Evaluation in Virtual Environments: Classifcation and Comparison of Methods", *Presence: Teleoperators and Virtual Environments*, vol. 11 (4), pp. 404 - 424, 2002.

[19] Kjeldskov, J., "Interaction: Full and partial Immersive Virtual Reality Displays", *In Proceedings of IRIS24*, pp. 587-600, 2001.

[20] Bowman, D., Datey, A., Ryu, Y., Farooq, U., and Vasnaik, O., "Empirical Comparison of Human Behavior and Performance with Different Display Devices for Virtual Environments", *In Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, pp. 2134-2138, 2002.

[21] Roberts, D., Wolff, R., Otto, O., Kranzlmueller, D., Anthes, C., and Steed, A., "Supporting Social Human Communication between Distributed Walk-in Displays", *In Proceedings of VRST 2004*, pp. 81-88, 2004.

[22] Carlsson, C. and Hagsand, O., "DIVE - A platform for multi-user virtual environments", *Computers & Graphics*, vol. 17 (6), pp. 663-669, 1993.

[23] Frécon, E., Smith, G., Steed, A., Stenius, M., and Stahl, O., "An Overview of the COVEN Platform", *Presence: Teleoperators and Virtual Environments*, vol. 10 (1), pp. 109 - 127, 2001.

[24] Frécon, E. and Stenius, M., "DIVE: A Scalable network architecture for distributed virtual environments", *Distributed Systems Engineering Journal (special issue on Distributed Virtual Environments)*, vol. 5 (3), pp. 91-100, 1998.

[25] Greenhalgh, C. M., Bullock, A., Frécon, E., Lloyd, D., and Steed, A., "Making Networked Virtual Environments Work", *Presence: Teleoperators and Virtual Environments*, vol. 10 (2), pp. 142-159, 2001.

[26] Mortensen, J., Vinayagamoorthy, V., Slater, M., Steed, A., Lok, B., and Whitton, M. C., "Collaboration in tele-immersive environments", *In Proceedings of Proceedings of the workshop on Virtual environments 2002*, Barcelona, Spain, pp. 93 - 101, 2002.

[27] Steed, A., Mortensen, J., and Frécon, E., "Spelunking: Experiences using the DIVE System on CAVE-like Platforms", in *Immersive Projection Technologies and Virtual Environments*. Wien: Springer-Verlag, 2001, pp. 153-164.

[28] Sitzman, K., "Likert type questions may be helpful for baseline inquiry", *AAOHN Journal*, vol. 51 (7), pp. 320, 2003.

[29] Slater, M., "Measuring Presence: A Response to the Witmer and Singer Presence Questionnaire", *Presence: Teleoperators and Virtual Environments*, vol. 8 (5), pp. 560-565, 1999.

[30] Bowman, D., Datey, A., Ryu, Y., Farooq, U., and Vasnaik, O., "Empirical Comparison of Human Behavior and Performance with Different Display Devices for Virtual Environments", *In Proceedings of Human Factors and Ergonomics Society Annual Meeting*, pp. 2134-2138, 2002.

[31] Hindmarsh, J., Fraser, M., Heath, C., and Benford, S., "Virtually Missing the Point: Configuring CVEs for Object-Focused Interaction", in *Collaborative Virtual Environments. Digital Places and Spaces for Interaction*. London, UK: Springer Verlag, pp. 115-139, 2001

Social Presence in Two- and Three-dimensional Videoconferencing

J. Hauber[1] H. Regenbrecht[2] A. Hills[2] A. Cockburn[1] M. Billinghurst[1]
[1]University of Canterbury, Christchurch, New Zealand
[2]University of Otago, Dunedin, New Zealand
joerg.hauber@hitlabnz.org, holger@infoscience.otago.ac.nz

Abstract

New three-dimensional videoconferencing systems are trying to overcome the artificial nature of two-dimensional desktop video conferencing. The phenomenon of Social Presence serves as a measure of how natural persons feel when they are connected with a distant other through a telecommunication interface. We present a study measuring the difference in Social Presence in three conditions: (1) desktop 2D videoconferencing, (2) desktop 3D videoconferencing, and (3) face-to-face communication in a real environment, each with three participants involved. We applied two Social Presence measurement scales in an experiment with 42 subjects and found that in one scale Social Presence is rated higher in 3D than 2D. Further results are discussed in the paper.

Keywords--- Social Presence, Co-Presence, Videoconferencing, Presence measurement

1. Introduction

The success of globally and locally distributed organizations heavily depends on their ability for remote collaboration. Therefore, videoconferencing (VC) technology plays an increasingly important role as it provides a rich communication environment in which a wide range of remote collaboration tasks can be successfully accomplished. The quality and the reliability of web based video conferencing tools has improved over recent years, aided by the explosion of the internet and advances in modern network technology. In order to save travel expenses and time, many organizations apply these tools on a global scale for online meetings and presentations, or simply to keep in touch.

However, compared with real face-to-face conversation, communicating through conventional videoconferencing tools is an artificial experience. This is due to the absence of eye-contact, lack of a shared social and physical context, and a limited possibility for informal communication. These mediated systems lack “media richness” and support for verbal and non-verbal communication [5].

Recently three-dimensional metaphors have been applied in videoconferencing applications in an attempt to simulate traditional face-to-face meetings. For instance, SmartMeeting [19] provides a highly realistic conference environment involving different virtual rooms with chairs, whiteboards, virtual multi-media projectors, and even an interactive chessboard. AliceStreet [1] makes use of a similar concept, although with a more minimalist virtual room design. Participants are represented here as rotating video planes sitting around a virtual table and watching each other or a shared presentation screen. Finally, in “cAR/PE!” [15] participants can even freely move within the virtual environment and are able to place and discuss 3D models on top of the virtual table. The common goal of all of these approaches is to improve the usability of remote collaboration systems by decreasing the artificial character of a remote encounter. This goal seems to be of particular importance for the acceptance of these systems, as Biocca et al. point out:

> “*The assessment of satisfaction with entertainment systems and with productive performance in teleconferencing and collaborative virtual environments is based largely on the quality of the social presence they afford*“
>
> *Biocca et al., 2001*

In the study we are presenting in this paper we investigate if three dimensional interfaces are indeed able to shape a more natural sense of “being together” with remote others than traditional systems by comparing a three dimensional and a two dimensional video-conferencing interface with respect to their support for Social Presence.

Our study focuses on interactions involving three participants, in contrast to comprehensive studies with two participants only (e.g. [9]). We are assuming that this will lead to a deeper discussion of mediated multi-user communication where more than two people interact, which is common in real-world situations.

Additionally, we want to examine the power of two Social Presence measurement approaches in discriminating effects between different interfaces. This question is not trivial, as traditionally, Social Presence measurement instruments are applied in cross-media comparisons such as chat versus audio or audio only versus audio-video and not cross-interface comparisons.

2. Background: Social Presence

2.1 Definitions

Mediated Social Presence describes a feeling of togetherness of remote persons that are connected through some form of telecommunication medium. Definitions of Social Presence include the sense of "being together" [8], the sense of "Being There with others" [17], or the "perceptual illusion of non-mediation" [12]. According to Lombard and Ditton this illusion of non-mediation occurs when a person fails to perceive or acknowledge the existence of a medium in his/her communication environment. Consequently, unmediated face-to-face situations are considered the gold standard in Social Presence. The degree to what extent a telecommunication medium can support a feeling of Social Presence depends on the communication channels it provides but also on additional cues that an interface affords.

Although the presented definitions of Social Presence help to understand the concept, they are too general to derive some concrete measurement instruments. Existing Social Presence measures therefore are built on more advanced conceptualizations. However, it must be pointed out that these conceptualizations have to be seen in the context with their main concerns and emphases and thereafter lead to some inconsistencies that exist under the umbrella term Social Presence.

Because of this lack of a precise scope of definition, a promising approach is the definition of the term social presence through the validation of different instruments. This will probably lead to deeper insights into the phenomenon and eventually leads to a comprehensive understanding of the underlying concepts as well as a well-founded definition.

Therefore, in the following we outline two measurement instruments together with their definition of Social Presence.

2.2 Semantic Differential measure

In the 1970s, Short et al. were the first who defined Social Presence as the "degree of salience of the other person" [18] in their book "The Social Psychology of Telecommunication". Their work was funded and motivated by organizations such as the UK post office in order to determine the relative effectiveness of different media channels for social communication. Their focus therefore is on the medium and the attitude of customers towards the medium from a market analysis point of view. They regard Social Presence as a stable subjective quality of the medium, assuming that every user of any given communications medium is in some sense aware of the degree of Social Presence it supports. This "mental set" towards the medium in turn affects the user's nature of interaction and for example the user's media selection.

In Short's approach, the preferred method for measuring Social Presence in the laboratory is the semantic differential technique [14]. Participants are asked to rate telecommunication systems on a series of seven-point, bipolar pairs such as "impersonal – personal", "cold – warm", and "insensitive – sensitive".

Media having a high degree of Social Presence are typically rated as being warm, personal, sensitive, and sociable. This approach is still the most common way of measuring Social Presence and it has been applied in many studies.

2.3 Networked Minds measure

A more recent theory of Social Presence is given by Biocca, Harms, and Gregg [2]. Their main unit of analysis is the perceived access to another intelligence. They define mediated Social Presence as "the moment-by-moment awareness of the co-presence of another sentient being accompanied by a sense of engagement with the other… As a global, moment-by-moment sense of the other, Social Presence is an outcome of cognitive stimulations (i.e. inferences) of the other's cognitive, emotional, and behavioral dispositions". Based on a comprehensive literature review, they identify "Co-Presence", "Psychological Involvement" and "Behavioral Engagement" as the theoretical dimensions of Social Presence and name their empirically determined factors (figure 1).

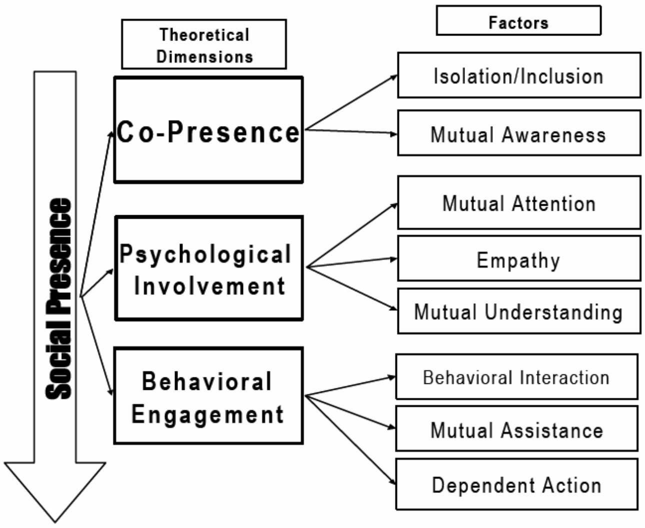


Figure 1: Factor structure of the Networked Minds measure of Social Presence [2]

Factor scale	Items	Example
Isolation/ Inclusion	2	"I often felt as if I was alone"
Mutual Awareness	4	"I hardly noticed another individual"
Mutual Attention	8	"I paid close attention to the other individual"
Empathy	6	"When I was happy, the other was happy"
Mutual Understanding	6	"The other understood what I meant"
Behavioral Interaction	6	"What I did affected what the other did"

Mutual Assistance	4	“My partner worked with me to complete the task”
Dependent Action	2	“The other could not act without me”

Table 1: Example items of the Networked Minds measure of Social Presence

The Networked Minds measure of Social Presence consists of a questionnaire which is built up by multiple items for each factor scale (see Table 1 for example items)

The important difference to Short’s concept of Social Presence is, that Biocca et al. understand Social Presence rather as a varying subconscious state of a person that is linked up with a distant other than a constant property of the medium that connects them. The questionnaire items target at the experience with the remote other as the main criterion and don’t assess a user’s subjective judgment about how well he or she thinks the medium supports Social Presence. This approach is more in line with other conventional subjective presence measures and promises a higher sensitivity and deeper insights in different points of interest in cross-media comparisons.

3. Method

We conducted a study to investigate how a two dimensional and a three dimensional videoconferencing interface affect the sense of Social Presence. We asked participants to work on a collaborative task in groups of three in three rounds with different conditions; one condition (FTF, figure 2) where they were collocated in one room and could talk to each other face to face, and two conditions (3D, 2D, figure 3 and 4) where they were located in separate rooms, connected though either the 3D or the 2D videoconferencing interface.

Figure 2 “Face-To-Face” (FTF) Condition

After each round, every participant filled in our questionnaire on Social Presence, which we used later for our data analysis.

Figure 3 Screenshot Condition “3D”

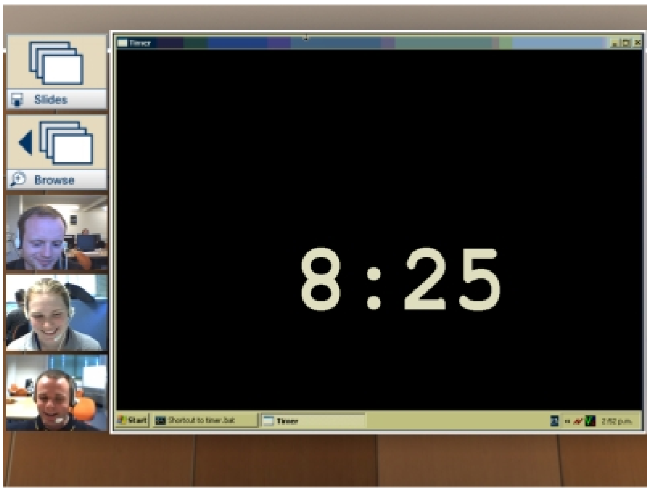


Figure 4 Screenshot Condition “2D”

3.1 Apparatus

For the mediated conditions (2D, 3D) three acoustically and visually separated rooms were prepared with identical standard desktop PC’s, monitors (TFT, 17”, 1280x1024), head-sets (stereo with mono microphone), and web cameras (USB, CIF resolution) (see figure 5).

All three PC’s were connected using a standard 1000 MBit/sec. network, although the available bandwidth was not necessary for the application (200 KBit/sec would have been sufficient).

Both mediated conditions (2D and 3D) consisted of variants of the same videoconferencing software “cAR/PE!” as described by Regenbrecht et al. [15]. This software represents all participants as video planes in a three-dimensional virtual environment as shown in figure 3. Users can freely navigate in the virtual room while their movements are directly mapped onto the position of their video planes in the environment. Participants can thus automatically convey spatial relationships between each other: They can be virtually close or far to each other, can face each other or the projection wall of the room, or can “sit” around a virtual table In addition the “cAR/PE!” software supports 3D-sound which further underlines the spatial character of the 3D environment.

Figure 5 Videoconferencing workplace

These features were fully available in the condition “3D”. At the beginning, all participants were placed around a virtual table. Afterwards they could freely navigate within the environment using the computer mouse (“head” rotation left/right and up/down, and movement forward/backward and left/right). If a participant got “lost” in the virtual room he or she could get back to the initial position at the virtual table by clicking on a home-button shown in an interactive menu on the bottom of the screen. Integrated into the virtual environment was a virtual screen which displayed a counter with the remaining time for the current round.

In condition “2D” the subject’s view into the environment was locked at a fixed position facing the timer screen (see figure 4). Video streams of the participants were shown on top of each other beside the timer screen, comparable to other conventional 2D desktop videoconferencing layouts.

In both conditions the same video and audio codecs were used (CIF resolution, 16 Bit stereo audio). The given video and audio quality was therefore constant in both mediated conditions and can thus be excluded from having any unwanted side effect. Video and audio are synchronized and the latency (loop) was about 300 msec. The size of the video for each participant in the environment in the “2D” condition was 6 cm x 4 cm measured on the monitor screen while in the “3D” condition this size varied according to the movements of the participants (own and others). In a view most participants intent to choose after a while (e.g. like in figure 3) comparable video sizes (measured on the monitor screen) as in “2D” were displayed (between 5 and 7 cm in width).

3.2 Participants

Forty-two subjects (36 male and 6 female) participated in the experiment. In 14 sessions each of three participants took part in three trials which gives a total of 126 trials. The age of the participants ranged from 19 to 63 years (median age 33 years).

Out of the 42 participants, three subjects reported to use videoconferencing tools regularly to communicate with their friends, further two subjects commonly used videoconferencing in their business context.

The participants were recruited by personal invitation mainly out of Information Science staff members. The assignment of participants to groups and time slots was based on self selection.

As an incentive a large chocolate bar was given to each participant at the end of each session.

3.3 Task

We chose the collaborative task “the Desert Survival Game” [11]. The main challenge of this task is to assign priorities to a given list of items such as a parachute, a cosmetic mirror or a compass, based on how useful the items would be for helping the group survive in a certain, given extreme situation (crash landing of an airplane in the desert). This task requires the people to work together as a team to resolve and interpret many uncertainties and to trade off all alternatives. According to the media richness theory [7], this sort of task requires a rich medium such as given in audio-video telecommunication and is thus appropriate for our experiment.

Another reason for choosing this task was the fact that the same task has been used in the pilot study of the Networked Minds measure of Social Presence [2] which we want to partly replicate and extend in our study. Choosing the same task guarantees a better comparability of results and ensures scientific stringency. In order to keep the task interesting and involving for the participants, we calculated the difference of the group’s ranking with an “expert solution” after each of the three conditions to give some feedback on how well their team was doing. Based on this interim score, we then encouraged them to further improve their result in the succeeding round by reconsidering the previous ranking.

The detailed game description was found at RogerKnapp [16] and was adapted in the following ways: (1) the number of items on the ranking list was decreased to 10 (from 15 in the original task), (2) values in miles and Fahrenheit were converted to km and centigrade.

3.4 The Questionnaires

Our *Social Presence Questionnaire* was applied after each of the three trials. It consisted of a combination of both measurement instruments as described in the chapters 2.2 and 2.3.

In the first part (38 items), we used all items of the Network Minds measure of Social Presence as described in 2.3 in randomized order.

In the second part, we put 9 bipolar pairs of the semantic differential technique as described in 2.2. Similar to the approach taken by Nowak and Biocca [13] we selected items directly out of Short's Social Presence measurement instrument. The bi-polar pairs chosen are:

Impersonal	-	Personal
Cold	-	Warm
Ugly	-	Beautiful
Small	-	Large
Insensitive	-	Sensitive
Colourless	-	Colourful
Unsociable	-	Sociable
Closed	-	Open
Passive	-	Active

In addition, we applied a *General Demographics Questionnaire* once to collect some details about the participants. This questionnaire assessed gender, age, simulator experience, previous use of telecommunication tools for business and private purposes, proficiency of English, if participants had played the task (Desert Survival Game) before and if they had prior experience with the cAR\PE! system which we used.

In addition, we applied a *General Demographics Questionnaire* once to collect some details about the participants. This questionnaire assessed gender, age, simulator experience, previous use of telecommunication tools for business and private purposes, proficiency of English, and if participants had played the task (Desert Survival Game) before and if they had prior experience with the cAR\PE! system which we used.

3.5. Procedure

The experiments were conducted during the first weeks of May 2005 at Otago University in New Zealand. For every one-hour session a group of three subjects were used. Upon arrival the participants could choose one of three seats at a table (marked as person 1, 2, 3). The participants were asked to read the *Participant Information*, explaining (1) the goal of the experiment (investigating differences in previous experiences with videoconferencing systems), (2) the general procedure, (3) the anonymity of the experiment, and (4) a participant consent text, which was to be signed by the subjects. Additionally the document contained the *General Demographics Questionnaire*.

After completion, a second sheet was handed out for reading: the *Participant Instruction*, which describes the Desert Survival Game.

Each participant had to take part in three rounds, one for each condition (FTF, 2D, 3D). The order of conditions was randomized beforehand (Latin Square). The task in each condition was the same (ranking of item list) and the group was told the interim result after each condition. One participant in each condition had the role of the "scribe", who had the list of items and who had to compile the group ranking list to be presented after each round.

In the 3D condition, the participants could navigate within the videoconferencing environment using a simple mouse interface. Participants were given an introduction of how to use this mouse interface and had then approximately 2 minutes to make themselves familiar with it. They were invited to think of the interface "as if they were together in a real room" and were encouraged to make use of the spatial cues the interface provides. A sheet explaining the mouse interface with pictures was put at the workplace as a further reference.

The 2D condition did not require any instruction. In both mediated conditions (2D, 3D), the subjects wore audio head-sets (see figure 5) which were explained and adjusted for best comfort.

After each condition, the subjects came back together and filled in the *Social Presence Questionnaire* on paper. The interim score of their ranking was announced and they continued to the next condition trying to further improve their result.

The experimenters played a passive role. They were only instructed to assist the participants in case of unforeseen circumstances or to help with the equipment. In addition, the experimenters made notes of their observations.

The last two sessions (6 trials) have been recorded on video tape for later use after agreement of the participants.

3.6. Hypotheses

We conducted three pilot test sessions (9 trials) with slightly altered setups and tasks before the actual experiment. Based on the first user reactions and the first measurement results, our general assumption was that a three-dimensional interface can support a higher sense of Social Presence than the two dimensional one, but supports less Social Presence than the Face-to-Face situation.

Regarding the first questionnaire results, the data we collected of the Networked Minds measure of Social Presence showed that not all items discriminated between the two mediated conditions but more clearly between the face to face and the mediated conditions. This lead to our experiment hypothesis number 1.

Hypotheses 1: Every factor of Social Presence, measured with the Networked Minds measure of Social Presence, is higher in the Face-To-Face condition than in both mediated conditions and at least several factor scores of Social Presence are higher with the three-dimensional interface than with the two dimensional one.

The semantic differential technique seemed to result in more uniform responses that tended to clearly discriminate between every condition. From this finding we derived hypothesis 2.

Hypothesis 2: Social Presence, measured with the semantic differential technique, is higher in the Face-To-Face condition than in both mediated conditions and Social Presence also is higher with the three-dimensional interface than with the two-dimensional one.

4. Results

The results presented in this chapter have been analyzed using SPSS version 11.

4.1 Reliability analysis

As all measured factors are multiple-item additive scales, a reliability analysis of the items in all factors was performed first. For this, Cronbach's Alpha was calculated for each variable (see tables 2 and 3).

Factor	Nr of Items	Alpha
Social Presence	9	0.93

Table 2: Test for internal consistency for the semantic differential measure of Social Presence

Factor	Nr of Items	Alpha
Isolation	2	0.54
Mutual Awareness	6	0.83
Mutual Attention	8	0.76
Empathy	6	0.70
Mutual Understanding	6	0.88
Behavioral Interaction	6	0.84
Mutual Assistance	4	0.74
Dependent Action	2	0.32

Table 3: Test for internal consistency for the Networked Minds measure of Social Presence

The alpha score for the factor Social Presence using the bi-polar pairs is very good, suggesting that the items measure a single uni-dimensional construct sufficiently well. Short identified this construct as Social Presence in his studies using the same items. The high alpha value also shows that an occasional appearing second orthogonal factor, which Short referred to as "aesthetic appeal", doesn't form in our case. Instead this factor seems to fuse with the dimension Social Presence so that all 9 items indeed describe the same dimension.

In the Networked Minds measure of Social Presence, the factors "Isolation" and "Dependent Action" reached an insufficient Alpha score and are therefore excluded from further analysis. This result of reliability is in line with the results of the pilot study presented by Biocca et al. [4].

No strong correlations were found between the different factor scores of both instruments. The strongest inter-correlation between scales of the two different measurement instruments appeared between the factor mutual understanding and the pair "cold-warm" in the 3D condition with a Pearson correlation factor of 0.58, p<0.001 and between mutual understanding and the pair "insensitive-sensitive" with a correlation factor of 0.63, p<0.001 in the 2D condition.

4.2. Comparing Means

The results of both measures are presented separately in the following two sections.

4.2.1 Networked Minds measure

The average score and standard error were calculated for every factor in the Networked Minds Measure of Social Presence and are displayed in figure 6. For a more detailed information about all sub-scores, please refer to Appendix A and B.

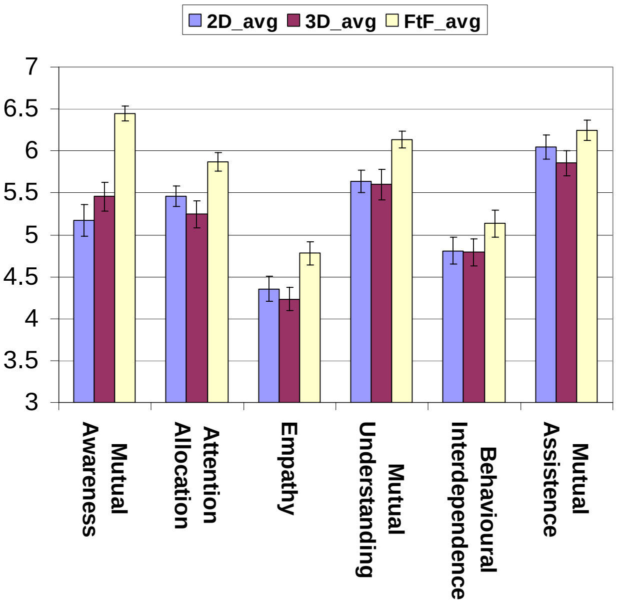


Figure 6: Mean differences and standard errors in the factors of Social Presence, as measured by the Networked Minds Measure of Social Presence

In every factor, the score of the Face-To-Face condition reached the highest value. Furthermore, every factor was analyzed in an analysis of variance with media (FTF vs. 3D vs. 2D) as a within-subject factor. The main effect of media was significant for all factors. Post-Hoc comparisons were performed using the Bonferroni adjustment for multiple comparisons. The scores of the face-to-face condition were significantly higher than the scores of the 3D condition in all factor scales. The scores for of the Face-To-Face condition was significantly higher than the 2D condition in the factors Mutual Awareness (p<0.01), Mutual Understanding (p<0.01), Attention Allocation (p<0.05), and Empathy (p<0.01). This result is slightly different from the findings of the Biocca's pilot test, which found significant differences between Face-To-Face and a 2D videoconferencing interface only in Mutual Awareness, Mutual Attention, and Mutual Assistance. No significant differences (p<0.05) could be found in any of the factors between the condition 3D and 2D.

H1 could therefore only be partly supported as not all factors were significantly higher in the face to face condition than in the 2D condition and there were no factors that showed any significant difference between the conditions 2D and 3D.

4.2.2 Semantic differential measure

The mean value and standard error of the dimension Social Presence as measured with the semantic differential technique, is displayed in figure 7.

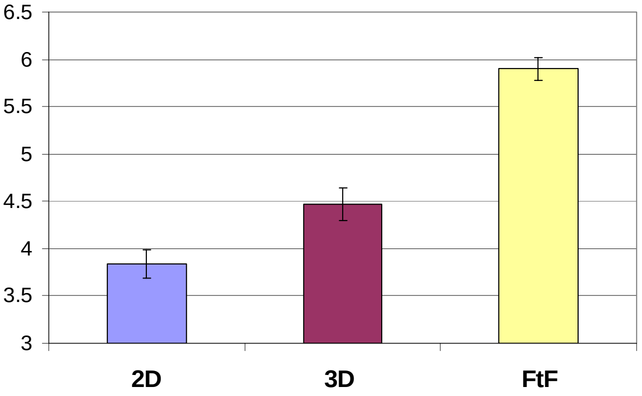


Figure 7: Mean Difference and Standard Error in Social Presence, measured with the semantic differential measure of Social Presence

The mean values of every condition were compared in an analysis of variance with media (FTF vs. 3D vs. 2D) as a within-subject factor. The main effect of media was significant for Social Presence, $F(2,82)=64.78$, $p<0.01$. After that, pair wise post-hoc comparisons were performed using the Bonferroni adjustment for multiple comparisons. The Social Presence mean score in the Face-To-Face condition ($M=5.90$, $SD=0.77$) was significantly higher than it was in the 3D condition ($M=4.47$, $SD=1.11$, $p<0.01$) and in the 2D condition ($M=3.84$, $SD=1.00$, $p<0.01$). Furthermore, the measured Social Presence in the 3D condition was significantly higher than the 2D condition ($p<0.05$).

The hypothesis H2 was thus fully supported. Social Presence, measured with the semantic differential technique is indeed higher in the Face-To-Face condition than in both mediated conditions and the three-dimensional interface can support a higher Social Presence than the two dimensional one.

5. Further Findings and Limitations

The control variables order, age, simulator experience, gender, proficiency in English, and experience with telecommunication technology were tested for further effects on all factors. No significant effects ($p<0.05$) were detected. (highest within-subject effects for Social Presence (measured with the semantic differential technique) , and gender $F(2,50)=1.99$, $p=0.15$, and Social Presence and age $F(6,50)=1.66$, $p=0.15$, both not significant.) This result shows that the semantic differential technique is quite robust against variable disturbances.

The experimenters wrote notes during the sessions about their observations. The intention for this procedure was to explore clues for further experiments in the field and to get some informal hints about the usability of the concepts tested. Although these observations lack empirical evidence, they are useful for explaining reasons behind the measures. Some of these anecdotal situations are:

While head-movement in reality is very fast and does not need any interface (except for the already learned one), its substitute in the 3D conferencing interface needs to be improved. Obviously the interface is not fast enough to meet the expectations of the participants. Turning the (virtual) head with a computer mouse is not fast and robust enough or deserves more training. For instance asking participants who frequently play computer games with their PC they would prefer a keyboard or combined keyboard/mouse interface because it is more “natural” for them.

There were clear indicators that users understood the spatial character of the interface in the 3D condition. For example, users frequently turned their avatar away from the other avatars towards the projection screen to see the timer and then back again, imitating a glimpse to the clock in a real room. Users also clearly liked and exploited the 3D sound for example, by adjusting the own avatar’s view direction towards an avatar who was out of the view while talking to them.

Very often participants tried to navigate to a “comfortable view” position (two others in view). The pre-set field of view of the software system used did not allow for three participants to see each other in that kind of view, so at least one participant had only one person in view and had to do a lot more virtual head rotations when wanting to see the communication partners. This could be improved in future versions of the system by either changing the virtual field of view in the environment (with a trade-off in correct perspective view) or by altering the navigation interface.

Occasionally, the sheet with the interim rankings was held in front of the camera by the “scribe” to show it to the other participants. Obviously there is some need to present something to others even in this mainly verbal task. Therefore, the approaches taken by AliceStreet [1], SmartMeeting [19] and Regenbrecht et al. [15] to include virtual presentation mechanisms into their environments seem to be logical. From an interface design point of view and considering these observations, a more “natural” interaction metaphor should be provided for the presentation of real world objects (like documents).

The “scribe” was almost always looking down to his/her sheet and therefore was not seeking face contact with the others. The communication took place on an almost audio-only channel. Surprisingly there was no effect on the results. This might lead to the assumption that the mediated environment was cognitively and emotionally "understood" within the first minutes of contact and later on taken for granted. So, the "scribe" was aware of the environment (and the perception of the others of it and of him/herself) even when not using the medium continuously. The three-dimensional audio capabilities of the system (to hear the other participants from their spatial position within the environment) could also have been contributing to this behavior and rating.

It was also observed that the display of the video stream of oneself in the 2D condition was valued as advantageous. It apparently gave some faith in using the system to know how others were seeing one. The same feature was present in the 3D condition (displayed video streams of all participants on the virtual room wall opposite to the presentation screen and table) but almost nobody made use of it. Perhaps it was simply too "laborious" to navigate to this place in the virtual environment.

We would also like to mention that evaluative studies of new media such as a 3D Videoconferencing system could also be distorted by a certain "Wow-Effect" by first time users. As Ijsselsteijn [10] points out in a review of the introductions of age-old media technologies, people's first responses to new and more realistic media such as the first photograph, first movie, or the first virtual environment have always been characterised as being very exciting, emotional, and intriguing. However, the reason for this is more grounded in the fact that previous expectations with and experiences of users were exceeded, rather than the sensory information that this medium provides could be improved. In our study we therefore have to be aware that this effect might also have an impact when we ask participants for their emotional attitudes of a new 3D-interface versus a common 2D one.

6. Discussion

With the presented experiment we have successfully replicated the Pilot Study of the Networked Minds measure of Social Presence in a three person setup. The results of our Networked Minds measure confirm, that the instrument is capable of discriminating the experienced Social Presence between unmediated and audio-video mediated communication. However, the instrument was not sensitive to comparisons within the two video conferencing interfaces. Also, we found the factors of *isolation* and *dependent action* failed the criteria for internal consistency. This suggests that the items in these factors should be reconsidered and modified in future studies. At this point it would also be interesting to run a factor analysis involving both our data and the Networked Minds Lab pilot study data sets to refine the current factor structure.

Using the semantic differential technique we succeeded to find a difference not only between the Face-To-Face versus the two audio-video mediated settings, but also between the two audio-video interfaces themselves in the Social Presence scale. This result confirms studies by Christie [6] where the hypothesis that the Social Presence dimension would discriminate even between variations of the same telecommunications medium was supported. The result of our semantic differential measure implies that the 3D-videoconferencing is rated to be more capable of supporting a high sense of Social Presence than the 2D version.

From an interface designer's point of view, this is a very encouraging result, suggesting that the semantic differential technique is sensitive enough to evaluate different interface features of telecommunication systems with respect to Social Presence as Short et al. defined it. For example in our presented experiment, the concept of 3D videoconferencing was encouraged as it seems that it has some positive effects on the user's attitude towards the medium. The instrument seems to be valid, elegant and robust, as it can be universally applied for different media and different tasks.

From a presence researcher's point of view, however, our result shows that in its current version, the Networked Minds questionnaire alone is not able to inform about how the experience of Social Presence is affected by telecommunication interfaces. Future evaluative studies should therefore try to add objective or physiological measurements as well.

A more robust and uniform theory and measure of the experience of Social Presence could advance many fields of telecommunications research including the exploration of design goals, properties and effects of telecommunication systems [4]. We believe that the Networked Minds theory of Social Presence is a good first step towards such a robust theory, but further efforts are required.

Having a look at the overall ratings in all factors apparently the Face-To-Face situation is still the gold standard as expected. The differences measured are clearly higher between the FTF and the mediated conditions compared to the differences between the mediated conditions (if any).

The increase in Social Presence, which we measured with the semantic differential technique, in the order 2D-3D-FTF, indicates that the more the system is similar to the FTF situation the higher the Social Presence. This leads to the assumption that an interface design towards three-dimensionality is a founded one.

We used 3-participant-groups, in contrast to many studies in social presence that involved only two interlocutors [2][9]. The result of our three person setup might give an idea, how the situation might scale to a larger number of people. Clearly, the more interlocutors are participating in a videoconference, the bigger is the need for them to stay aware of both, the situation and the others. Our result could be a first indicator that providing spatial

Video Conferencing systems which support a higher sense of Social Presence could be especially useful in multi-person scenarios.

It can reasonably be expected that a truly three-dimensional task would even further increase ones sense of Social Presence. E.g. the task described by Regenbrecht et al. [15], where the design of different 3D car models had to be evaluated, would probably benefit from the three-dimensionality of the environment itself. Further investigations are needed here to prove this assumption.

While the focus of this study was set on perceptual issues, namely Social Presence, the question remains if task performance can be increased by using three-dimensional user interfaces rather than two-dimensional ones. Empirical evidence here would clearly strengthen the argument for this new kind of environments. We assume that task performance will positively correlate with communication quality in computer-mediated communication. Further research is needed here.

7. Conclusions

We have shown that Social Presence increase from two- and three-dimensional mediated to real face-to-face communication. While the natural face-to-face meeting situation is still by far the benchmark for all mediated systems the introduction of three-dimensionality in computer-mediated communication is a well-founded step.

While the instrument given by Short et al. [18] is a reliable and elegant one to measure the main dimensions of this study, further work is needed towards a robust theory and measure regarding all other dimensions. This was stated by Biocca at al. [4] and could be reinforced here.

Based on our results, further findings, and observations in this study we believe that research and development in three-dimensional video-conferencing can be seminal.

We hope that researchers and practitioners in the field will benefit from our findings and that we can contribute with this study and in the future to more effective, efficient, and enjoyable computer-mediated communication interfaces.

Acknowledgements

We would like to thank Gordon Wetzstein, Graham Copson, and all the participants in our experiments for their great support. We would like to acknowledge the very fruitful discussions with Thomas Schubert and Dean Owen during earlier versions of this article. In addition, we would like to thank DaimlerChrysler Research and Technology for supporting our work and the three anonymous reviewers. The German Academic Exchange Service, DAAD, financially supports the first author.

References

[1] AliceStreet (2005). Online product description. http://www.alicestreet.com/. Last accessed 31 May 2005.

[2] Biocca, F., Harms, C., & Gregg, J. (2001). The Networked Minds Measure of Social Presence: Pilot Test of the Factor Structure and Concurrent Validity. Paper presented at Presence 2001, 4th international workshop, Philadelphia.

[3] Biocca, F., Burgoon, J., Harms, C., & Stoner, M. (2001). Critaria and scope conditions for a theory and measure of social presence. Paper presented at the Presence 2001, 4th international workshop, Philadelphia.

[4] Biocca, F. Harms, C. and Burgoon, J.K. (2003). Towards a More Robust Theory and Measure of Social Presence: Review and Suggested Criteria. Presence, 12(5), 456-480.

[5] Burgoon, J., Buller, D., & Woodall, W. (1996). Nonverbal Communication: The Unspoken Dialogue (2nd ed.): McGraw-Hill.

[6] Christie, B. (1973). Appendix M. In P.C. Goldmark et al. The 1972/73 New Rural Society Project. Research report available from Fairfield University, Connecticut.

[7] Daft, R. L., & Lengel, R. H. (1984). Information richness: a new approach to managerial behavior and organizational design. In L. L. Cummings & B. M. Staw (Eds.), Research in organizational behavior 6 (pp. 191-233). Homewood, IL: JAI Press.

[8] de Greef, P., & IJsselsteijn, W. (2000). Social Presence in the PhotoShare Tele-Application. Paper presented at Presence 2000 - 3rd International Workshop on Presence (27-28 March 2000), Delft, The Netherlands.

[9] Garau, M. (2003). The Impact of Avatar Fidelity on Social Interaction in Virtual Environments. PhD thesis, University College London. http://www.cs.ucl.ac.uk/staff/m.garau/papers/garau_phd.pdf

[10] IJsselsteijn, W. (2003). Presence in the Past: what can we learn from Media History? In G. Riva, F. Davide & W. IJsselsteijn (Eds.), Being There - Concepts, effects and measurements of user presence in synthetic environments (pp. 4-14). Amsterdam: Ios Press.

[11] Lafferty, J. C., & Eady, P. M. (1974). The desert survival problem. Plymouth, Michigan: Experimental Learning Methods.

[12] Lombard, M., & Ditton, T. (1997). At the heart of it all: The concept of presence. Journal of Computer Communication, 3(2).

[13] Nowak, K.L., & Biocca, F. (2003). The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and Social Presence in virtual environments. Presence 12(5), 2-35.

[14] Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). The measurement of meaning (Vol. 1). Urbana: University of Illinois Press.

[15] Regenbrecht, H., Lum, T., Kohler, P., Ott, C., Wagner, M. T., Wilke, W., et al. (2004). Using Augmented Virtuality for Remote Collaboration. Presence, 13(3), 338-354.

[16] RogerKnapp (2005). Scouting Games. Online Resource. http://www.rogerknapp.com/download/games.htm. Last accessed: 31 May 2005.

[17] Schroeder, R. (2002). Social interaction in virtual environments: Key issues, common themes, and a framework for research. In R. Schroeder (Ed.), The social life of avatars: Presence and interaction in shared virtual environments. London: Springer.

[18] Short, J., Williams, E., & Christie, B. (1976). The social psychology of telecommunications. London: John Wiley & Sons.

[19] SmartMeeting (2005). Online product description. http://www.smartmeeting.com/. Last accessed 31 May 2005.

Appendix A: Descriptive statistics table for the Social Presence Items

Social Presence (sem.dif.)	FTF			3D			2D		
	mean	std. error	std. dev.	Mean	std. error	std. dev.	mean	std. Error	std. dev.
cold -warm	6.29	0.14	0.89	4.43	0.24	1.53	3.93	0.22	1.40
impersonal-personal	6.31	0.19	1.20	4.29	0.22	1.44	3.95	0.21	1.34
unsociable-sociable	6.64	0.10	0.62	4.71	0.21	1.33	4.33	0.20	1.30
closed-open	5.98	0.20	1.28	4.24	0.23	1.48	3.86	0.24	1.57
ugly-beautiful	4.88	0.15	1.00	4.31	0.22	1.41	3.67	0.18	1.16
colourless-colourful	5.33	0.23	1.48	4.64	0.21	1.34	3.71	0.24	1.55
passive-active	6.24	0.16	1.01	4.90	0.23	1.49	4.26	0.24	1.58
insensitive-sensitive	6.10	0.15	0.98	4.43	0.21	1.36	3.71	0.17	1.09
small-large	5.33	0.23	1.51	4.29	0.22	1.44	3.12	0.20	1.31

Appendix B: Descriptive statistics table for all sub factors in the Networked Minds Questionnaire

Networked Minds Measure of Social Presence factors	FTF			3D			2D		
	mean	std. error	std. dev.	mean	std. error	std. dev.	mean	std. Error	std. dev.
Mutual Awareness	6.45	0.09	0.61	5.46	0.17	1.11	5.17	0.19	1.22
Mutual Understanding	6.13	0.10	0.68	5.60	0.18	1.17	5.64	0.13	0.85
Mutual Assistance	6.24	0.11	0.76	5.86	0.16	1.05	6.05	0.12	0.77
Empathy	4.79	0.14	0.89	4.23	0.14	0.93	4.35	0.15	0.98
Attention Allocation	5.87	0.12	0.75	5.24	0.15	0.99	5.46	0.14	0.89
Beh. Interdependence	5.13	0.16	1.00	4.79	0.16	1.06	4.81	0.16	1.05

Session 7

22nd September, 2005

Presence Theory and Experimentation

17.30-18.00 *Agency and Presence: a Common Dependence on Subjectivity?*
Gerardo Herrera[1], Rita Jordan[2] and Lucía Vera[1]
[1]Autism & Learning Difficulties Group, Robotics Institute, University of Valencia, Spain
[2] School of Education, University of Birmingham, United Kingdom

18.00-18.30 *When Presence and Emotion are related and when they are not*
Jonathan Freeman[1], Jane Lessiter[1], Katherine Pugh[1] and Ed Keogh[2]
[1] Department of Psychology (i2 media research), Goldsmiths College, University of London
[2] Department of Psychology, University of Bath

18.30-18.45 *Schemata, Narrative and Presence*
Dan Pinchbeck and Brett Stevens
Department of Creative Technologies, University of Portsmouth, UK

18.45-19.00 *The role of content preference on thematic priming in virtual presence*
Ilda Ladeira[2], David Nunez[1] and Edwin Blake[2]
[1] Department of Psychology, University of Cape Town, South Africa
[2] Department of Computer Science, University of Cape Town, South Africa

Agency and Presence: a Common Dependence on Subjectivity?

Gerardo Herrera[1], Rita Jordan[2] and Lucía Vera[1]
[1]Autism and Learning Difficulties Group. Robotics Institute. University of Valencia,
[2]Autism Team. School of Education. University of Birmingham
{Gerardo.Herrera@uv.es, r.r.jordan@bham.ac.uk, Lucia.Vera@uv.es}

Abstract

The analysis of agency, a very close concept to presence, is of great help for acquiring insights into how the sense of presence is acquired in the developing child and also about the experience of presence itself. Empirical evidence coming from Cognitive Developmental Research together with the positive outcome of people with autism (who are not generally able to act "as if") when participating in Virtual Environments suggests that presence is more about 'experiencing agency' than 'Pretending to be there' or than constructing and reconstructing mental models in real time. It is considered that these phenomena shed some light on the current issues of Presence Research and open up new fascinating philosophical and psychological ones, both in relation to Presence and Autism.

Keywords --- **agency, presence development, autism, interaction, affordances, subjectivity**.

1. Introduction

1.1. Agency, action perception, imitation and autism

From the point of view of the user of Virtual Reality (VR) systems, agency has been referred to by Murray [1] (p.126) as "the satisfying power to take meaningful action and see the results of our decisions and choices". The individual-environment relationship and the potential for action of the environments, very close concepts to agency, have been the focus of attention of other researchers in this field such as Spagnoli et al. [2], and also Zahorik et al. [3].

The field of cognitive developmental and autism research provides other more complete attempts to define agency as in that of Russell [4], who defines agency as the exercise of a capacity for first person experience that has four integral features: the first two describe types of information-processing and control that the agent must achieve, and the other two describe the kind of self-knowledge that is available to agents and to agents alone. These features (that will be analysed in the third section of this paper) are:

A. Locating the cause of altered inputs in one's body rather than in the world –"action-monitoring".
B. The perceptual sequences brought about by acting are reversible; but those experienced in perceiving environmental change are irreversible.
C. Our actions are known non-observationally whereas the world is known by observation.
D. Agents have a privileged knowledge of their own "tryings" which they lack when observing the "tryings" of others (although the existence of 'mirror neurons' (Rizzolatti)[5] provides a possible mechanism through which the linkage between one's own and others' actions might be made apparent).

Brewer [6] used the term 'experiencing self' (akin to 'Presence') to characterise our typical moment to moment awareness of ourselves in the process of perception of the world, which is a more comprehensive conceptualisation than that of Steuer [7] (p.73) who characterised presence just as the *sense of being in a place*. The understanding of presence that emerges from Brewer's conceptualisation fits better with Russell's definition of agency and puts both concepts very close to each other. Therefore, we will use presence in that way throughout this paper.

We understand that what we will call 'Tangible Presence' can occur in natural/physical Realities (where natural presence takes place), in Technologically Mediated Realities (such as Augmented/Mixed Reality or Real Environments equipped with Ambient Intelligence) and in Virtual Realities. Tangible Presence is possible both remotely (as in Telecommunications and Teleoperations, i.e. *Telepresence*) and locally (as in Augmented Realities), and both alone or in social contexts (Co-Presence).

1.2. Hypothesis about the role of Agency in Presence

Analysing how the sense of presence is acquired through typical development and in autism offers some insights into the concept of presence concept and leads towards understanding of the role of agency both in the development and in the experiencing of presence in any environment and by any individual.

Several authors such as Biocca & Delaney [8]; Kalawsky [9]; Sheridan [10,11]; Welch, Blackmon, Liu, Mellers, & Stark [12] have reached consensus in seen presence as a multidimensional concept. We argue further that there is a multidimensional continuum that goes from Absence to Presence of this sense of being engaged in the perceived environment and that agency is a regulating variable that usually correlates with the level and type of presence obtained.

We also argue that presence is a subjective measure and, as a consequence, it adopts different forms for each person in different moments or situations and with different technologies.

Finally, we propose that the agency based model (which is more connected with experience) is a better model than *constructivism* (proposed by Nunez [13) for describing and explaining the experience of participating in Virtual and Real Environments, since it has greater potential to be an empirically (rather than metaphorically) based model.

2. Agency and the development of a natural sense of presence

Agency and presence cannot be understood fully by paying attention solely to the environment or to the individual; it is necessary to consider the relationship between them. As individuals with autism find it difficult to cope with the environment (perceptually, culturally and socially) we will also analyse this human condition within this section.

For obtaining more insight into that feeling of presence that we have in natural settings, it is useful to separate the 'sense of presence' from the child's '*development* of a sense of presence' for which exercising agency is fundamental, and this will help us to know more about presence in Technology-Mediated and Virtual Realities.

2.1. The case of typical developing children

There are different mechanisms in which the development of a sense of presence is supported. Without the intention of being exhaustive, we outline those that we find more relevant for the scope of this paper.

The role of sensory perception in the development of a sense of presence: The perceptions we receive from our senses have a very important role in the configuration of the sense of being there as they keep us 'connected' with reality at every moment. Relevant here is the concept of *affordances* as noted by Zahorik et al. [3]. *Affordances*, as Gibson suggested [14, define the opportunities for perception and action offered by the environment in the context of the individual's capacities: they are things that one perceives directly (without the need of a mental representation process). This is characterised by Valenti [15] (p.90) as '*It is more of something we live in rather than we think*'. Loveland [16] claims that discovering and acting upon available *affordances* is an essential process in development. In general, people suffering from learning difficulties (with or without autism) will also have an impoverished exercise of agency as their knowledge of the *affordances* (physical, cultural or social) of the environment will be very limited and so they have a limited range of actions to execute through agency. Gibson's [14] passive view of perception, as derived from the *affordances* of the world around, has been challenged by the view of those like Russell [4] who claim that the infant must become aware of his/her own actions in order to truly 'perceive' (i.e. to make sense of sensory information).

The role of exercising agency in the development of a sense of presence: The development and continuous update of our mental world, then, is fully connected with our interaction with the environment (exercising our agency). Russell (op cit.) shows how sensory perception is bound up with action and that presence comes from agency, through the capacity thus afforded of distinguishing self from other perceptions. Mirror neurons, a subset of action-coding neurons identified in the premotor cortex in monkeys by Rizzolatti [5], show activity both in relation to specific actions performed by self and in matching actions performed by others (see the work of Williams [17] for further explanation and implications in humans). This work also contributes to our understanding of how the distinction of the self from others (and the notion of others) is fostered further through reciprocal imitation. However, in typical development at least, the child does not act autonomously on the world but does so initially as a 'social unit' with those who help tutor the sense of both personal and social agency. As Hobson [18] makes clear, the sense of self and of others and the capacity to make human sense of the world depend on the quality of social interactions through which such understandings develop. The claim is that it is through inter-subjectivity that we are enabled to take a subjective stance and thus have a sense of presence. As Halliday [19] suggests, the baby is taught how to mean, and so missed opportunities to engage in these acts of emotionally charged mutual agency (as happens in autism) lead to a failure in culturally appropriate perception.

The role of connecting the experience with feelings and emotions: The work of Damasio [20] on the biological representation of emotion in the brain, gives further insight into the development of 'the sense of what is' as Damasio phrases it. Damasio sees the subjective experience of the world as being the root of consciousness. He characterises emotions as being represented in the brain at three different levels, occurring in a timed sequence during a single emotional event. The initial stage (called 'emotions' by Damasio) is purely represented at sub-cortical levels of arousal and there is no subjective consciousness of this stage. The next stage (Feelings 1) is represented also subcortically but reflects reactions to the first stage and is mediated through the hormonal system; this stage is also unconscious. It is only at the third stage (Feelings 2) where emotion processing subcortical areas are linked, via the hypothalamus and amygdala loop, to the cortex that the individual becomes aware of his/her emotion and is able to name it and control its expression. Thus, conscious experience of the world and the sense of oneself as an agent within it, are linked to this emotional level of representation. The relationship to the development of first order presence is not clear, but it is clear that conscious awareness is a necessary condition of presence.

2.2 The case of autism

Autism is a pervasive developmental disorder of brain functioning (definition of American Psychiatry Association [21]). According to Wing [22], the main symptoms of autism are:

- Deficits in social reciprocal interaction
- Deficits in verbal and non verbal communication
- Limited range of activities and interests

2.2.1. Why consider Autism?

In developmental psychology, it is common to study people with developmental disorders to find out about the existence of very basic abilities of typical human development. Studying people who have differences in their interaction with the environment (perceptually, culturally and socially), as occurs with people with Autism Spectrum Disorders (ASD)[1], would help us to understand both our and their relation to the real or virtual world better. This is especially the case if take into account the work referred to above that links autism with deficits in the basic mechanisms that might lead to a sense of agency and hence, of presence. Individuals with autism (possibly all those with ASD, although it is clearer in those with classical autism defined by Kanner [23]) also appear to have problems with 'first-order' presence i.e. with a subjective experience of the world in which they act as agents (see Grandin [24]; Powell & Jordan [25]).

There are a number of studies aimed at educational intervention for people with autism where VR settings have been used as a medium for developing this sense of agency, as in that of Herrera et al [26] or Parsons [27]. As we will examine later, the data collected from the experience of people with autism facing VR situations is a very useful secondary result of those studies in the sense that they help us to better understand the experience of typical developing people and people with autism when participating in these settings.

Recent work on 'mirror neurons' from Nishitani et al. [28] and from Oberman et al. [29] suggests that this mechanism, which affords the final aspect of agency (feature D above), may be missing or at least dysfunctional in autism. Oberman et al. (ibid) and Williams et al. [17] make the case for such a fundamental deficit being the basis for the sequalae of social and interpersonal problems that characterise autism.

Sensory disability is also a source for obtaining some insight about presence. Oliver Sacks [30] describes the case of a man whose sight was restored after 45 years of blindness. After such a long period, his world had been built up through other senses (cited, Bogdashina [31]) and thus visual stimuli still did not play a major role in his understanding of the world even after his sight was restored. He was able to see but not to decipher what he was seeing [30]. This is only an extreme case of how the same environment (even with the same perceptual inputs) can produce different kinds of presence depending on the previous experience of each individual; it is an example of the subjective nature of presence. The difficulties suffered by people with hearing difficulties in hearing themselves, in order to modulate their speaking, also supports the perception/action cycle view of agency.

2.2.2 Autism and Presence

Nadel [32] makes the case for a connection between early imitation and a sense of agency, citing the psychological and neuroimaging experiments that have demonstrated that there are some common neural and cognitive mechanisms underlying perception of action, action generation, action simulation, action recognition and action imitation. She further points out that human beings as young as 2 months are selective in their imitation of biological agents as opposed to mechanical actions with similar perceptual qualities. She suggests, using Russell's characterisation of agency above, that this kind of imitation may be the basis for distinguishing self from other agency.

Rogers et al. [33] claim that children with autism have difficulties with spontaneous imitation of others but Nadel was able to enhance the capacity to imitate other children, in children with autism, by first giving them experience of synchronous imitation of their actions by a robot.

Jordan has described the problems of 'first-order' presence of people with autism as lacking an 'experiencing self' (see Jordan [34]; Jordan & Powell [35]; Powell & Jordan [25]), after Brewer's [6] use of this term to characterise our typical moment to moment awareness of ourselves in the process of perception of the world. This theoretical notion is allied to (although not identical with) Russell's [4] characterization of autism as a failure to develop 'social agency' and Hobson's [18] account of how early failures in social and emotional processing lead to later failures to differentiate the self (and other) from the experience. People with autism have themselves described their experience of the world as like watching a video (as in Grandin [24]). The result is a unique 'objective' view of life, which makes it easier (unlike any other group) to recall what they have witnessed than what they have experienced (Millward et al. [36]), poor spontaneous recall of personal episodes alongside phenomenal cued, rote and procedural memory and a world view that Baron-Cohen [37] suggests is at the extreme end of systemetising as opposed to empathising.

The reasons for the failure in first order presence in autism may be manifold. At one level it may relate to the way emotions are processed in the brain. We know that the areas of brain functioning linking emotional with cognitive processing are disturbed in autism and we know that there is often (always?) extreme delay in developing understanding of their own emotional states (and thus the emotional states of others). It may be that, without that cognitive emotional link, it is not only that it is hard to be consciously aware of emotions but also to be emotionally aware of cognition i.e. to develop a subjective 'presence' in the world.

We also know that people with autism have difficulty in becoming aware of their own intentions (perhaps through failures in 'efference copying' –see next section for an

---

[1] All along this document, in many instances, we talk about autism to refer briefly to Autism Spectrum Disorders

explanation of this concept: Frith & Frith [38]) and that this is one of the breakdowns in neural functioning that also occurs in schizophrenia. In the latter condition there has been a stage of normal development before the illness so that a loss of that awareness is interpreted, by the individual, as what we might call ‘false secondary mediation’. Thus, the person who is used to the feeling of intending his/her own actions reacts to the lack of that feeling by imputing technological or ‘other’ agency in their own actions (they feel ‘controlled’ by radio waves, extra terrestrials and so on). The person with a developmental disorder like autism, however, may never have experienced that sense of agency or intention so they do not react with delusions and paranoia to its loss, but instead are far less engaged in the world and far less aware of their own engagement when it does happen.

If mirror neurons are absent or dysfunctional in autism and Asperger syndrome, as growing evidence suggests (see above), then there is a further barrier to developing that sense of an ‘experiencing self’ (i.e. presence). The capacity to engage in reciprocal imitation not only leads to bonding but to the capacity to distinguish ones own actions from those of others and to learn the difference between being a passive viewer of life and someone who is emotionally and physically engaged.

3. Agency in Technology-Mediated and Virtual Realities

Here we will review the implication of agency in Technology-Mediated and Virtual Realities and, for this analysis, we will first adopt the limitations of the state-of-the-art technology and then we will adopt a focus that anticipates a future position where there are no such technological limitations. We will do this in order to outline a model that may last through time, not restricted to the current state-of-the-art.

We will also propose ways of obtaining an ‘augmented agency’ by augmenting the potential of each core feature of Russell’s model of agency. This will be done, respecting those features by manipulating them in the natural dimension (that goes from non agency to natural agency) but going beyond natural limits to reach augmented agency. Examples of other non-natural possible ways of manipulating agency will also be examined.

3.1. Types of information and control that the agent must achieve.

As indicated above, Russell [4] grouped those features of agency into two pairs: The first pair (A and B) describe kinds of information-processing and control that the agent must achieve.

Manipulating Feature A: Locating the cause of altered inputs in one’s body rather than in the world: “action-monitoring”.

To illustrate some of the limitations of some of the state-of-the-art technologies for obtaining a natural-like (“natural”) agency in relation to this feature, we find that when a VR helmet is used, there may be efficiency limitations in the head tracking system used –in some low frequency electromagnetic trackers– since the changes in our visual input are not naturally correlated to head movements (which is a cause of altered input that comes from one’s body). In this case, it would be difficult to solve the problem of self-ascription versus world-ascription of the changes in the visual input (a natural ability that is called ‘efference copying’ by Host and Mittelstaedt, cited Gallistel [39]). Given the fact that, according to Russell [4], visual experiences are, to some extent, a function of what we do, if there are interferences in this action monitoring, this will lead to a disturbed sense of agency.

Another problem is the imperfection of the available representations of the user’s body in VR. Some authors such as Tang, Biocca & Lim [40] have already suggested that the absence of representations of the user’s body in the VR environment may lessen the sense of spatial presence compared to the Augmented Reality environment. In fact, the variety of Augmented Reality that consists of seeing oneself (with a VR helmet with embedded subjective video-camera and real-time video processing and reproduction in the helmet displays) allows the participants to experience agency in a natural way, although it also includes artificial additions. A variety of VR that includes body representation is that of Fernandez and Gimeno [41] where an infrared motion tracking system and a Cave Automatic Virtual Environment (CAVE) have been used for obtaining the information about the user’s body and drawing it in a virtual mirror, thus increasing the body perception of the user.

Using Joysticks for moving around the virtual environment (VE) is not a way of promoting a natural sense of agency, although there are examples of adapted joysticks that are natural for certain applications (such as a steering wheel controller of a driving simulator). Thus, there are implications in the particular interfaces used, if a “natural” presence is the goal.

We can believe that, at some time in the future, these limitations will be completely solved in Virtual Environments and then we would be able to configure this Feature of agency in a way that produces a natural sense of agency and presence. We would also be able to exploit the potential of technology by manipulating this feature in a way that produces other kinds of artificial agency and presence. For example, we can modify this feature of agency just by inverting the positive and negative (x, y) values of the movements of our head and then produce an inversed efference copy that would lead to a different (and perhaps uncomfortable) artificial agency. An example in the direction of augmented agency would be augmenting the visual perception of our body, such as having transparency in the skin of our virtual body and being able to see our heart beating when the physiological measure of our heart rate exceeds a given value. This would be similar to the visual effect recreated in the film ‘Amelie Poulain’ of Jean Pierre Jeunet (min 38) [42], with the difference that in this film the director uses this effect to communicate something to the audience, whereas the combined use of virtual and electrophysiological technologies could be a way to use visual information to augment proprioception of the user him/herself.

The kind of agency that will be generated in this way will be artificial (i.e. not natural) in the same sense that the agency that is generated with the limitations in the efficiency of some head trackers of the current state-of-the-art technologies is artificial. Nevertheless, as the ‘Amelie’ example is situated beyond the natural limit of feature A, the perception of ourselves when interacting with the environment would be augmented and we could call this ‘augmented agency’.

Manipulating Feature B: The perceptual sequences brought about by acting are reversible; but those experienced in perceiving environmental changes are irreversible.

State-of-the-art VR environments (such as 3D games) perfectly incorporate this possibility, so it is not necessary to wait to have natural agency in relation to this feature in the future. Again, it is possible to manipulate this possibility to obtain an artificial experience of agency. As an example of this, we could say that in reality we can ‘undo’ our stream of visual input just by going back again with our eyes over the previously seen stimuli and that, thanks to Gaze Tracking Technologies, even currently it is possible to construct a gaze-contingent virtual environment that is voluntarily configured in a way that the user cannot pass his eyes back over and find what he/she had previously seen in his/her visual perceptual sequence again (for example, putting an apple where he/she has just seen an orange or even, at a more basic level, manipulating colours and shapes).

A way of promoting ‘augmented agency’ by manipulating this variable would be to increase our capacity to reverse our perceptual sequences going backwards further than the natural limits of our short term memory (reviewing our perceptual sequences of previously lived minutes, hours, days, months or years).

3.2. Self- knowledge that is available to agents and to agents alone.

The second and final pair of features (C and D) pointed out by Russell [4] describes the kind of self-knowledge that is available to agents and to agents alone:

Manipulating Feature C: Our actions are known non-observationally whereas the world is known by observation.

Shopenhauer [43] claimed that we know everything representationally except facts about our will. Russell [4] explains that the representations about our actions (e.g. I am doing X) are not gleaned from self-observation: they are known immediately, in the sense of “without inference”.

This feature is fully related to the degree of sensory immersion in VEs. If we want to obtain a “natural” sense of agency in a virtual environment, we should always act in the first-person. However, the state-of-the-art technologies do not allow us to obtain a Quality of Immersion comparable to the one we have naturally in the real world (although that is not the case with Augmented Reality). Quality of Immersion (Schubert et al. [44]) refers to immersion that includes sensory factors (Witmer & Singer, [45]), multimodal presentation and consistency of multimodal information (Held & Durlach, [46]), but (we suggest) not necessarily to the environmental richness or other non-sensory related features.

With the advance of technology, these difficulties will be solved at some time in the future and then it will be possible to obtain a natural first-person experience of interaction with the VR world. Once again, potential manipulation of this agency variable can be outlined. A possible example would emerge if we distort the way our actions modify the virtual world (i.e. producing inverted effects to those of the same action in the real world, such as making it necessary to grasp a VR object if we want to release it), then this feature would also be challenged by forcing us to know our actions through observation or inference and thus obtaining an agency that, at least at first, will be really different to the natural one.

An example of augmented agency would be just giving the user more potential for interaction than what he/she has in reality by allowing him/her to move VR objects with his/her eyes using a gaze tracking system. In this line also is the work of Duncan et al. [47] where they use electroencephalogram signals for what they call ‘thought-controlled music systems’.

Another example would be to let the user obtain knowledge of the world non-observationally by allowing him/her to adopt the subjective points of view of others (and swapping these with his/her own at his/her will).

Manipulating Feature D: Agents have a privileged knowledge of their own “tryings”, which they lack when observing the “tryings” of others.

O’Shaughnessy [48] defines strong knowledge of agents as knowledge whose falseness is impossible to imagine. Russell [4] explains that the agent’s knowledge of what he or she is trying to do in goal-directed action has a degree of first-person authority similar to the first-person authority of an experiencer of a sensation (such as pain) and claims that, through the exercise of agency, one gains the conception that agents have immediate and incorrigible knowledge of some aspects of their mental life.

In order to widen the field of view and interaction with the user’s body, some existing VR games (such as Tomb Raider [49]) include an avatar that is fully managed by the user in what it is called the *third-person mode* of controlling the game. Although it is a third-person perceptual point of view, the user does not establish differences between him/her and the avatar he/she controls, so the user adopts the same first-person conceptual point of view. Even with this, the sense of agency in relation to this feature would be artificial rather than natural because of these perceptual differences and because of the requirement for the user to identify him/herself with the avatar he/she is controlling in a third-person perceptual point of view.

For obtaining augmented agency and presence here, we may want to give the user access to the privileged knowledge of the "tryings" of others by constructing what we may call a 'shared subjectivity' in which a user can transfer him/herself to (and acquire some control over) the subjectivity of another. This could be as simple as remotely controlling someone else's computer mouse and keyboard, or as complex as controlling some movements and actions of another's virtual body.

In a virtual social framework where several agents participate in a 'shared subjectivity' basis it would be possible, for example, to have a face to face conversation with other agents also knowing that behind the eyes of that agent there can be several human people. The deeper into our cycle of perception/action technologies we go the greater the possibilities in this direction.

4. Discussion

4.1. Agency and it's correlation with Presence

Once agency and its potential have been analysed, in order to clarify the differences between presence and agency we start by going back to the role of agency in the development of a sense of presence.

As we stated in the second section, exercising agency is a necessary companion in the journey that enables us to take a subjective stance and thus have a sense of presence but, once the capacity of experiencing presence has been developed, do we still need agency to experience presence?

4.1.1. Affordances, Agency and Presence

One consideration is the state-of-the-art limitations to the realism of stimuli and perception (the cycle of ourselves perceiving the environment and the environment perceiving us). Another is the potentiality of the environment itself for interacting with us. This is a good complementary concept to presence in that it has the potential of putting together many of the other components of presence.

Defined by Gibson [14] and previously approached in presence research by Zahorik et al. [3] *affordances* relate to the action-supportive information of the content of a given environment. Experiencing agency also means being able to put all our repertoire for action into practice and, if we do not perceive this possibility, then our agency is impoverished.

If we have expectations about the contents of a given environment and the objects we find fit those expectations, then the affordances provided by those objects can be seen as enriching our sense of presence. If those objects are tools through which we gain access to the affordances of other (different) objects then, once we have them in our hands, they will act as an extension of ourselves increasing or modifying our potential Agency in that environment. We may say that in relation to the sense of being 'there', finding what we expect to find is something that brings us 'nearer there' and not finding it moves us further away 'from there'.

Zahorik et al. [3] claim (p. 87) that '*Presence is tied to action in the environment*' and further that '*Successfully supported action in the environment is a necessary and sufficient condition for presence*'.

Although we believe that Zahorik et al.'s assertions are generally right, without taking away importance from the role of agency in presence, we may say that even in reality the demands for action vary from one environment to another and depend also on the previous experiences and individual profile of the participant. For example, the employee of a repair shop would feel a high intensity demand for action in that environment but if he/she is not used to eating popcorn (or some similar activity) while watching a film, the demands for action he/she will receive in the cinema will be kept to a minimum. A clearer example occurs when someone with paralysis is in a non-accessible environment where he/she cannot do any single action. Both examples illustrate how, at least in some situations, there can be presence without potential for action.

Agency, as well as attention and other variables, helps very much to fill up our '*moment to moment awareness of ourselves in the process of perception of the world*' (i.e. presence) [6] but in certain situations it is not a necessary condition for this. As indicated in the second section, individual differences deviate more from the typical in those who cannot typically 'perceive' because they do not have the typical awareness of their actions [4] (those who have not acquired a typical sense of presence).

4.1.2. Tangible presence vs. Imaginary Presence

The imaginary experience of being in another place that we experience when we read a book or in daydreams has been claimed to be a form of presence that it is known as 'the book problem' (Schubert [50]). Although it can metaphorically be considered as presence, this kind of experience is certainly not about '*our moment to moment awareness of ourselves in the process of perception of the world*' (Brewer [6]) as this world would be imaginary rather than real or even artificial. As the main component in this kind of "presence" is provided by imagination, we prefer to label this experience as being 'imaginary presence'. In the borderline of tangible and imaginary presence we would find those films or those non agency-able contents where perception still plays an important role and conspires with imagination to obtain that feeling of being there.

When we are using state-of-the-art VEs that include interferences from the real world, we can also fill those lapses of agency or presence with our imagination. This would be similar to when there is a power cut and the lights are turned off, and we have to move in the darkness to find a candle or a lighter. We will go back to this 'power cut' problem in the next section.

The amount of working memory dedicated to the task would also contribute to a higher involvement and then to a higher sense of presence (Nunez [51]), but this would only be necessary in those situations in which the state-of-the-art technologies still have failed to obtain a “natural” presence (if a natural-like presence is pursued).

Even in relation to social presence (feeling or being there with other people) there can be a high component of ‘imaginary presence’. A very well experimented situation in developmental psychology is the ‘Sally Ann’ Test from Wimmer and Perner [52], where participants’ abilities for attributing false beliefs are assessed. In this test the participant sees how a doll (Sally) puts her toy in a box and, while she is out of that room, the other doll (Ann) changes the location of the toy putting it into a basket. The aim of this test is to ask the participant about Sally’s belief as to the location of the toy while going back to recover it (the false belief of thinking she will find it where she left it).

Modifying the previous experimental situation slightly, we can have a hypothetical virtual environment with several rooms, each containing several objects distributed in a given order that we can alter. If we are the participants and the experimenter tells us that it is a collaborative environment (where other people are supposed to be –but they are not– participating in the same way), then after altering the order or distribution of the objects in one of the rooms, if we go to another room and back to the previous one a few minutes later, finding a different distribution than the one we set up, then we may have a feeling of social presence (suspecting that another participant has changed our distribution), but again this will only be a product of our imagination.

We can say that in Presence Research where we have to set up the technology for obtaining a sense of presence: the more an experience rests on our imagination, the less robust and consistent is the presence it provides.

4.2. Social Agency supports the development of Presence and the experience of Social Presence

In the second section we reviewed the role of experiencing social agency in the development of the ‘experiencing self’. In early ages, the simple situation of being imitated by an adult can be seen as a form of socially-mediated agency. Features C and D of Russell’s Model of agency reflect the contrast between our knowledge about ourselves and our knowledge about others.

Extending Brewer’s conception of presence [6], we can understand social presence as our moment-to-moment awareness of ourselves in the process of perception of the *social* world (the word *social* is ours), which again would include both directions of the perception/action cycle: awareness and inferences about the subjectivity of the others *and* feeling that the others are aware of ourselves.

Going back again over autism, some authors (such as Russell [4]) propose that the deficits of Social Agency of these people have their origin in an impaired development of the sense of self or self-awareness. As we mentioned before, people with autism have themselves described their experience of the world as something like watching a video (Grandin [24]). If they cannot experience their feelings and reactions to the world in the first person, then it would be difficult to empathise with the feeling of others (putting themselves in the same shoes as another person).

Hobson [18] suggests that early failures in social and emotional processing lead to later failures in differentiating the self (and other) from experience, and thus their sense of presence (experiencing self) can be impaired.

Social awareness has been the focus of attention in some educational intervention with VR. In the research carried out by Parsons et al. [27], whether or not individuals with autism adhere to particular social conventions (in a Café and a Bus) in Virtual Environments was assessed. Different degrees of success were found, with results suggesting that some individuals with an ASD, low verbal IQ and weak executive ability require the most support to complete tasks successfully in the Virtual Environment. Participants in the research developed by Herrera et al [26] have also re-created (within the VE) a limited and basic range of social routines they can manage (such as greeting and saying goodbye to the employees in the Virtual Supermarket).

The origin of these appropriate behaviours towards unfamiliar (virtual) people might only be in the context of that predictable and structured way of socially interacting. The repertoire of behaviours of virtual characters in these experiments is very limited and thus predictable. Perhaps in both situations (real and virtual), people with autism are challenged to interact with social stimuli by putting their impaired (but not null) capacity for experiencing agency into practice, and differences in performance arise as a consequence of the differences in the degree of predictability and structure of those stimuli with unfamiliar people.

Social Presence is being studied in several research works on presence. The kind of measures of social presence that came from some authors of Presence Research (Garau et al. [53]) fits well with this, as they include co-presence feeling, participant behaviour in response to other agents and other agents’ awareness of the participant.

Given the fact that the children of today are becoming more and more familiarised with technology from the very beginning, and that social agency drives ‘self awareness’ development (see second section), it seems appropriate to include as many opportunities for social agency as possible in any technological product designed for children.

With co-presence being obtained from avatars or from artificial intelligence agents, our action in any environment should trigger others’ reactions and this can be adapted for each individual. Together with this, the effect of our activity must persist throughout time, not only physically but also socially: endurably affecting our relationships with other co-participants or agents.

4.3. Exercising Agency or Constructing Mental Models: Experiencing Presence or Pretending to be there?

The moment-to-moment experience of interacting in real or virtual environments can be understood in different ways, with some of them based on empirical evidence and others being just metaphorical, although they are useful for understanding some concepts.

In the research field of developmental psychology, some authors (such as Baron Cohen [54]; Leslie [55]) have proposed modular models of the mind (and theoretical conceptions of autism) that have been demonstrated to be useful for developing autonomous robots (Adams et al. [56]) but, as pointed out by Russell [57], they fail to correspond with human functioning partly because of their empirical inconsistency. Russell [4] suggested that these positions are more like philosophical doctrines rather than empirical hypotheses. Loveland [16] argues that, from an ecological psychology perspective, those tests of 'Theory of Mind' actually measure the subject's ability to perceive what a particular situation affords to another person directly, thus indicating the varied interpretation of the findings and the failure of testability in many studies.

The usefulness of these meta-representational models of mind, however, is not restricted to the development of autonomous robots or to give a view about minds. It has been demonstrated to be useful for teaching people with autism about mental states (Herrera et al [26]), by using think bubbles where mental content about imagination was represented.

In the Presence Research field, some publications have emerged around 'cognitive constructions' (Nunez [13]) and also around 'spatial constructions' (Wirth et al [58]). They consider their proposals to be attempts to understand the role of cognition in presence. The Cognitive Constructivist view (Nunez, ob cited) also includes a higher level of meaning about the environment within the model and can be useful for that possible research proximate to social psychology where constructivism plays a major role.

But does interacting with VR environments (as Nunez suggested [13]) consist of constructing and reconstructing mental models? If we were immersed in a completely novel environment (such as a virtual environment of complex molecules and DNA strings without being biochemical experts), would we quickly construct and reconstruct mental models about these 'strange things'? Even if we were experts in a given environment (such as reality), would we continuously construct and reconstruct mental models (spatial and meaningful) about everything? Would, for example, an expert car driver who is used to driving on the right side of the road (as in Spain) continuously imagine how it would be to drive this way when driving on the left side of the road (as in the UK or Ireland)?

Or, on the contrary, does this continuous construction and reconstruction of mental models (or of some of its parts) only occur in such situations as the 'power cut' where imagination plays a major role?

To answer this question, the experience of people with autism participating in VE is of great help.

Difficulties and delay in understanding symbolism, especially in relation to symbolic play, have long been documented as characteristic of people with ASDs. It is not clear whether such difficulties and delays represent a core deficit in imagination, as some have proposed, or whether they result from other aspects of autism (Jordan [59]) such as communication or social difficulties. Whether it is one origin or the other, it is commonly accepted that people with autism obtain low scores when they are asked to "act as if…" that can be measured with psychological tests such as the Test of Pretend Play (ToPP:Lewis & Boucher [60]).

There was an experiment carried out by Labajo et al [61], with a sample of 34 participants with autism (mean age: 13·6 years; mean score ABC [62] Test 63·85 points), aimed at assessing the acceptance of VR devices (VR Helmet, data-gloves and positioning-trackers) and the VR environment by people with autism. In that study, after following a period of using analogous materials (ski glasses and gloves) and providing structured information in advance, 86 per cent of the participants accepted VR and interacted naturally with the environment. With smaller samples, other studies (Strickland [63]; Herrera et al [26]; Parsons et al [27]) have also found good levels of management within virtual environments by people with autism using a variety of devices (from mere flat screens to immersive helmets).

Some authors of Presence Research (Nunez [13]; Slater [64]) have compared Presence and Pretence in the sense that presence *is taken as when the subject is acting and thinking "as if"* in the virtual environment. Although this can be accepted metaphorically speaking, if we consider that people with autism (who are not generally able to act "as if") do not find it difficult to manage in VR settings, then it seems that when we are participating in a Virtual Environment we are not pretending (at least in the sense of second-order meta-representing referred to by Leslie [55]), it seems more likely that we are just exercising our agency in relation to an approximate version of that reality we know (i.e. it is intuitive) but on a first-order (i.e. non meta representational) basis.

When participating in a VE where the experience appears to be 100% natural, this interpretation does not prevent us from rationally knowing (analytically) that it is not "real" in the sense proposed by Biocca [65] and Slater & Steed [66]. Thus, as occurs sometimes when we are dreaming, we can ask ourselves whether or not we are living a dream and then rationally analyse all the information available to obtain an answer (How did we get there?, Is that possible?, etc.).

5. Conclusions

Considering and analysing agency in the field of Presence Research has consequences both for Presence itself and for Autism research.

5.1. Conclusions for Presence Research

5.1.1. Presence is about experiencing Agency

Constructivism offers a theoretical frame for understanding presence but has the disadvantage of diverting the focus of Presence Research towards aspects that are more related to imaginary presence than to the kind of tangible presence that seems to be pursued in Presence Research. Empirical evidence coming from Cognitive Developmental Research and from the participation of people with autism in Virtual Environments suggests that presence is more related to experiencing agency than to pretending to be there.

5.1.2. Presence is a subjective concept

As a product of our individual features and our accumulated experience in relation to the world, all of us have different conceptual systems and, therefore, the same stimuli can have a different effect or meaning for any of us: presence is a subjective concept. The enormous variety of presence measures proposed by different authors or groups of people may be a consequence of this subjective characterisation.

As perception is a core component of the sense of presence, analysing presence in certain collectives such as people with sensory disability (e.g. blind or deaf people) or people with impairments in sensory perception (people with autism) helps us to better understand how presence can vary from one person to another. As concluded in the fourth section, individual differences show more deviation from the typical in those who cannot typically ‘perceive’ because they do not have the typical awareness of their actions [4] (those who have not acquired a typical sense of presence).

5.1.3. Agency generally correlates with Presence

As presence is partially rooted into agency, manipulating variables associated with agency is a way of obtaining different kinds of presence (such as “natural” or artificial presence) that offer some insight into their nature. To say there are different kinds of presence is not to make evaluations judgements. There is no kind of presence better than another; every kind of presence has its own pros and cons depending on the objectives we have. For example, the situation of experiencing a “natural” agency and presence in a VE would not be an advantage in a future where augmented agency existed, especially in those competitive collaborative virtual environments where other participants would enjoy the advantages that augmented agency would bring when compared to the natural one.

5.1.4. Implications for measuring Presence

If we accept that agency is a component or a regulating variable of presence, then we can think about adapting Agency Psychological Measures to obtain partial measures of presence.

Some authors have already established subjective measures that, although they are not specifically designed for this, slightly relate to agency. Among them we can find Rice [67] who speaks about assessing a medium’s “capacity for *immediate feedback*” (the italics are ours) which is very related to the perception/action cycle involved in exercising agency.

Any possible psychological instrument that would emerge in the future for assessing an individual’s agency could be reversed in order to assess the potentiality of a given environment for experiencing agency inside it. Instruments aimed at assessing individual relationships with the environment could be used as well.

If there is Potentiality for Agency, even with one’s arms folded, the environment should make us feel our potentiality for action; there must be technological elements to support our Perception/Action cycle in the VE. For this to happen, the system must respond to our exploratory initiations, we have to see the effects of our actions and our emotional state, we have to be felt by the other co-participants and the effect of our activity must persist throughout time both physically (persistence of the changes we produce) and socially (endurably affecting our relationships with other co-participants).

Accurately knowing what natural agency means (Russell’s features) would help us to measure the existence of “natural” presence in VE when we wish to obtain such a feeling. For this purpose, experiments that compare base-line natural presence with artificial presence would be of help. The use of Functional Magnetic Resonance would also be of help for this when used to check if the same areas of the brain are being activated when exercising agency in reality and in VE. Those technologies will also be useful to obtain more insight into the implications of presence.

As we stated before, autism can be considered as an interesting condition for Presence Research. If the researcher wants people with autism to participate in his/her experiments, whether Presence Research is the primary objective or not, then ethical concerns should be considered and collaboration of accredited professionals in autism should be ensured.

5.2. Conclusions for Autism Understanding and Intervention

Our interest in these phenomena is not just at a theoretical and philosophical level; we are concerned with establishing the conditions that can help people with autism increase their capacity to become aware of themselves and others and to learn more effectively from their experiences.

There is already evidence that high emotional involvement in a task does seem to ‘work’ in putting people with autism in touch with their own experiences (Grandin [24]; Sherratt [68]) and that explicit structure can help them interpret and deal with their experiences (Peeters [69]). There is also evidence that computer assisted learning is an effective medium for them, for a variety of reasons (Murray [70]). This fits with the work developed in the past few years by Herrera et al [71] where they have developed a VR environment for individuals with autism which has already had some success (Herrera et al [26]) in teaching individuals with autism about mental states during an intensive period of three months. If we look at this in relation to secondary mediation and the role of ‘artificial presence’ then we have the interesting case of the individual with autism being seemingly *more* involved in the virtual environment and better able to participate in it than in the ‘real’ environment. The previous experiences of people with autism participating in VR settings suggest that participating in these environments was facilitatory in developing a sense of agency and thus presence. Can they be said to show ‘presence’ in this virtual situation when they do not show ‘presence’ in reality? How can we measure ‘*Presence*’ in such cases when we have no baseline ‘presence’ with which to compare? Are the examples of agency in the virtual environment transferring to the real environment, examples of the secondary mediation drawing attention to primary mediation and in that way helping the individual develop their sense of agency and intention? Is the ‘artificial’ situation more ‘real’ to people with autism because they are more engaged in it?

The analysis of how Russell’s agency features can be manipulated (third section) leads to some other possible future developments for helping people with autism to develop agency. The possibility of obtaining a “shared subjectivity” (by modifying feature D) would be of great help for the teacher and the individual with autism to share a subjectivity and then, under the guidance of the teacher, let the user with autism increase his/her participation gradually until he/she becomes the main protagonist of that shared subjectivity.

Some authors, such as Vygotski [72,73] or Rivière [74], have suggested an interpersonal origin of some intrapersonal functions such as those in which imagination is supported. In fact, some theories such as the one of Jordan [59] suggest that social play, impaired in the condition of autism, is the confluence of two development paths that are affected in autism: the social and emotional development, and the cognitive development of play. Both paths influence each other and so it can be expected that improvements in one of the components also will have repercussions on the other. In consonance with these conceptions are the empirical results obtained by Herrera et al. (in press) when assessing the educational benefit of their VR tool for promoting play in a small sample of participants, which suggest a key role for the child’s ability to initiate social contact in the development and generalisation of cognitive play. The framework of ‘shared subjectivity’ outlined above would be a unique and valuable opportunity to teach individuals with autism to redirect their agency towards social sources. Exercises for connecting the experience with feelings and emotions should also be an intervention goal. For this aim, the work developed by Rey et al. [75] can be a base for how virtual environments can be used to induce emotions and then for teaching people with autism about contingencies between what they perceive and what they feel.

Finally, can studying the unique way that people with autism respond to secondary mediation help us understand the process in general and the factors that lead to natural or artificial presence? In autism it appears that making the environment clear and structured and giving the individual control over the speed at which it is processed, makes it accessible. Adding to this, visual cues to the thought processes underlying the person’s own agency (distinguishing functional acts, playful acts and imagination) seems to enable the person to pay attention to their own role and thus become involved in a more subjective way. The fact that the virtual environment is attractive and enjoyed by the participants may also play a role but that is hard to quantify at present. It may be that the ‘presence’ then displayed by people with autism under these conditions is not the same as others who are ‘neglecting’ the (secondary) mediation rather than (as in autism) ‘discovering’ the whole aspect of mediation. Yet it might be that this neglect is mediated by the same variables (accessibility, involvement, enjoyment) as its obverse. These are empirical questions that nevertheless lead to fascinating philosophical and psychological questions.

5.3. General conclusions: Beyond natural Presence

The experience of people with autism participating in technology-mediated and virtual realities is of double benefit: the positive outcome of each educational intervention and the insights into the implications of those technologies for the Presence Research community. As we have seen in the third section, it is possible to obtain Augmented Agency through technologies but is it possible to augment presence? Is it possible to obtain a sense of ‘being there’ that is more intense than our everyday sense of presence? Naming it *Hyperpresence*, Biocca [65] has already pointed to new alternatives for communication between individuals for obtaining it. Again, the knowledge coming from developmental psychology suggests other aspects to consider and gives fundamental cues about how to obtain such a sense of presence: As babies, when we come into the World we are not equipped with a (full) sense of presence. It is through development that we acquire such a capacity and the extension to which we develop it depends both on the environment opportunities (cultural and social) and on our personal intellectual potential. With Technology-Mediated and Virtual Realities we can improve our potential in all the variables involved in presence and we can think that, by augmenting them, we can go beyond the natural limits to acquire an Augmented Sense of Presence, as the moment-to-moment awareness of ourselves in relation to the world will be augmented.

If, while departing from null presence, we have successfully travelled across the developmental journey to reach the sense of presence that we know, could we develop it –with the appropriate stimuli– beyond the natural limits? Would the differences to our current sense of presence then be as big as the differences we have when we compare our typical experience of presence with that (still incomplete) provided by technology?

We have seen how people suffering from sensory disabilities have a sense of presence that is qualitatively different to the typical, and that the sense of presence of people with autism (who fail on interpersonal experiences) is quantitatively minor (what we may call *hypopresence*). If the sense of presence develops interpersonally, we may suspect that the way of leading it beyond its natural limits (hyperpresence) is, again, interpersonal.

Acknowledgments

Thanks to people with autism for their company and for helping us to know more about ourselves. Thanks to the members of the Autism & Learning Difficulties Group of the Robotics Institute for their collaboration in the research projects related to this paper. We would also like to thank the anonymous reviewers of this paper for their suggestions.

References

[1] J.J. Murray. Hamlet on the holodeck: The Future of narrative in cyberspace. New York: The Free Press, A Division of Simon & Schuster, Inc. 1997 ISBN:0262 631873

[2] Zahorik, P.,& Jenison, R. L. (1998). Presence as Being-in-the-World. Presence, 7 (1): 78–89.

[3] Spagnolli A., Gamberini L., Understanding the Human Involvement in Mediated Interactive Environments ‘PsychNology Journal, 2005, 3(1): 6 – 15).

[4] J. Russell. Agency: its role in mental development. Erlbaum (UK) Taylor&Francis, 1996 ISBN: 0 86377 228 5

[5] Rizzolatti G. and Arbib MA. Language within our grasp. Trends Neurosci. 1998; 21, 188-194

[6] W.F. Brewer. What is autobiographical memory? In D. Rubin (Ed.), Autobiographical memory (pp. 25-49). Cambridge: Cambridge University Press, 1986.

[7] J.S. Steuer. Defining virtual reality: Dimensions determining telepresence. Journal of Communication, 42(4), 73-93. 1992.

[8] F. Biocca, B. Delaney. Immersive Virtual Reality Technology. In F. Biocca & M. R. Levy (Eds.), Communication in the Age of Virtual Reality. (pp. 57-124). Hillsdale, NJ: Lawrence Erlbaum Associates, 1995.

[9] R. Kalawsky. A Tool for Evaluation of Contributory Factors Associated with Presence in Spatially Immersive Environments. Presented at the *BT Presence Workshop*, 10-11 June 1998. Document retrieved from the Internet, http://sgi-hursk.lut.ac.uk/~avrrc/presence/vrsart.htm.

[10] T.B. Sheridan. Musings on Telepresence and Virtual Presence. Presence: Teleoperators and Virtual Environments, 1(1), 120-125, 1992.

[11] T.B. Sheridan. Further Musings on the Psychophysics of Presence. Presence: Teleoperators and Virtual Environments, 5(2), 241-246, 1996.

[12] R.B. Welch, T.T. Blackmon, A. Liu, B.A. Mellers, L.W. Stark. The Effects of Pictorial Realism, Delay of Visual Feedback, and Observer Interactivity on the Subjective Sense of presence. Presence: Teleoperators and Virtual Environments, 5(3), 263-273, 1996.

[13] D. Nunez. A Constructivist Model of Presence. *Proceedings of Presence2004 Workshop in Valencia (Spain).* 2004.

[14] E.J. Gibson, A. Walker. Development of knowledge of visual-tactual affordances of substance. Child Development, 55, 453-460. 1984.

[15] S. S. Valenti, J.M.M. Good. Social Affordances and Interaction I. Ecological Psychology, 3 (2), 77-97. 1991.

[16] K.A. Loveland. Toward an Ecological Theory of Autism. Chapter of ‘The development of autism: Perspectives from theory and research’. New Jersey: Erlbaum Press. 2001.

[17] J.H.G. Williams, A. Whiten, T. Suddendorf, D.I. Perrett. Imitation, mirror neurons and autism Neuroscience Biobehaviour Review 25, 287-295. 2001.

[18] P. Hobson. The Cradle of Thought: Exploring the Origins of Thinking. Oxford University Press. 2002

[19] M.A.K. Halliday. Learning how to mean: Explorations in the development of language. London: Edward Arnold. 1975.

[20] A. Damasio. The feeling of what happens: body and emotion in the making of consciousness London, Harcourt Brace. 2000.

[21] Diagnostic and Statistical Manual of Mental Disorders 4th ed. American Psychiatric Association. Washington, DC: American Psychiatric Press, Inc. 1994.

[22] L. Wing, J. Gould. Severe Impairments of Social Interaction and Associated Abnormalities in Children: Epidemiology and Classification. Journal of Autism and Developmental Disorders 9: 11-30. 1979.

[23] L. Kanner. Autistic Disturbances of affective contact. Nervous Child, 2, 217.50. 1943.

[24] T. Grandin. Thinking in pictures and other reports from my life with autism. Vintage Books. Randomhouse Inc. 1995.

[25] S. Powell, R. Jordan. Being subjective about autistic thinking and learning to learn. Educational Psychology, 13, 359-370. 1993.

[26] G. Herrera, F. Alcantud, R. Jordan, A. Blanquer, G. Labajo, C. de Pablo. Development of symbolic play through the use of VR tools in children with autism. In Autism: the International Journal of Research and Practice (in Press).

[27] S. Parsons, P. Mitchell, A. Leonard. Do adolescents with autistic spectrum disorders adhere to social conventions in virtual ºenvironments? In Autism: an International Journal of Research and Practise. 2005, 9: 95-117.

[28] N. Nishitani, A. Avikeinan, R. Hari. Abnormal imitation-related cortical activation sequences in Asperger syndrome Ann. Neurology 55, 558-562. 2004.

[29] L.M. Oberman, E.M. Hubbard, J.P. McCleery, E.L. Altschuler, S. Vilayanur, V.S. Ramachandran, J.A. Pineda. EEG evidence for mirror neuron dysfunction in autism spectrum disorders Cognitive Brain Research. (in press).

[30] O. Sacks. An Anthropologist on Mars. London: Picador. 1995.

[31] O. Bogdashina. Sensory Perceptual Issues in Autism and Asperger Syndrome, 2003. ISBN 1 84310 166 1

[32] J. Nadel. Early imitation and the emergence of a sense of agency. In L. Berthouze et al. (Eds.) *Proceedings from the Fourth international Workshop on Epigenetic Robotics: modelling cognitive development in robotic systems* 117, p15-16 Genoa, 2004.

[33] S. Rogers, S. Hepburn, T. Stackhouse, and E. Wehner. Imitation performance in toddlers with autism and those with other developmental disorders. Journal of Child Psychology and Psychiatry 44(5) 763-781, 2003.

[34] R. Jordan. Autistic Spectrum Disorders: an introductory handbook for practitioners London, David Fulton. 1999.

[35] R. Jordan, S. Powell. Understanding and teaching children with autism Chichester, Wileys. 1995.

[36] C. Millward, S. Powell, D. Messer, R. Jordan. Recall for self and other in autism: children’s memory for events experienced by themselves and their peers. Journal of Autism and Developmental Disorders, 30, 15-28. 2000.

[37] S. Baron-Cohen. The Essential Difference: the truth about the male and female brain London, Penguin books. 2004.

[38] C. D. Frith, U. Frith. Elective affinities in schizophrenia and childhood autism In P.E. Bebbington (Ed.) Social Psychiatry: theory methodology and practice New Brunswick, Transaction publishers. 1991.

[39] C.R. Gallistel. The organisation of action: A new síntesis. Hillsdale, NJ: Lawrence Erlbaum Associates Inc. 1980.

[40] A. Tang, F. Biocca, L. Lim. Comparing Differences in Presence during Social Interaction in Augmented Reality versus Virtual Reality Environments: An Exploratory Study. *Proceedings of the Seventh Annual International Workshop ‘Presence 2004’*, Valencia. 2004.

[41] Marcos Fernández and Jesús Gimeno (2005) Master Thesis: Study and development of a motion capture system applied over synthetic characters in real time, using infrared cameras. Computer Science Departement, University of Valencia.

[42] Jean Pierre Jeunet (2001) Film: 'The Fabulous Destiny of Amelie Poulain'. Victories Productions, Tapioca Films – France 3 Cinema and MMC independent GmbH

[43] A. Schopenhauer. Die Welt als Wille und Vostellung I. (The world and will and representation Vol.1). New York: Dover Press. 1844/1966.

[44] T. Schubert, F. Friedmann, H. Regenbrecht. Decomposing the Sense of Presence: Factor Analytic Insights. *Proceedings of the 2nd International Workshop on Presence,* 1999.

[45] B.G. Witmer, M.J. Singer. Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence, Vol. 7, No. 3, June 1998,* 225–240.

[46] R. Held, N. Durlach. Telepresence. *Presence: Teleoperators and Virtual Environments, 1* (1), 109–112. 1992.

[47] Duncan, E. R. Miranda, and K. C. Sharman. Mind the music: Towards thought-controlled music systems. In Proceedings of the V Brazilian Symposium on Computer Music, volume 3, pages 133-141, 1998.

[48] B. O,Shaughnessy. The will (Volume II). Cambridge: Cambridge University Press. 1980.

[49] TOMB RAIDER URL: http://www.tombraider.com/

[50] T. Schubert, J. Crusius. Five Theses on the Book Problem: Presence in Books, Film and VR. Proceedings of the *Fifth Annual International Workshop Presence 2002, Porto, Portugal - October 9-11*. 2002.

[51] D. Nunez. Working Memory and Presence: Reconsidering the Role of Attention in Presence. Proceedings of *Presence2004 Workshop in Valencia (Spain).* 44-48. 2004.

[52] Wimmer and Perner (1983) Beliefs about beliefs: Representation and the constraining function of wrong beliefs in young children's understanding of deception. Cognition, 13, 103-128.

[53] M. Garau, M. Slater, D.P. Pertaub, S. Razzaque. The Responses of People to Virtual Humans in an Immersive Virtual Environment. *Presence, Vol. 14, No. 1, February 2005,* 104–116.

[54] S. Baron-Cohen. Mindblindness Cambridge, Cambridge University Press. 1995.

[55] A.M. Leslie. Pretence and Representation: The Origins of Theory of Mind. Psychological Review 94: 412-26. 1987.

[56] B. Adams, C. Breazeal, R.A. Brooks, B. Scassellati. Humanoid Robots: A new kind of Too.. IEEE Intelligent Systems. 2000.

[57] J. Russell. Cognitive theories of autism In G. Harrison & W. Owen (Eds.) Cognitive Deficits in Brain Disorders London, Erlbaum. 2002.

[58] W. Wirth, P. Vorderer, T. Hartmann, C. Klimmt, H. Schramm, S. Boecking. Constructing Presence: Towards a two-level model of the formation of Spatial Presence experiences. Presentation on the *6th Annual International Workshop on Presence - Presence 2003*, Noviembre 2003, Aalborg, Denmark.

[59] R. Jordan A review of the role of play in theory and practice in autistic spectrum disorders. Autism: the International Journal of Research and Practice (special edition), 7. 2003.

[60] V. Lewis, J. Boucher. The test of pretend play. Manual. London: Psychological Corporation. 1997.

[61] G. Labajo, G. Herrera, J. Tamarit, E. Tornero. Realidad virtual y autismo. *Actas del Congreso de AETAPI* 2000.

[62] D.A. Krug, J.R. Arick, P.J. Almond. Behavior checklist for identifying severely handicapped individuals with high levels of autistic behavior. Journal of Child Psychology and Psychiatry, 21, 221-229. 1980.

[63] D. Strickland, L.M. Marcus, G.B. Mesibov, K. Hogan. Brief report: Two case studies using virtual reality a learning tool for autistic children. Journal of Autismand Developmental Disorders, 26**(**6), 651-659. 1992.

[64] M. Slater. Presence or pretence? Poster at the Presentation at the *6th Annual International Workshop on Presence*. 2003

[65] F. Biocca. The cyborg's dilemma: Embodiment in virtual environments. Journal of CMC, 3(2), 12–26. 1997.

[66] M. Slater, A. Steed. A virtual presence counter. Presence: Teleoperators and Virtual Environments, 9(5), 413–434. 2000

[67] R.E. Rice. Task analyzability, use of new medium and effectiveness: A multi-site exploration of media richness. *Organization Science, 3*(4), 475-500. 1992.

[68] D. Sherratt. Developing pretend play in children with autism: A Case Study. Autism: The International Journal of Research and Practice. Volume 06 Issue 02 - June 2002.

[69] T. Peeters. Autism: From Theoretical Understanding to Educational Intervention (JA Majors Company, 1997) .

[70] D. Murray. Autism and information technology: therapy with computers In S. Powell & R. Jordan (Eds). Autism and Learning: a guide to good practice London, David Fulton. 1997.

[71] G. Herrera, R. Jordan, G. Labajo, J. Arnáiz. Manual de Uso de la Herramienta 'Voy a Hacer como si…' Editado por la Universidad de Valencia. (2004)

[72] L.S. Vygotski. Internalización de las funciones psicológicas superiores. En L.S. Vygotski, El desarrollo de los procesos psicológicos superiores. 1932 Barcelona: Ed. Crítica traducción de 1979.

[73] L.S. Vygotski. Play and its role in the mental development of the child Soviet Psychology, 12, 6-18, 1933, (traducción al inglés de 1966)

[74] A. Rivière. Interacción y símbolo en personas con autismo en Infancia y Aprendizaje, 8, 3-25, 1983. Reproducido en el segundo volumen de las Obras Escogidas de Angel Rivière (2003).

[75] B. Rey, M. Alcañiz, J.A. Lozano, R. Baños, C. Botella, N. Lasso, J. Osma. Emotionally controlled Virtual Environments: A New Tool for Enhancing Presence through Media Content Manipulation. *Proceedings of Presence2004 Workshop in Valencia (Spain)*. 2004.

When presence and emotion are related, and when they are not

Jonathan Freeman[1], Jane Lessiter[1], Katherine Pugh[1], & Ed Keogh[2]
[1]i2 media research, Department of Psychology, Goldsmiths College, University of London;
[2]Department of Psychology, University of Bath
{J.Freeman@gold.ac.uk, J.Lessiter@gold.ac.uk, K.Pugh@gold.ac.uk, E.M.Keogh@bath.ac.uk}

Abstract

This paper proposes a theory generating predictions for when presence and emotion are related, and when they are not. The theory explains apparently conflicting reports on the relation between presence and emotion that have appeared in the literature – some reporting a relationship, others reporting no relationship. The key assertion of the theory is that presence and emotion are related for arousing content only. For content to be arousing to a media consumer it must be perceived as personally relevant and significant – either viscerally or carrying more complex meaning. Of course, this is not to say that all content that is personally relevant and significant need be arousing – it could be personally relevant, significant and arousal reducing. In addition to explaining contradictory results previously reported in the literature on the relation between presence and emotion, the theory explains new data presented in this paper. Consistent with our own and other authors' previous reports, we present new data showing that media form manipulations (specifically screen size and participants' ability to navigate within a VE) influence presence. We did not find related influences of the same manipulations on participants' emotional responses to the media experiences. Finally, whilst we do not present our theory as evidence in favour of the evolutionary rationale for presence recently presented in the literature it is encouraging to note that the two theories are compatible.

Keywords--- **presence, emotion, arousal, significant, meaningful, navigation, screen size, media, form, content**

1. Introduction

Through combining Barfield, Zeltzer, Sheridan and Slater [1] and Lombard and Ditton's [2] definitions of presence, we have previously presented a unified definition of presence, as "a participant's sense of being there in a mediated environment, arising from a perceptual illusion of non-mediation" [3]. Determinants of how present a participant feels in a mediated environment can be categorised into user and media characteristics [4]. User characteristics can include the participant's perceptual, cognitive and motor abilities and certain personality traits – which can vary with the age and sex of the user [4].

Media characteristics can be split into media form variables and media content variables. Media from refers to the physical properties of the medium e.g. the amount of sensory information presented and the extent to which the participant can control and modify the environment [4].

Media content refers to the overall theme or story represented by the medium [4] including the objects, actors and events depicted [5], and includes the inherent interest of the content, its relevance to a person, its familiarity and its naturalness.

The majority of presence research to date has focused on evaluating the effects on presence of manipulating aspects of media form e.g., Freeman et al. [6]; Freeman et al. [7]; Welch et al. [8]; Slater and Usoh [9]; Slater Usoh, and Chrysanthou, [10]; Hendrix and Barfield, [11] and IJsselsteijn, et al. [12] – such as 3D presentation, the inclusion of shadows in VEs, screen size, and interactivity. Given that the recent study of presence has its roots in advances in computing, telecommunications and broadcast technologies, it is perhaps not surprising that technical aspects related to media form were initial research foci.

As reported in [3], and previously [13] [14], several recent studies from a range of theoretical and methodological starting points – including semiotic, phenomenological, qualitative depth interviews, and quantitative factor analytic studies - converge on a definition of presence comprising three dimensions (or factors). We have labelled the three dimensions as:

(1) Sense of Physical Space: a participant's sense of being located in a contiguous spatial environment, *determined primarily by aspects of media form;*

(2) Ecological Validity (naturalness): a participant's sense of the believability and realism of the content – that it is real; *determined by aspects of media form and media content*; and

(3) Engagement: a participant's sense of engagement and interest in the content of the mediated environment; *determined primarily by media content.*

Variations in both media form and media content have been shown previously to influence various dimensions of presence (e.g. [15]). Sense of Physical Space is primarily determined by aspects of media form, Engagement primarily by aspects of media content, and Ecological Validity (or the realness of the experience) by a mixture of the two.

Emotions are transient states of feeling, of relatively short duration, having a rapid onset; they are usually caused by specific events. They result from appraisals of the significance of what has happened for personal well-being: the more relevant an event, the more emotive it can be. Psychologists have found that a wide range of emotive stimuli can induce short term mood changes. These include films/stories [16], music [17] and emotive sentences [18].

That content can affect both emotion and (at least some dimensions of) presence suggested a need for research to understand the relation between the two. In previous psychological research on mood induction and emotion a prime focus has been on the effects of content on mood elicited. Previous research has also provided evidence for a relation between presence and emotion in certain contexts – through experimental research on presence and arousal [19], [20] and in a clinical context using virtual reality environments as therapeutic tools [21], [22].

The study used an environment which, as we reported at last year's conference, was successful at eliciting its target emotion of relaxation. In the current studies we manipulated media form aspects of the presentation of Relaxation Island – a novel virtual environment (VE), conceptualised, designed and specified by i2 media research ltd and Goldsmiths College and developed by project partners at the Interactive Institute (Sweden) as part of the EC funded project, Engaging Media for Mental Health Applications (EMMA). The environment is novel in that it combines relaxing narratives with visual and auditory representations of imagery that might also facilitate relaxation (e.g. calming sea waves, sounds of a tropical island) within an interactive VE. Previously we have reported that Relaxation Island was effective in ameliorating (negative) mood states of stress and anxiety and promoting positive mood states (of happiness and relaxation) [3]. The VE, called 'Relaxation Island', is described in detail elsewhere [23]. In brief, it comprises several *zones* ('waterfall', 'beach 1' 'beach 2', and 'cloud'). Each zone has been developed to facilitate the delivery of *instructional narratives* based on one of two theoretical approaches to modifying negative thinking and anxious mood state: standard 'controlled' breathing techniques (SBT) and narratives based on acceptance and commitment therapy (ACT; which promotes the idea of 'just noticing', 'accepting' and 'experiencing' bodily sensations rather than trying to control them). Both the ACT and SBT techniques are usually 'eyes closed' (audio-only) techniques, where participants listen to an instructional narrative and have the freedom to imagine any visual (or other sensory) element required. It has also been previously reported that when the presentation of the full VE was compared with presentation of just the ACT narrative there was a significant difference between conditions in change in relaxed mood and presence ratings, all being higher in the full VE [3].

In the current report we again focus only on conditions in which we presented participants with the ACT (and not the SBT) narrative within the VE. Our focus was on the effects on presence and emotion of allowing a participant to navigate within the full Relaxation Island environment, and of manipulating screen size within the same environment where navigation was permitted. Our key research questions were:

- What was the effect of being able to self-navigate on presence and emotion/mood?
- What was the effect of screen size on presence and emotion/mood?
- Is there a relationship between presence (a participant's sense of being there in the mediated environment) and her/ his emotional response to the content of the environment?

2. Method

2.1. Participants

Thirty participants (15 male and 15 female) aged between 18 and 43 (mean age = 25 years) took part in the study. The majority were students or staff of Goldsmiths College, University of London. All were recruited by advertising around Goldsmiths College. Participants reported to the experimenter that they satisfied the inclusion criteria that they: (a) were not taking any form of prescribed medication (except oral contraceptives), (b) were not suffering any diagnosed emotional/psychological disorder, (c) were not receiving any form of psychological therapy/counselling, (d) had normal (or corrected to normal) vision, and (e) had a good grasp of the English language. Each participant received an incentive of £10 for taking part in the study.

2.2. Design

Three independent groups (n=10 in each) were run in the study reported here, with the following conditions:

Group 1: small screen – no capacity for participant to navigate (navigation conducted by experimenter)
Group 2: small screen – navigation by participant
Group 3: large screen – navigation by participant

In all 3 groups participants were exposed only to the ACT (and not the SBT) narrative. In the results section below, we report on the influence of navigation on presence and emotion ratings (by comparing Group 1 with Group 2) and on the influence of screen size (by comparing Group 2 with Group 3).

2.3. Measures

Mood Measures

- Positive and Negative Affect Schedule (PANAS: Watson, Clark & Tellegen [24]: 20 items, 10 for the positive and 10 for the negative affect scale);
- Visual Analogue Scales (VAS: a variation of Gross & Levenson's [16] measure) assessing seven discrete emotions: happiness, anger, disgust, relaxation, fear, sadness, surprise

Presence Measures

- ITC-Sense of Presence Inventory (ITC-SOPI: Lessiter, Freeman, Keogh & Davidoff [4]: 44 items);
- UCL-Presence Questionnaire (UCL-PQ: Slater, Usoh & Steed, [25]: 3 items) - *results from the UCL PQ are not reported in this paper.*

2.4. Procedure

Prior to its commencement, this study received approval from Goldsmiths College Ethics Committee. Participants were randomly allocated to one of the 3 conditions.

Navigation (self/no self)

Participants who did not self navigate (Group 1) were simply told to keep their eyes open and that visuals were going to appear on the screen. The experimenter navigated the participant to the deckchair in the beach zone using the keyboard at the computer.

Screen size (large/small)

Viewing distance for both screen sizes: 210cm

Large screen size:

- projected size of 129cm by 96cm
- 37.5 * 28.5 degrees visual angle

Small screen size:

- projected size of 38cm by 29.5cm
- 11.0 * 8.5 degrees visual angle

The different screen sizes were created by adjusting the size of the projected image from a projector on to a large white projection screen situated 210 cm in front of the participants when they were seated in the lab. The small screen size approximates to the view of a TV screen, the larger size was defined by the limits of our equipment and laboratory. The update rate of the programme varied between 22.5 and 30 frames per second with a refresh value rate of 60 Hz, both values were the same across conditions.

On arrival at the lab, participants were taken into an office space. They were told that the study involved questionnaire completion and having an 'experience' which could involve looking at something presented to them on a screen. They were first asked to complete an Ethics Form, which requested them to agree that they satisfied a number of inclusion criteria.

After consent was obtained, participants were asked to complete a battery of pre-test questionnaires as detailed above (Section 2.3 Measures), and other psychological screening questionnaires. The emotion scales were presented immediately pre-test to account for any effects on mood of completion of the other screening questionnaires, to establish an accurate pre-test mood rating.

Participants were then taken into the laboratory. They were asked to sit on a sofa located at a distance of 210 cm from a projection screen and were handed an instruction sheet that explained: "You are about to take part in a short experience. You will either be asked to sit with your eyes open or with your eyes closed. During the experience you may or may not receive verbal instructions. If you are asked to sit with your eyes closed, please try to imagine a scene consistent with any instructions you may receive". The lights were dimmed and they were then instructed that they were to keep their eyes open.

Participants in the self-navigation conditions (Groups 2 and 3) were handed a wireless keyboard and instructed that they could use the arrow keys to move around in the environment they were about to experience. They were asked to make their way to 'beach 2' which would be signposted in front of them when the environment was displayed. The experimenter then presented the environment, and participants navigated their way to the beach zone (see Figure 1).

Participants in the no-self navigation condition (Group 1) were simply told to keep their eyes open and that an environment was going to appear on the screen. The experimenter navigated the participant to the deckchair in the beach zone using the keyboard at the computer. This was positioned to the far left of the participants allowing the experimenter to navigate the participants in Group 1 to the beach zone without being in the participants' view. The view on the screen was the same as if the participant was navigating themselves. In order that participants in both conditions took part in media experiences of approximately equal duration, and to trigger elements of the narrative (built into the VR program) the experimenter navigated through the environment from the same starting point (the signpost) and followed the same path (to 'beach 2') as did participants in the self-navigation conditions.

Figure 1 Beach Zone 2 (Relaxation Island)

On arrival at the beach zone the participant was 'seated' in a deck chair located near the sea shore with a view of the sea and a palm tree (swaying in the breeze) positioned on the right of the display. Once in the chair, the self navigation groups' navigation capacity was restricted to panning left and right. The pre-recorded narrative then began. All participants experienced the full audio-visual VE with the ACT narrative, which included the sounds of the waves and the sea lapping at the shore and the tropical sounds of birds and insects. During the experience, the experimenter remained in the room, silent and seated at the computer to the far left of the participants, out of their view.

There were four main sections of narrative each divided by long pauses to allow the participant to focus on the exercise. The narrative first welcomed the participant and commented on the presented environment (the ocean, waves, sun, breeze, golden shores). It was explained that the exercise would focus on a breathing technique. Participants were asked to just notice what their body and mind provided them with. In the second piece of narrative, participants were instructed to just notice their breathing. The act of taking a breath and exhaling was described and they were asked not to change their breathing but to simply notice it. The third section of narrative instructed that if their mind was drifting to other things, to gently bring it back to just noticing their breathing. And finally, the fourth narrative explained that they were coming to the end of the session. It was suggested that they could practice this breathing technique at any time and any place by visualizing the beach. The entire presentation in the beach zone lasted 7 minutes and 20 seconds. All participants were then instructed by the experimenter that the experience was over.

The participants then completed the post-test battery of measures, fixed in the following order: VAS, PANAS, ITC-SOPI, UCL-PQ. Participants finally completed the VAS and PANAS once again on reflection of their mood *during* the experience. They were then paid for their participation. The entire session lasted approximately one hour, including a short debrief. Participants completed the presence and mood questionnaires post-test to avoid disrupting their experience and having to divide their attention between the VE and the questionnaires. Because presence is a subjective experience it has been argued that using questionnaires that rely on subjective report is the most efficient way to measure it. [5]

3. Results

3.1. Subjective Mood

The "experience" irrespective of capacity to navigate or screen size, significantly changed discrete emotion ratings

A one sample t-test was run on all 3 groups together (n=30). The results indicated that the presentation of the VE regardless of the condition significantly increased change (post-pre) in Visual Analogue Scale (VAS) rated positive discrete mood ratings of relaxation ($t_{(29)} = 5.68$, $p < 0.01$) and significantly decreased change (post-pre) in VAS rated discrete mood ratings of anger ($t_{(29)} = -4.37$, $p < 0.01$)and sadness ($t_{(29)} = -3.5$, $p < 0.05$).

3.2. Navigation and Presence

Participants who self-navigated in the environment gave significantly higher ratings of Engagement, and significantly lower ratings of Negative Effects

An independent samples t-test was run to compare presence scores of participants who self-navigated in the environment (Group 2: small screen) compared with those who did not self-navigate in this environment (Group 1: experimenter navigated on their behalf, small screen).

In terms of the ITC-Sense of Presence Inventory (ITC-SOPI) ratings, there was a significant difference between the navigation groups on Engagement ($t_{(18)} = 2.33$; $p < 0.05$): participants who self-navigated gave significantly higher ratings than those exposed to the same experience but who did not self-navigate. There was also a significant difference in Negative Effects ratings ($t_{(18)} = -3.14$; $p < 0.01$): participants who did not self-navigate gave higher ratings than those who self-navigated. Sense of Physical Space and Ecological Validity ratings were also higher, though not significantly, among participants in the self-navigation groups (see Figure 2 and Table 1). The authors expect this to be an issue of power – if more participants took part these effects too would be significant.

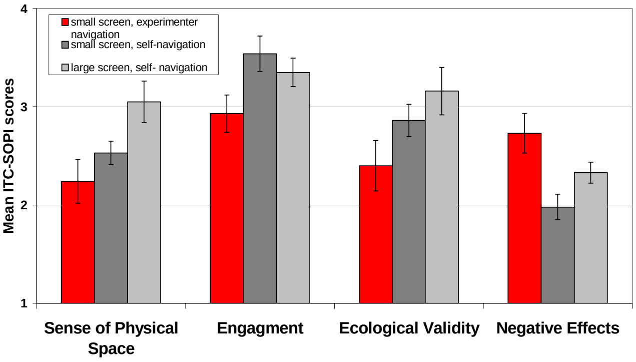


Figure 2 The effect of navigation and screen size on ITC-SOPI scores

3.3. Navigation and Emotion

There were no significant effects of self-navigation on emotion/mood

There were no significant differences between the two groups' Visual Analogue Scales (VAS) or Positive and Negative Affect Schedule (PANAS) (post-pre) change scores. Navigation did not influence the emotion/ mood change elicited by the experience (see Table 1).

3.4. Screen Size and Presence

Relative to Small screen presentations, Large screen presentations resulted in significantly higher ratings of Sense of Physical Space, and higher, though not significantly, ratings of Negative Effects

The two self-navigation conditions (Groups 2 and 3, facility to self navigate) that differed in screen size were explored to understand the effects of the screen size manipulation on presence scores.

In terms of the ITC-SOPI ratings, there was a significant difference between the screen size groups on Sense of Physical Space ($t_{(18)} = 2.15$; $p < 0.05$): participants exposed to the large screen version of the VE gave significantly higher ratings than those exposed to the same experience but on a small screen. The difference in Negative Effects ratings almost reached significance ($t_{(18)} = 2.07$; $p = 0.053$): participants presented with the large screen experience gave higher ratings than those exposed to the small screen. There was no significant difference between the screen size groups in their ratings of Engagement or Ecological Validity (see Figure 2 and Table 1).

Table 1 Presence (ITC-SOPI) and mood/emotion (post-pre) mean scores for all groups

	Group 1 **Small screen** **No self navigation**	**Group 2** **Small screen** **Self navigation**	**Group 3** **Large screen** **Self navigation**
ITC-SOPI: Sense of Physical Space	2.24 (.70)	2.53 (.38)	3.05 *(.67)*
ITC-SOPI: Engagement	2.93 (.60)	3.54 *(.57)*	3.35 *(.46)*
ITC-SOPI: Ecological Validity	2.4 (.81)	2.86 *(.52)*	3.16 *(.76)*
ITC-SOPI: Negative Effects	2.73 (.63)	1.98 *(.41)*	2.33 *(.34)*
VAS: Happiness	-1.26 (12.19)	-.59 *(22.65)*	8.07 *(11.23)*
VAS: Anger	-6.92 (7.88)	-5.57 *(6.64)*	-9.98 *(12.91)*
VAS: Disgust	-4.47 (9.05)	-2.45 *(4.37)*	-6.55 *(14.12)*
VAS: Relaxation	17.53 (27.76)	20.38 *(20.06)*	25.36 *(11.11)*
VAS: Fear	-12.19 (19.74)	-6.83 *(7.30)*	-6.55 *(13.08)*
VAS: Sadness	-8.14 (14.16)	-6.38 *(12.63)*	-12.20 *(15.47)*
VAS: Surprise	1.12 (14.16)	5.72 *(16.74)*	-.048 *(10.89)*
PANAS: Positive Affect	-4.6 (6.9)	-6.50 *(6.29)*	-3.30 *(4.76)*
PANAS: Negative Affect	-2.7 (2.21)	-1.20 *(2.70)*	-2.40 *(4.62)*

3.5. Screen Size and Emotion

There were no significant effects of screen size on emotion/mood

An independent samples t-test was run to compare mood/emotion scores of participants exposed to the large screen compared with those presented with the small screen version of the VE. There were no significant differences between the large and small screen size groups on VAS or PANAS (post-pre) change scores (see Table 1).

3.6. Relation between presence and emotion

Pearson's r correlations were run to explore the relationship between ITC-SOPI presence and VAS/PANAS emotion/mood across groups in which the presence manipulation was effective in producing significant results in the expected direction. The screen size manipulation produced significant differences in Sense of Physical Space ratings, with participants who experienced the large screen presentation reporting higher presence ratings along this dimension than participants in the small screen (self navigation) group. Correlations were run between ITC-SOPI presence and the subjective emotion ratings from participants in these two conditions.

Just one significant correlation emerged between ITC-SOPI Engagement and VAS (post-pre) Happiness ($r = .56$; $p < 0.05$; $n = 20$): as participants' engagement in the experience increased, so did their happiness ratings. This correlation is likely to be an artefact of the similarity in scale content; the Engagement scale of the ITC-SOPI includes items relating to participants' enjoyment of their media experience.

4. Discussion and Conclusions

The study described here was designed to explore the relationship between presence and emotion, in conditions where presence was manipulated through variations in media form. Participants from fully independent groups who had the capacity to navigate in an environment and who experienced the environment on a larger display - Sheridan's (1992, [26]) ability to position sensors, and extent of sensory information, respectively – gave higher presence ratings than participants who experienced the environment on a smaller display and than those who were not given the capacity to navigate within the environment. More specifically, participants who self-navigated in the VE gave higher Engagement ratings on the ITC - Sense of Presence Inventory and participants who experienced the environment on a larger display gave higher Sense of Physical Space ratings. These findings support previous research by Welch et al. [8]; Hendrix and Barfield [11] and IJsselsteijn et al. [12] who have also found that screen size and the navigation within a VE can increase presence ratings. An important point to note here is the power and usefulness of carefully designed and validated presence questionnaires in independent groups designs.

Related changes in emotion were not observed. This finding stands in contrast to previously published results in the literature showing that presence and emotion are related. Comparison of the design of the studies raised an interesting explanation, supported by other data from our laboratory. Previously we have reported no relation between presence and emotion for neutral (non-emotive) stimuli [15]. In the current study, we have reported no relation between presence and emotion for arousal reducing stimuli. In all studies where a relation between presence and emotion has been reported, arousing stimuli have been used (such as Meehan's experiments using Slater and colleagues virtual pit [20]). The theory we propose is therefore that *presence and emotion are related only for arousing stimuli*.

For content to be arousing to a media consumer it must be perceived as personally relevant and significant – either viscerally or carrying more complex meaning. Of course, this is not to say that all content that is personally relevant and significant need be arousing – it could be personally relevant, significant and arousal reducing.

We propose that there is an intuitive theoretical basis to explain why presence may enhance users' emotional responses to arousing media experiences, relating to the nature of arousal. For highly arousing stimuli a user's typical response is to be on alert, ready to respond to positive or negative events that would require an action (flight/flight or approach behaviour), and attention becomes heightened to that environment. In essence a user's ability to act, or perceived ability to act, in response to a stimulus is relevant for arousing stimuli. As presence is a user's sense of "being there" in a mediated experience – and hence their perceived ability to act within the experience - there is a rationale to expect it to be related to the extent of arousal elicited by an arousing stimulus. For non-arousing or arousal reducing stimuli there is no call to action on the user. As such, a user's perceived ability to interact with an environment, their attentional allocation to and sense of presence in relation to an environment may be expected to be less relevant for non-arousing stimuli. Of course, as per the initial expectations of the project, emotional responses to non-arousing stimuli may still be stronger in higher presence environments (as per the direction of the results presented here) but less so than are emotional responses to arousing stimuli.

The theory we propose is directly testable and falsifiable and we plan to investigate the topic further in our future research. One encouraging point to note is that it is consistent with the evolutionary rationale for presence recently presented by Riva, Waterworth and Waterworth (2004)[27].

Acknowledgements

This research was conducted within the EMMA (Engaging Media for Mental Health Applications) project: IST-2001-39192- supported by the FET Presence Research Initiative.

References

[1] Barfield, W., Zeltzer, D., Sheridan, T.B., & Slater, M., (1995). Presence and Performance Within Virtual Environments. In Barfield, W., & Furness III, T.A. (eds.) *Virtual Environments and Advanced Interface Design.* Oxford: Oxford University Press.

[2] Lombard, M., & Ditton, T. (1997). At the heart of it all: the concept of presence. *Journal of Computer Mediated Communication, 3*(2).

[3] Freeman, J., Lessiter, J., Keogh, E., Bond, F.W., & Chapman, K. (2004). *Relaxation Island: virtual and really relaxing.* Proceedings of the Seventh Annual International Workshop Presence 2004, Valencia, Spain, October 2004.

[4] Lessiter, J., Freeman, J., Keogh, E., & Davidoff, J.D. (2001). A Cross-Media Presence Questionnaire: The ITC Sense of Presence Inventory. *Presence: Teleoperators and Virtual Environments, 10*(3), 282-297.

[5] IJsselsteijn, W. A., de Ridder, H., Freeman, J., & Avons, S. (2000). *Presence: Concept, determinants and measurement.* Proceedings of the SPIE, Human Vision and Electronic Imaging V, 3959, 520-529.

[6] Freeman, J., Avons, S. E., Meddis, R., Pearson, D. E. & IJsselsteijn, W. A. (2000). Using behavioural realism to estimate presence: A study of the utility of postural responses to motion-stimuli. *Presence: Teleoperators and Virtual Environments, 9*(2), 149-165.

[7] Freeman, J., Avons, S. E., Pearson, D. E., & IJsselsteijn, W. A. (1999). Effects of sensory information and prior experience on direct subjective ratings of presence. *Presence: Teleoperators and Virtual Environments, 8*(1), 1-13.

[8] Welch, R. B., Blackman, T.T ., Liu, A., Mellers, B. A. & Stark, L. W. (1996). The effects of pictorial realism, delay of visual feedback, and observer interactivity on the subjective sense of presence. *Presence: Teleoperators and Virtual Environments, 5*(3), 263-273.

[9] Slater, M., & Usoh, M. (1993). Representations systems, perceptual position, and presence in immersive virtual environments. *Presence: Teleoperators and Virtual Environments, 2*(3), 221-233.

[10] Slater, M., Usoh, M., & Chrysanthou, Y. (1995, January). *The influence of dynamic shadows on presence in immersive virtual environments.* Presented at the Second Eurographics Workshop on Virtual Reality, Monte Carlo.

[11] Hendrix, C. & Barfield, W. (1996). Presence within Virtual Environments as a function of Visual Display Parameters. *Presence: Teleoperators and Virtual Environments, 5 (3),* 274-289.

[12] IJsselsteijn, W. A., de Ridder, H., Freeman, J., Avons, S & Bouwhuis, D. (2001) Effects of Stereoscopic Presentation, Image Motion, and Screen Size on Subjective and Objective Corroborative Measures of Presence. *Presence. 10* (3), 298-311.

[13] Freeman, J. (2003). *Converging evidence on the structure of presence.* 6th International Workshop on Presence/ Presence 2003, 6-8 October, Aalborg University, Denmark.

[14] Freeman, J. (2004). *Implications for the measurement of presence from converging evidence on its structure*. 54th International Communications Association Conference, New Orleans, USA, May 28th.

[15] Dillon, C., Keogh, E., & Freeman, J. (2002) *It's been emotional: affect, physiology and presence*. Proceedings of 5th International Workshop on Presence (Presence 2002), Porto, Portugal, October 2002.

[16] Gross J. & Levenson, R.W. (1995). Emotion elicitation using films. *Cognition and Emotion, 9,* 87-108.

[17] Clark, D.M. (1983). On the induction of depressed mood in the laboratory: Evaluation and comparison of the Velten and musical procedures. *Advanced Behavior Research and Therapy, 5,* 27-49.

[18] Velten, E. (1968). A laboratory task for induction of mood states. *Behaviour Research and Therapy, 6,* 473-482.

[19] Lombard, M., Ditton, T. B., Grabe, M. E., & Reich, R. D. (1997). The role of screen size in viewer responses to television. *Communication Reports, 10*(1), 95-107.

[20] Meehan, M. (2000). *An objective surrogate for presence physiological response.* Presented at PRESENCE 2000 - 3rd International Workshop on Presence, Delft, The Netherlands, 27-28 March 2000. [Full text available at http://www.cs.unc.edu/~meehan/presence2000/MeehanPresence2000.htm]

[21] Rothbaum B. O., Hodges L., & Smith S. (1999). Virtual reality exposure therapy abbreviated treatment manual: Fear of flying application. *Cognitive & Behavioral Practice.* 6:234-44.

[22] Riva G., Wiederhold B., & Molinari E. (Eds.). (1998). *Virtual environments in clinical psychology and neuroscience: Methods and techniques in advanced patient-therapist interaction.* Amsterdam: IOS Press. Online: http://www.cybertherapy.info/pages/book2.htm.

[23] Waterworth, E., Häggkvist, M., Jalkanen, K., Olsson, S., Waterworth, J., Antonsson, E., Chapman, K., Keogh, E., & Freeman, J. (2004*). Eyes Open or Closed: Is a virtual beach more relaxing than an imagined beach?* Presented at the Fourth International Conference on Design and Emotion, Middle East Technical University, Ankara, Turkey; 12 July 2004.

[24] Watson, D., Clark, L. A. & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54,* 1063-1070.

[25] Slater, M., Usoh, M., & Steed, A. (1994). Depth of Presence in Virtual Environments. *Presence: Teleoperators and Virtual Environments, 3,* 130 - 144.

[26] Sheridan, T.B. (1992) Musings on Telepresence and Virtual Presence. Presence: *Teleoperators and Virtual Environments, 1(1),* 120-125.

[27] Riva, G., Waterworth, J.A., & Waterworth, E.L. (2004). The Layers of Presence: a bio-cultural approach to understanding presence in natural and mediated environments. *CyberPsychology & Behavior 7*(4): 402-416.

Presence, Narrative and Schemata

Daniel M. Pinchbeck & Brett Stevens
University of Portsmouth
dan.pinchbeck@port.ac.uk, brett.stevens@port.ac.uk

Abstract

This paper documents the early stages of research into the effect of manipulating narrative upon reported sense of presence. It is argued that presence, rather than a state, should be defined as an indicator of the ongoing development of relationships of significance between the user and the perceived environmental stimuli. If the manipulation of narrative affects reported sense of presence according to existing measurement techniques, it suggests that presence is dependent upon post-perceptual constructs, such as schemata, and that a perspective that only considers presence as a "perceptual illusion of non-mediation" is flawed. This paper will conclude by setting out the empirical studies planned to explore the hypothesis that narrative, schemata and content are important factors in generating a sense of presence.

Keywords--- **Presence, Schemata, Scripts, Narrative**

1. Introduction

To view presence as either a unitary, or a uniquely perceptual construct is untenable. Instead, it appears more likely that presence is an emergent property of a combination of cognitive and perceptual processes and stimuli. Whilst recognizing Lombard & Ditton's oft-quoted definition of presence as a "perceptual illusion of non-mediation" [1], examining presence from an additional level of abstraction provides an opportunity to develop a new model that avoids the inherent contradictions and instabilities in the construct. Fundamentally, presence indicates that a relationship has been established between stimuli within an organism's frame of reference. This is true for all of the major types of presence suggested to date: self/environmental (the organism exists in relation to the environment), object (the object exists in relation to the organism) and social (other organisms exist in relation to the organism). The key factor is significance – presence indicates that the stimuli are deemed worthy of inclusion within the primary frame of reference for the organism.

Approaching presence from this perspective, several clarifications can be made. Rather than drawing a line between perceptual and cognitive processes, such as presence and absence [2] or suggesting that emotional response and presence are orthogonal [3], a definition of presence as an indicator of successful relationships of significance circumnavigates the problematic issue of how presence relates to constructs such as suspension of disbelief, attentional investment and object identification. This paper documents the early stages of developing this model. First, a theoretical framework based upon manipulation of schemata will be argued to support an indicator model of presence. Following this, empirical studies designed to test this model will be introduced.

2. A Schematic-Indicator model of Presence

2.1. Presence and evolution

The debate over the relationship between presence and mental representation is long and complex. On one hand, it has been argued that presence and content are unrelated [4]. On the other, models have been proposed that bind presence to various psychological, neurological and cognitive constructs. These include emotion [5], reality judgment [6], and "successfully supported action within an environment" [7]. Whilst it is increasingly recognized that presence cannot be successfully defined as a unitary construct [8], determining an exact mix of contributory processes is equally problematic [9]. What is obvious, however, is that any definition of presence must fulfill four basic criteria:

1. It must conform to current neurological and physiological understanding. In other words, it must fit the physical facts as we know them.
2. It must take into consideration operationally valid models of mind (including perception and cognition). That is, if it contradicts existing theory that is based upon empirical evidence it must provide justification or, at the least, explanation, for doing so.
3. It must provide means of empirical testing, i.e., there must be means to confirm or deny its claims.
4. It must provide opportunity for further development and application to systems and procedures. In other words, it must be useful.

A perceptual model of presence fulfils these criteria, but begins to fail when examined in more detail. Primarily, the flaw is logical. Presence is generally agreed to be a subjective state or, at the least, an emergent property of neurological activity. Subjectivity requires a self/other relationship to exist; it is a phenomenological state that requires some form of consciousness. Consciousness, clearly, is not reliant upon perception to occur, or sensory deprivation environments would be profoundly different experiences (the effect of long-term sensory deprivation upon consciousness is another matter). Botella et al [6] similarly argue that all perception is, in fact, mediated, so the idea of a perceptual illusion of non-mediation is internally incoherent. Secondly, it is recognized that simply increasing sensory input fidelity does not automatically imply greater presence beyond a certain threshold [4], which suggests, at the least, room for something else to be occurring. Thirdly, there is evidence to suggest that the manipulation of imposed narrative within an environment [10] and that conceptual priming [11] affect levels of presence. Finally, there is the still unresolved 'book problem'. Simply stating that reported presence from media with low immersive capabilities is not presence but something fundamentally different, if indistinguishable when using existing measures, is an unacceptable theoretical stance.

On the other hand, attempting to define presence as a distinct modular, or even combinational cognitive module or process has proved equally unsatisfactory. The fact that presence can be used to describe constructs ranging from emotion to activity has led to a situation where almost anything can be described as presence, and the field, as Slater has rightly pointed out, loses focus and meaning. Surveying the literature, the question one is most frequently confronted with is not "what is presence?" but, on the contrary, "what isn't?"

When defining presence, it is important to distinguish between first- and second-order mediation, as identified by the International Society of Presence Research's explication statement. According to the ISPR, first-order mediation is the natural process of information pick-up from an environment, via a mediating process of perception and perceptual processing. This is to be distinguished from second-order mediation, i.e., through some form of technological artefact. It is the illusion of non-mediation at the second level that concerns presence researchers. This, it could be suggested, renders Botella et al's argument redundant. There are issues with this stance however, that must be recognized. For example, does technological mediation include wearing glasses to correct optical defects? Does this mean, then, that spectacle wearers are subject to second-order mediation more-or-less permanently? By this definition, our lack of attention to the mediating technology means that we experience presence whenever our vision is thus corrected. This is clearly not a satisfactory position**.** Floridi [12] suggests that the traditional models of presence are rooted in the notion of Epistemic Failure, that is, a cognitive failure by the individual to spot the mediation. He notes that not only is there an inherent Cartesian dualism at the root of the model, but that the "conceptual reduction of a broad spectrum of phenomena to a single unifying frame of interpretation" is riddled with contradictions and inconsistencies.

The model proposed in this paper is presence as an indicator of a particular type of organization of environmental stimuli and information. What is clear from the last two decades of presence research is that perception, emotion, attention, arousal, suspension of disbelief, consistency of signal, memory, fidelity of stimuli and so on, all play a part within presence. Crucially, presence is a reported output of an emergent state, and focusing upon what purpose such a state may serve enables a circumnavigation of the issue of its exact constituent parts and processes, without rendering the outcome valueless. Tackling presence as an evolutionary development has been suggested before [13] and this approach to the issue of presence from the perspective of evolution, and an evolutionary model of consciousness, is key to a better understanding of how to work with it.

The first step in the formulation of this model is to draw a direct causal link between consciousness and presence. Without consciousness, there can be no presence, as there will be no subjective states. Sanchez-Vives & Slater [14] are amongst those who argue that presence research is an important new perspective from which to tackle consciousness studies, though they accept that the exact nature of the relationship is unclear. For example, is presence simply spatial consciousness? It can be argued that this is not the case, that consciousness is necessary for presence, but the opposite is not necessarily true – self/other relationships, and therefore subjectivity do not have to be based upon spatial relativity. Presence is a means of managing spatial consciousness.

Secondly, it should be asked what all versions of the construct have in common. The answer is evident: all are rooted in the notion of a relationship of significance being established between the self and external stimuli.

Thirdly, the evolutionary benefit in establishing and successfully managing such relationships of significance (it should be noted that a distinction is being made between 'significant' and 'attended to') parallels theories for the development of mind, cognition and consciousness put forward by cognitive scientists, neuroscientists and philosophers of mind.

To summarise, an organism that has an awareness of its own boundaries and is able to establish a conceptual relation to the environment and objects within the environment has a survival advantage over one that is environment-blind. Thus, frames of reference and the establishment of relationships of significance within them, confer a distinct evolutionary advantage. Together, they form an emergent state that positions the organism in relation to stimuli and enable information structuring, including recognition, memory and predictive trialing to occur. This model avoids the question of reality of the stimuli and with it the more complex question of whether the environmental stimuli being responded to are external / perceived, or internal / represented. All that matters is they have been incorporated into the frame of significant reference for the organism.

Presence, according to this model, is the name given to the reported output of this state in action, and we can re-conceptualize the three example types of presence referenced at the start of this paper as follows:

1. Self / Environmental - a spatial relationship of significance is established, positioning the organism within a field of stimuli.

2. Object – an object, whether ‘real’ or computer generated is incorporated into the field of significant stimuli, establishing a relationship between it and the organism that may lead to interaction (or avoidance).
3. Social - other organisms are identified within the environment and distinguished as agents, implying a different set of schematic relationships than static or mindless stimuli.

2.2. Presence and schemata management

A frame of significant reference can be viewed as a snapshot of the organism’s management of active schemata: nominally, the current state of internal affairs. Not only are objects currently deemed significant identified and mentally represented, but also the relational concepts that bind them together are active. In other words, the frame of significant reference is a perceptually orientated semantic network [15], essentially the same as a single instance of an active frameset or script [16, 17]. The notion of a frame of significant reference also fits Schank’s dynamic memory theory [18] and, although not explicitly connectionist, can easily be understood from that perspective.

These parallels are important, as they provide a means to import knowledge and models, developed by scholars investigating knowledge representation, into the presence debate. In other words, they provide means with which to codify relational content. Although we recognise that most presence researchers would credit the importance of content in generating presence, actual references to content within studies are sparse, and no attempt has been made to stratify the component elements and factors of content in relation to presence. Narrative theory provides one approach to this deconstructive process; scripts and frames suggest a complimentary approach that may assist in modeling the phenomenon. Dynamic memory, which develops these ideas further, allows a theoretical bridge to be established between presence and the organisms pre-existing tendencies, knowledge and, crucially if one sees perception as an active, directed process [19, 20], assumption and expectations.

It is assumed that at any given moment, multiple scripts will be active for any organism, although only a number of these may be attended to at any given point. This allows for modularity and parallel processing, both of which appear crucial to handling the massive information loads inherent to cognition and, more specifically, consciousness. No contradiction is apparent between the notion of a frame of significant reference (or attended script-instance) and less cognitively orientated models of consciousness, memory and information processing, such as Damasio’s somatic marker hypothesis [21]

The assumed existence of multiple scripts, along with other key assumptions of this model – the existence of a subjective conscious state, a modular semi-computational mind, active information exchange between organism and environment – all require evidential support. A short examination of the literature reveals that this support exists. With this support in place, a theory of presence based upon scripts or schematic management must then fulfill the four criteria set out at the beginning of this paper. The model proposed does indeed appear to do so.

The relationship between presence and schematic management remains to be explained. Presence, according to this model, is a perspective, a particular methodology of measurement and analysis of the existence and success of perception-orientated scripts. The underlying consensus across existing presence research is that stimuli (virtual or otherwise) may or may not trigger an emergent, subjective and psychological state, roughly conceptualized as ‘sense of being’, to a greater or lesser extent, with a broad range of factors impinging upon this emergence. This state – the active processing of relational concepts, scripts, frames, schemata or conceptual dependencies – is ongoing, and presence is a set of tools, an outlook through which to approach it.

If presence is recorded therefore, we should expect to see evidence of schematic relationships of significance occurring. Put another way, the book problem should come as no surprise and rather than being an issue, should be taken as demonstrating that virtual environments and other media share the capacity to influence an organism’s representation of its surroundings and establishment of networks of attentional resources, contextual interconnections and predictive scenarios. However, it is not as Waterworth & Waterworth state: “Presence seems to have become just another word for conscious attention. In trying to solve the so-called book and dream-state problems that baby of presence has been thrown out with the bathwater of conscious attention” [22]. According to the schemata model, if there is a shared definition, it is between conscious attention and relationships of significance within an active script-set. Presence, on the other hand, is just what it has always been, prior to and beyond the definitional debate, a series of tools that demonstrate that ‘something is occurring’ when subjects experience virtual environments and stimuli.

2.3. Presence and narrative

It is therefore suggested that what has been problematically termed presence is in fact evidence of relationships of significance amongst received stimuli. Central to this argument is the notion that these emerge from an interplay of perceptual, cognitive, emotional and experiential factors. A program of empirical studies is needed, and has been developed, to test this model. In order to further place these in context, a brief description of narrative and its relationship to presence is required.

Narrative here is defined as “the semiotic representation of a series of events meaningfully connected in a temporal and causal way” [23]. Narrative is understood as an artificial encoding of a series of linearly organized, causal relationships. There is, of course, an additional line of enquiry that needs to encompass the specific issues surrounding interactive narratives, but it falls outside the scope of this paper. In essence, a narrative is a highly formalized script the reader can accept at various levels. There are two perspectives that can be inferred from this:

the first being that mental scripts are inherently narrative. Indeed, narrative psychology, concerning itself with the "storied nature of human conduct" [24] is an established psychological perspective. The second implication is that narrative artifacts allow readers to import formalized scripts into a network of existing schematic relations. This second idea resonates with both Baar's Global Workspace Theory [25] and Dennett's reading of Gregory's Potential / Kinetic Information theory from the perspective of the evolution of mind [26].

Defining narrative as a particular subset of script, one that utilizes primarily linear, temporal and causal relationships, sites it within the overall framework of schematic frames of reference. Although keeping a flexibility of definition that describes narrative as a grammar, rather than an artifact, thus allowing it to be deployed across media (including mental architecture), the definition nevertheless enables us to explicitly test the model proposed.

A simple hypothesis can be drawn from the model: manipulation of narrative will affect presence as measured by a representative sample of existing tools. If this hypothesis is confirmed, then it suggests that presence cannot be purely perceptual. There may very well be an illusion of non-mediation, but it emerges from an engagement with content as well as form.

If presence is affected by the manipulation of narrative, it follows that what is being observed is the result of schemata, or scripts. This provides evidence to support the model of presence as a measurement of these relationships of significance.

This second inference will, of course, require independent validating: in other words, alternative tests that provide strong evidence for the existence and development of schematic relationships will need to show correlation of results. Identifying these correlational tests will be a significant challenge in the research plan.

3. Empirical Studies – a research plan

For the next twelve months, a series of research exercises and empirical studies have been planned to both test this hypothesis and investigate it in more detail. The first stage of this is to assemble a group of narrative experts from both traditional and interactive content development, who will examine a series of narrative variations on a single fabula. According to the Russian Formalist school of narrative theory, a fabula represents the actual events contained in a story, as opposed to the sjuzet, the version of events as related by the narrator. In essence, therefore, the aim of this study is to attempt to rate the impact of sjuzet manipulation upon its overall intensity. For example, a narrative may be broken down into narrator, plot, character. Each of these aspects may then be subdivided: plot into pace, causality and level of disorder [27]. These subcategories can then be individually manipulated, to create versions of a fabula to a defined brief. The same basic sequence of events can be effectively re-configured as distinct narratives. The outcome of this study will be an attempt to rate a sample of these narratives according to intensity. A separate subject group will then independently rate these for intensity to confirm the working scale.

A virtual environment will be built according to the specifications of the base fabula (i.e., it must allow for all the actions and events contained within the fabula itself). The second study in the program will pilot this environment and three of the identified narratives, agreed as having high, low and conflicting narrative intensity. The results of this study will enable a greater understanding of the constraints of the environment and subject expectations to be brought to the main study scheduled for the early part of 2006. Primarily, the pilot study will test the basic hypothesis – that a variation in narrative intensity affects presence. Confirmation of the hypothesis will then naturally lead to the question of which aspects of narrative manipulation are most crucial to this effect: is, for example, a first-person perspective more effective at generating presence than third-person; or is priming material that builds named characters into the narrative more effective than one which suggests other characters, but only describes them according to plot function?

The subsequent empirical study will present the environment to a larger subject group, who will be subdivided into narrative variation groups. Each of these narratives will be characterised by an emphasis upon a key factor identified from the initial studies and will be delivered to subjects as priming material. Subjects will be asked to complete a simple task within the environment and factors relating to their ability to do so will be measured, as will objective outcomes (time spent in environment, etc). Presence questionnaires will be used as the primary source of data. Subjects will also be videoed and their actions in the environment captured in real-time to enable further analysis of points of specific interest. A final variable in the study will be the embedding of a contradictory narrative element in one of the groups. It is already documented that contradictions in stimuli and interaction negatively affect presence [28] and the study will test whether this holds true for content-based contradictions as well.

Nunez and Blake [11] found that priming did not directly affect presence, but was a "mediating variable". A recent study by Banos et al [29] suggests that whilst imagination alone can generate presence, reinforcement from perceptual cues (i.e., a virtual representation of the same environment) is necessary to maintain it. The primary goal of the research program detailed above is to break these findings apart further and determine which aspects of content are crucial and which less so to generation and maintenance of presence. Its aimed outcome, therefore, is a practically applicable framework that VE designers can use to enhance presence. The combination of using narrative and priming to maximize the subject's own contribution to their sense of presence has clear advantages, especially in the light of the apparent ceiling of the effect of display realism upon presence [14, 30].

Manipulating narrative as a means of influencing the schemata that subjects engage with the VE through is only a small part of the field of study directly focusing upon the role of content in generating presence. Already, there is a substantial body of research examining the nature of narrative within interactive and virtual environments, with particular focus on the use of AI agents to control narrative and allow the levels of interactivity demanded by the user [31, 32]. It will be interesting to see how possible it is to control not just the illusion of non-mediation, but the illusion of freedom to act within a VE, something which games designers have been working on for some time. In a sense, it will be approaching the problem from the other direction: if presence requires the user to be able to carry out 'successfully supported action' [7] in an environment, it is worth considering how to manipulate the expectations of what actions are possible, therefore reducing the technical load on the system.

Finally, subsequent studies will need to examine the difference between methods of deploying and manipulating narrative, not just in terms of priming media (audio, text, video etc), but real-time, internal adjustments to the experience. In other words, what types and intensities of cues delivered in situ to the subject damage presence, and which further enhance it.

Conclusions

Presence is generally agreed to be a desirable emergent factor in immersive computer environments, but remains elusive in definition. Multiple models have been proposed, but although there is increasing agreement that the idea of a unitary construct is misleading, a commonly agreed set of standards is still unforthcoming. This paper has suggested that by examining the evolutionary purpose of what is currently loosely called 'presence', one finds that the construct can actually be explained as a network of relationships of significance, contextual dependencies or scripts. Presence, it is then argued, is an indication of these scripts emerging and functioning.

A set of four criteria that must underlie any theory of presence has been put forward, and it is suggested that a schematic-indicator theory of presence meets all four conditions. To provide empirical data to support this theory, a series of studies are proposed.

Stepping back from presence and viewing it at a more abstract level allows unification between current theories. It is not, as has been suggested [22], that presence has been confused with conscious attention. Rather that presence is a means for conscious organisms, by way of networks of conceptual relationships and shifting frames of significant references, to interact with their environments, real and virtual, external and internal, actual and abstract.

Acknowledgements

The authors would like to acknowledge the contribution of Dr Steve Hand and Dr Darren Van Laar to the formation of ideas contained in this paper.

References

[1] M. Lombard, T. Ditton. At the Heart of It All: The Concept of Presence. JCMC (3)2 1997

[2] J.A. Waterworth, E.L. Waterworth. Focus, Locus and Sensus: The three dimensions of virtual experience. *Cyberpsychology and Behaviour 4(2)*. 2001

[3] M. Slater. Letter to the Editors. *Cyberpsychology & Behaviour 7(1)*. 2003

[4] M. Slater. A Note on Presence Terminology. www.presence-connect.com. January 2003

[5] M.P. Huang, N.E. Alessi. Presence as an Emotional Experience. In *Medicine meets Virtual Reality: The Convergence of Physical and Informational Technologies Options for a New Era in Healthcare. Westwood, J.D., Hoffman, H.M., Robb, R.A. & Stredney, D. (eds)*. IOS Press 1999

[6] C. Botella, R.M. Banos, M. Alcaniz. A Psychological Approach to Presence. *Presented at Presence 2003, 3rd Inter. Workshop on Presence.* March 2003

[7] P. Zahorik, R.L. Jenison. Presence as Being-in-the-World. *Presence: Teleoperators and Virtual Environments 7(1).* 1998

[8] J. Freeman. Implications for the Measurement of Presence from Convergent Evidence on the Structure of Presence. *Presented at (Tele-)Presence Panel, ICA 2004* New Orleans, USA May 2004

[9] R. Kalawsky. The Validity of Presence as a Reliable Human Performance Metric in Immersive Environments. *Presented at Presence 2003, 3rd Inter. Workshop on Presence.* March 2003

[10] R.P. Darken, D. Bernatovich, J.P. Lawson, B. Peterson. Quantitative Measures of Presence in Virtual Environments: The Roles of Attention and Spatial Comprehension. *Cyberpsychology and Behaviour 2(4).* 1999

[11] D. Nunez, E. Blake. Conceptual Priming as a Determinant of Presence in Virtual Environments. In *Proceedings of the 2nd International Conference on Computer Graphics, Virtual Reality, Visualisation and Interaction in Africa (AFRIGRAPH).* 2003

[12] L. Floridi. The Philosophy of Presence: From Epistemic Failure to Successful Observability. *Presence: Teleoperators and Virtual Environments.* (in press)

[13] J.A. Waterworth, G. Riva. The Strata of Presence: Evolution, Media and Mental States. *Presented at Presence 2003.* October 2003

[14] M. V. Sanchez-Vives, M. Slater. From presence to consciousness through virtual reality. Nature Reviews: Neuroscience (6)4. 2005

[15] P. Quinlan. Semantic Memory. (1966) In *Semantic Information Processing. Minsky, M (ed.).* MIT 1986

[16] M. Minsky. A Framework for Representing Knowledge. In *The Psychology of Computer Vision. Winston P.H. (ed)* McGraw-Hill 1975

[17] R. Schank, R.P. Abelson. Scripts, Plans, Goals and Understanding. Hillsdale 1977

[18] R. Schank. Dynamic Memory. Cambridge University Press 1982

[19] J.J. Gibson. The Ecological Approach to Perception. Houghton Mifflin Company 1979

[20] Neisser, U. Cognition and Reality: Principles and Implications of Cognitive Psychology. W H Freeman and Company 1976

[21] A. Damasio. Descartes' Error: Emotion, Reason and the Human Brain. Macmillan 1994.

[22] G. Riva, J.A. Waterworth. Presence and the Self: a cognitive neuroscience approach. www.presence-connect.com. April 2003

[23] S. Onega, J. Landa (eds.). Narratology: An Introduction. Longman 1996
[24] T.R. Sarbin (ed.). Narrative Psychology: The Storied Nature of Human Conduct. Praeger 1986
[25] B.J. Baars. In The Theater of Consciousness: The Workspace of the Mind. Oxford University Press 1997
[26] D. Dennett. Darwin’s Dangerous Idea: Evolution and the Meanings of Life. Penguin 1995
[27] S. Keen. Narrative Form. Palgrave Macmillan 2003
[28] W. Barfield, D. Zeltzer, T. Sheridan, M. Slater. Presence and performance within virtual environments. In *W. Barfield and T. A. Furness, III (Eds.), Virtual environments and advanced interface design.* 1995.
[29] R.M. Banos, C. Botella, B. Guerrero, V. Liano, M. Alcaniz & B. Rey. The Third Pole of the Sense of Presence: Comparing Virtual and Imagery Spaces. *Psychology Journal* 3[1]. 2005.
[30] P. Zimmons, A. Panter. The Influence of Rendering Quality on Presence and Task Performance in a Virtual Environment. *Proceedings of the IEEE Virtual Reality 2003* 2003.
[31] R. Michael Young, M. Riedl. Towards an Architecture for Intelligent Control of Narrative in Interactive Virtual Worlds. *Presented at IUI'03, Florida, USA ACM 1-58113-586-6/03/0001.* Jan 2003.
[32] S. Louchart, R. Aylett. The Emergent Narrative theoretical investigation. *Proceedings of Narrative and Learning Environments Conference, NILE04 Edinburgh, UK.* 2004.

The role of content preference on thematic priming in virtual presence

Ilda Ladeira
CVC Laboratory
University of Cape Town
iladeira@cs.uct.ac.za

David Nunez
Dept. of Psychology
University of Cape Town
dnunez@humanities.uct.ac.za

Edwin Blake
CVC Laboratory
University of Cape Town
edwin@cs.uct.ac.za

Abstract

We set out to test the possibility that thematically priming participants with exposure to a familiar, contemporary introduction VE (with a hip-hop theme) could increase their levels of presence in a culturally unfamiliar, historical VE (a San storytelling VE). Our findings show that the relationship between priming and presence are more complex than previously thought. Specifically, for those participants who were primed with the hip-hop introductory VE, only those who chose hip-hop music as their favorite music genre derived any benefit from the introductory VE in terms of presence scores (measured on the Igroup questionnaire). This implies that thematic priming interacts with personal preference and that introductory VEs of this sort do not necessary improve the presence experience for all users.

1. Introduction

This paper reports on an interesting theoretical finding which occurred as a side-effect of an investigation on thematic priming. This investigation involved developing a virtual environment (VE) to present the oral storytelling tradition of the San people (a nomadic hunter-gatherer group indigenous to southern Africa) in an appropriate historical and physical setting. One of our goals was to create a strong sense of presence in the users of this VE.

One possible way to improve the presence experience of a VE user is to use thematic priming [1]. This method involves cognitively preparing users for a VE experience by presenting them with materials thematically related to the VE's content (the priming manipulation) prior to their experience. For high-fidelity environments, this method has been shown to increase presence scores [1]. From a presence theory perspective, thematic priming is an interesting concept. It is argued to operate when the content of the priming manipulation interacts with the content of the VE to influence presence [1]; this implies that the content of a VE can be a predictor of presence.

If presence is a function of content factors, then it is reasonable to suggest that for different users, the same VE might lead to different presence experiences. This is because content is likely to be interpreted and understood subjectively, based on an individual's previous experiences and knowledge of the content [2]. For instance, if the content is too unfamiliar or obscure, users may fail to extract enough meaning from the VE, thus compromising their presence experience. It might therefore be useful in such instances to introduce the content of the VE in more familiar terms, following the principles of constructionist learning [3]. Using familiar material as conceptual priming could theoretically provide a scaffolding of activated cognitive constructs upon which the unfamiliar VE can be understood, which would, hopefully, enrich the presence experience. Since familiarity will vary from individual to individual, it is important, from a methodological perspective, to control for this by obtaining some measure of a particular user's familiarity with the material used for priming.

From a practical point of view, we were interested in the possibility that providing some form of thematic priming could improve the presence experiences of our users. Most of our target audience would be young urban adults; thus, we were concerned that the historical San storytelling we wished to convey using VR would be too culturally remote for our users to understand. Also, as the story to be told in the VE is mythological, we were also concerned that the strangeness of the story might alienate users, and reduce their presence. In an attempt to maximize our target audience's presence, we constructed an introductory VE to preface the San VE and prime the users. This introductory VE provided some information on the San and their oral storytelling tradition.

2. Approach

To create a San storytelling VE, a traditional San fireside milieu was re-created to present the story in an appropriate historical context. This VE consisted of a large cave where a user can join a San gathering around a fire and listen to a traditional San story told by a San storyteller actor. As mentioned above, we wished to make this San VE as effective as possible and we therefore tested whether the use of a familiar introductory VE could increase presence in the San VE. Since we were aiming to convey content that was likely to be culturally unfamiliar in the San VE, we wished to investigate whether a more familiarly-themed introduction VE could improve users' experience of the San VE. We hypothesized that a contemporary theme was likely to be culturally familiar to users and might thus create an effective transition to the historical San storytelling VE.

Hip-hop was chosen as the contemporary theme for the introductory VE, since it is a well-known subculture whose style is easily recognizable. We investigated the similarities between San and hip-hop, and found that storytelling, painting, music and dance are key aspects in both. With regards to storytelling, hip-hop's mc-ing (also known as rapping) is a means of telling stories similar to the San tradition of oral storytelling performances.

3. Virtual Environments

In order to test the effect of using an introductory VE, two different virtual storytelling scenarios were created for comparison:

San VE (abbreviated as *NI*): A visual and audio desktop VE in which a San storyteller tells a San story to a gathering and the user around a fire (Figure 1). This condition had no introductory VE.

Figure 1: The historical San storytelling environment. The storyteller (left) tells the story to the audience, which responds and reacts.

In this VE, users are placed in a mountainous, outdoor environment; in the distance a gathering of San people can be seen sitting around an animated fire in a large cave. As the user walks toward the gathering, a San man from the gathering greets the user extends an invitation to join the group in listening to a story. The storyteller then begins to tell a traditional San story. The storyteller actor is animated throughout the narration; these animations were rotoscoped from the recorded actions of the actress who provided the storyteller's voice. The San gathering reacts during the story narration by exclaiming and gesturing to the storyteller and to each other.

The story was taken from a Bleek and Lloyd collection, one of the most comprehensive San story archives containing San stories transcribed from various San storytellers in the late 19th century [4, 5]. This collection contained two versions of a story about how Kagg'n, the praying mantis, created the eland and the moon. These versions were combined to form the story told in the San VE.

The cave was modeled after the Cederberg mountains of the Western Cape province in South Africa, once inhabited by the San [5]. The caves walls near the gathering were textured with San rock art which related to the story [6]. Various San objects, such hanging bags, a grinding stone and quivers were also placed in the cave. Detailed sketches were made for each character using photographs of the San. The cave, San actors and story text were all reviewed and refined with the help of an archeologist, Prof. J. Parkington, to ensure as much authenticity as possible.

San VE with Introductory VE (abbreviated as *I*): This is the NI environment described above, but was preceded by an introductory VE with a contemporary, culturally familiar theme, in this case hip-hop (Figure 2). The introductory VE consisted of an urban environment, where the user encounters a hip-hop actor, with a radio. The actor is rapping about the San people and the story that will be told in the San VE. Again, the recorded actions of a real-life actor, in this case a hip-hop musician, were rotoscoped to animate the hip-hop actor. Behind the hip-hop actor, in the VE, there are a number of graffiti covered walls and, nearby there is a door with the word "San" in graffiti on it.

As the user walks toward the actor, he stops rapping and speaks to the user; he tells them about the San and their storytelling tradition. This monologue, along with the earlier rap, serves as preparation for the content which the users will encounter in the historical San VE. He then directs the user to the door, which opens to reveal the historical San VE. The user is then able to enter the San VE described above. Thus both VE's contain content on the San, this links the introductory and San VE's thematically.

Figure 2: This hip-hop scene (in which the actor performs a rap and monologue about the San) precedes the historical San environment.

4. Study

A study was conducted to test the effect of the introductory VE; this was done by comparing the presence experienced by users in the two virtual storytelling scenarios. 58 undergraduate university students from both Science and Humanities faculties took part in this study. 30 participants experienced the San VE without an introductory VE (*NI*), while 28 experienced the San VE with the introductory VE (*I*). The mean age of participants was 20.42 years (s=1.64). 23 participants were female (40%), while 35 were male (60%).

The study took place in a quiet experimental room with four desktop computers, two running the *NI* scenario and two running the *I* scenario. Each computer ran the VEs at a frame rate of 25-30 frames per second. Four participants were taken into the experimental room at a time and were randomly assigned to the *I* or *NI* condition. The use of the mouse and keyboard for navigating the VEs was described and participants were provided with a training VE in which to practice until they felt comfortable with the controls. Participants were then informed that they would be experiencing a San story in a VE where they would be free to navigate as they wished. They were asked to put on headphones, the light in room was dimmed and participants experienced the VEs under the supervision of an experimenter.

After their virtual storytelling experience, presence (P) was measured using the Igroup Presence questionnaire. [7] Each participant's personal interest in and affinity for hip-hop was also measured. This was done since the choice of hip-hop as the introductory VE's theme was essentially arbitrary; any other well-known contemporary subculture could have been selected. Therefore, as discussed in Section 1 above, we wished to control for and consider any influence participants' personal interest, or lack of interest, in hip-hop might have had on their response to the introductory VE. Two measures of hip-hop interest (HI) were used:

- A multiple choice item asking participants to choose their favourite music genre from the following options: classical, hip-hop, alternative, rhythm and blues, rock and jazz.
- 5 items (7 point Likert-type response scale with 'fully agree' and 'fully disagree' as anchors) which measured enjoyment of hip-hop and rap music as well as familiarity with hip-hop as a popular contemporary subculture.

5. Results

The descriptive statistics for the study are shown in Table 1. A one-way analysis of variance (ANOVA) was performed using P scores as the dependent variable and condition (*I* or *NI*) as the categorical predictor. No significant effect was found (F=0.038, df=1, $p < 0.847$). However, when one considers only the participants in the I condition (i.e. those who experienced the hip-hop introduction), we see an effect on favorite music genre. A significant difference exists in favor of those who chose hip-hop as their favorite genre (N=6) over those who chose other genres as their favorite (N=24) (Mann-Whitney U=24.5, Z=2.47, $p < 0.013$). As a control for the possibility that the hip-hop preference variable was simply the manifestation of some latent factor which might affect presence directly, we conducted a similar analysis on the NI condition. This revealed no significant difference on music genre preference (hip-hop favorite: N=5; other favorite: N=23; Mann Whiteny U = 47, Z = 0.63, $p < 0.53$). A summary of the comparisons performed across the entire sample and within the NI and I conditions is shown in Table 2.

From these results, one may conclude that the difference arises only from the combination of music preference and introductory VE condition.

There were no significant relationships between presence scores and the 5 Likert items measuring enjoyment and familiarity with hip-hop culture, either across introductory VE condition, or inside either condition.

Group	Valid N	Mean	Std. Dev.
NI and I	58	59.48	12.71
NI	28	55.22	11.86
I	30	59.17	13.65

Table 1: Descriptive statistics presence (P) in the entire sample (NI and I), in the group with no introductory VE (NI) and the group with introductory VE (I).

Group	Comparison	*p*
NI and I	NI *vs.* I (ANOVA)	<0.847
NI	Hip-Hop *vs.* Other (Mann-Whitney U)	<0.53
I	***Hip-Hop vs. Other (Mann-Whitney U)***	***<0.013***

Table 2: Summary of effects on presence (P) in the entire sample (NI and I), within the group with no introductory VE (NI) and within the group with the introductory VE (I). Significant effects (*p*<0.05) are shown in bold and italic.

6. Discussion

6.1 Interaction of priming and preference

This study set out to test the effect of using a thematically familiar introductory VE (the hip-hop VE) to improve the sense of presence in an unfamiliar VE (the San VE). Instead, it revealed some interesting data with regards to the interaction of content preference and thematic priming on presence. We found that using an introductory VE as thematic priming does not necessarily serve to improve presence in all VE users. In this particular case, priming was only effective for users who had a pre-existing interest which overlapped with the thematic content of the introductory VE (the priming manipulation). This is evident from our result that participants who were primed with the introductory VE *and* showed a strong interest in the hip-hop exhibited higher presence scores than those who were primed and showed no particular interest in hip-hop. Conversely, for those who received no introductory VE, music preference had no effect. This clearly shows an interaction between thematic priming and individual content preference.

There has recently been some debate in the literature about the relative contribution of VE content to presence (see for instance [8, 9]). Our result opens up the possibility that there is an interaction at work: user preferences seem to mediate how content is processed. This interaction model, although more complex, allows for the possibility that users' previous experience, knowledge and preferences, can play a role in their presence experiences. This possibility will need to be further examined by means of studies which carefully control for the degree of previous knowledge or preference of each participant. Since there may be more at work than the combination of content preference and priming, it would be helpful to test all possible combinations of effects in order to confidently identify the factors at work. For instance, users' level of interest in priming and VE content could be measured *before* a VE experience to test their effect on presence. In the context of the study presented in this paper, this initial testing would have allowed us to test the effect of existing interest in both hip-hop and the San on presence.

We have not categorically proven an interaction of content preference and priming. Indeed, the results of this study would be more generalisable if they were replicated with different types of content. After all, it may be that individuals preferring classical music will not be affected by classically themed priming in the same way. Bear in mind however, that no priming effects were observed where any of the other music genres, classical, alternative, rhythm and blues, rock or jazz, were selected as favorites. This suggests that it was the match between participants' content preference and priming that increased overall presence and where there was a mismatch, presence was not affected.

One potential criticism of this study is that the subjects in the introductory VE condition confused the instruction of how to complete the presence questionnaire – rather than respond about the entire experience (introductory VE and San VE), they responded with regards to the introductory VE only, as the familiarity would have grabbed their attention. If the subjects responded only about their experience in the introductory VE, it still shows that their interest in hip-hop interacted with the hip-hop content of the introductory VE to increase their presence scores for the introductory VE. Thus, although one may argue that this study does not show definitively that introductory VEs are capable of producing a priming effect, it shows quite convincingly, we believe, that VE content and user experience factors interact during a presence experience.

It must also be noted that the five Likert items measuring enjoyment and familiarity with hip-hop culture showed no effect on presence. One might have expected to see these items give a similar effect as the music preference item. However, these five items do not directly measure personal preference for hip-hop, but rather the participant's perceptions of the subculture in a more objective sense (for instance, one of the items is "Fashion inspired by hip-hop culture is cool and looks good"). In other words these items were intended to measure whether subject recognized hip-hop as a familiar subculture along with their interest in hip-hop. This effectively reduces the validity of the item as a measure of personal preference for hip-hop. It is rather a measure of a combination of preference for and knowledge of hip-hop. We believe that finding only an effect on hip-hop as a favorite music genre indicates that preference for the theme of the introductory VE played a stronger role than just familiarity with the theme.

6.2 Practical implications

From a pragmatic perspective, the findings of this study are somewhat disappointing. Before conducting this study, we had hoped that a culturally familiar introductory VE would serve to improve the VE experience for all subjects. However, the fact that content preference plays a role suggests that the content of priming environments needs to be carefully chosen in order to maximize potential benefits. The data suggest that priming VEs do not necessarily increase presence for all users. If one aims the content of a VE too narrowly, one limits the potential benefits which presence may bring to the VE user. It may be that we can eliminate the role of content preference by ensuring that priming VEs are not themed too strongly; but then one may lose the potential benefit gained by appealing to users' personal interests. It may thus be preferable to examine the content interests of the user population and tailor the priming materials to match their interests. It might be possible to create several possible introductory VEs to any theme, and select one for the user based on their particular preference or experience.

References

[1] D. Nunez and E. H. Blake, *Conceptual priming as a determinant of presence in virtual environments., 2nd International Conference on Computer Graphics, Virtual Reality, Visualisation and Interaction in Africa (AFRIGRAPH 2003)*, Cape Town, South Africa, 2003, pp. 101-108.

[2] J. Bruner, *Acts of meaning*, Harvard University Press, Harvard, CN, 1990.

[3] L. B. Resnick, *Shared cognition*, in L. B. Resnick, J. M. Levine and S. D. Teasly, eds., *Perspectives on Shared Cognition*, American Psychological Association Press, 1991, pp. 1-20.

[4] L. Lloyd, W. Bleek and D. Bleek, *The Mantis and his friends: Bushman folklore.*, Maskew Miller, Cape Town, 1923.

[5] J. Parkington, *The Mantis, the Eland and the Hunter*, Creda Communications, Cape Town, 2002.

[6] S. Brown, I. Ladeira and C. Winterbottom, *The Effects of Mediation in a Storytelling Virtual Environment., International conference on storytelling 2003*, 2003, pp. 102-111.

[7] T. W. Schubert, F. Friedmann and H. T. Regenbrecht, *The experience of presence: Factor analytic insights.*, Presence: Teleoperators and Virtual Environments., 10 (2001), pp. 266-281.

[8] M. Slater, *A note on presence terminology.*, Presence Connect, 3 (2003).

[9] C. Fencott, *Presence and the Content of Virtual Environments, 2nd International Workshop on Presence*, Essex, U.K., 1999.

Presence for Sale: The Competitive Edge of Using VR

Carolina Cruz Niera
Stanley Chair in Interdisciplinary Engineering and a Professor in the Industrial and Manufacturing Systems Engineering Department

Abstract

Virtual reality is no longer a discipline exclusive to academics and researchers. Industry has realized the potential of this technology to obtain a competitive edge on the race to place products in the market. Virtual reality offers companies the ability to design, analyze, evaluate, and deploy a new product entirely within a digital world. This digital world provides a powerful communication tool in which designers, engineers, marketers, and customers can experience the product in the context pertinent to each one of them. They can carry discussions and understand each other through the visual, auditory, and even haptic product representations. Through VR, potential product flaws can be identified earlier, customers' preferences can be studied, and bolder designs can be addressed. The ability to bring humans into a digital world that "looks and feels real" and that contains an augmented reality of the product and its operation environment opens new and unexplored opportunities to leverage industry competitive expertise. This talk will review the presenter's experiences on taking VR out of the research area into the commercial world, focusing on a range of experiments conducted to characterize the critical elements of VR as a working environment for commercial products. The talk will continue with a discussion on the business value of VR and the presentation of several success stories on current commercial uses of this technology. The talk will end with a look into the future trends and expectations of VR in industry.

Session 8

23nd September, 2005

Interactivity and Usability – Theory and Practice

10.00-10.30 *Towards a Model for a Virtual Reality Experience: the Virtual Subjectiveness*
Narcís Parés and Roc Parés
Experimentation on Interactive Communication Audiovisual Institute, Universitat Pompeu Fabra Barcelona, Spain

10.30-11.00 *A Virtual Playground for the Study of the Role of Interactivity in Virtual Learning Environments*
Maria Roussou and Mel Slater
Department of Computer Science, University College London, UK

11.00-11.15 *Analysis of Subject Behavior in a Virtual Reality User Study*
Jurgen P. Schulze[1], Andrew S. Forsberg[1] and Mel Slater[2]
[1] Department of Computer Science, Brown University, USA
[2] Department of Computer Science, University College London, UK

11.15-11.30 *Exploring the relationships between the usability of a medium and the sense of Spatial Presence perceived by the user*
Andreas Nosper[1], Katharina-Maria Behr[1], Tilo Hartmann[1] and Peter Vorderer[2]
[1] Department of Journalism and Communication Research, Hanover University of Music and Drama, Germany
[2] Annenberg School for Communication, University of Southern California, Los Angeles USA

Towards a Model for a Virtual Reality Experience: the Virtual Subjectiveness

Narcís Parés, Roc Parés
Experimentation on Interactive Communication, Audiovisual Institute,
Universitat Pompeu Fabra, Barcelona, Spain
{npares | rpares}@iua.upf.edu

Abstract

After analyzing how VR experiences are modeled within Human Computer Interaction (CHI) we have found there is a deep theoretical gap. Similarly to how the scientific community has defined CHI models for multimedia applications, it would be very important to have such models for VR –obviously the standpoint cannot be the same because multimedia and VR applications differ in essence–. Indeed, there is no formal model to unify the way in which scientists and designers of VR applications define their experiences. More specifically, apart from the isolated initial scheme defined by S.R. Ellis [1][2], and a low level model defined by Latta and Oberg [3], there is no model to fully describe the relationship with which the user will be experiencing the VR application.

In this paper we shall explain where exactly we think this gap is found, which elements and concepts are involved in the definition of a model of experience and finally propose a definition of a model that we think, eventually, will fill this gap.

Keywords--- **Virtual Reality, Virtual Environment, experience, model, CHI**.

1. Introduction

Scientists have been trying to find and define the full potential of VR over the last thirty years. Many have restricted their search to the field of simulation. Others have observed that VR is more than a simulation technology and have widened their views to study the full spectrum of fields (e.g. [4][5][6][7][8]). The lack of a formal, timeless, lasting definition of VR, together with the strong emphasis on technology, have, in our opinion, hindered the definition of a model that expresses fully and coherently the relationship between a user and a VR experience. Such a model would help and guide designers, scientists and developers involved in VR to use this technology/medium in a justified and rich way. This would hopefully minimize applications where VR is used with a sensationalist and/or poor approach. For instance, it could help leisure creators in understanding how to use VR without falling in the temptation of using it as a substitute for, say, cinema or theater. It could also help scientists in designing the type of experience they really want the user to have, for her to correctly understand a concept, get a task done or get trained.

In this paper we will take a view of VR not as a mere technology, but rather as a medium. A communication medium in the sense of it being able to convey meaning, to transfer knowledge and to generate experience. We appreciate some efforts done in this direction, but we want to differentiate our position with respect to these. Specifically, we do not see it from the mass media approach as Biocca, et al. do [4][9], nor do we see VR as a point-to-point communication medium as Steuer [7] does. Therefore, our understanding of VR will be that of an interactive communication medium that is generated in real time. The intention behind approaching the analysis of VR as a medium is to be able to explore its potential distinctive properties independently from a specific technology of the moment. This, on the one hand, should avoid falling in past situations such as when VR was thought to include only those applications that used an HMD and a data glove; definitions that have later been found obsolete and have had to be reformulated to include CAVEs and many other technologies. On the other hand, if any true contribution is to be made by digital media, it must come from exploiting their new and intrinsic qualities and not from imitating or substituting what can already be done in other media [10][5][6]. Also we want to define the model of VR without being constrained by current applications in order to leave all future possibilities intact. In this sense we will not restrict our approach to simulation uses of VR.

2. Where is the gap?

In the history of VR, S.R. Ellis [1][2] defined what we could consider the strongest attempt to formalize a structural model of a VR application and the elements that participate in this structure. According to Ellis, an environment is composed of three parts: content, geometry and dynamics. Many software libraries and tools have used this model as a basis for their user-application paradigm and, although some have slight variations, the elementary concepts have not changed significantly (e.g. WorldToolKit from Sense8, Inc. [11]; DIVE from Swedish Institute of Computer Science [12]; etc.). However, this approach models only the structure of the virtual environment and does not model the "experience" of the user; i.e. it does not fully describe how the VR is "mediated" to the user. Indeed, Ellis defines content, as a set of objects and actors that are defined by state vectors which identify all their properties (physical, structural, etc.). He then distinguishes a particular actor that he calls the *self*. This self is "a distinct actor in the environment which provides a point of view from which the environment may be constructed.” ([1] pp. 322). In this definition we see that there is an initial attempt to define a structure that describes the user experience. However, this structure is not complete in the sense that he defines it as an internal element of the environment that is not specifically linked to the exterior (the user). We shall detail further these aspects below, in our model.

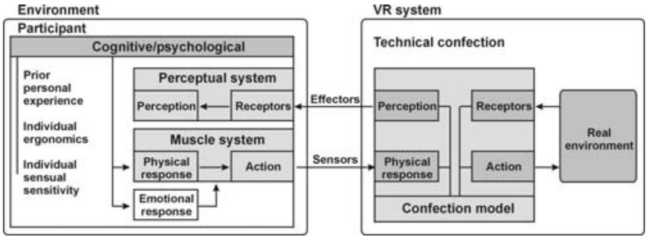


Figure 1: Diagram summarizing the “human view” and “technical view” of a VR system by Latta & Oberg [3].

On the other hand, Latta and Oberg [3] defined their “Conceptual VR Model” in such a way that, although they do emphasize a user-centered approach, they stay at a very low abstraction level of the perception of the user and therefore the “experience” is only defined at a sensorial and effector level. In other words, the experience is analyzed from the point of view of the perception of digital stimuli by the user, and how the motor system of the user may influence the management of the application. Fig. 1 summarizes their schemes from the “human view” and the “technical view” of the VR system.

In this scheme we can see how they focus on the interface at a very low description level and very much centered on the physical interfaces.

These were very interesting first approaches, however, ten years and many application areas, new technologies and new definitions of VR, have gone by since then. The gap we have found is therefore at a high level of description. We think it is important to obtain a model of how the user experience is mediated in a VR application; what can make the experience different, how can the user understand it and how she can influence the states and actions of the application.

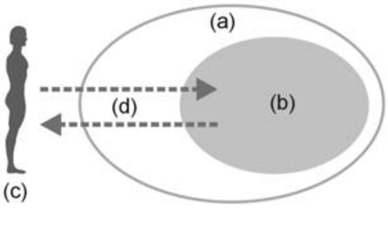


Figure 2: Simple CHI diagram of: (a) VR system (b) Virtual Environment (c) user and (d) interactive communication.

So, how can we typify this gap? Fig. 2 shows a very simple diagram of the CHI structure that relates a VR system (a) and a user (c). The user is related to the system by a two-way communication; that is, an “interactive communication” (d). It is therefore important to stress that the user is not related to the VE (b) by a mere “action/reaction” type of interaction. This is a key change in focus because by putting the accent on communication, as opposed to technology, we are explicitly referring to the exchange of information through a medium. This means that some sort of “filtering action” is giving a specific “view” of the communication experience to the user: the mediation. Hence, depending on how this mediation is defined, the user will *understand* the whole communication experience in one way or another.

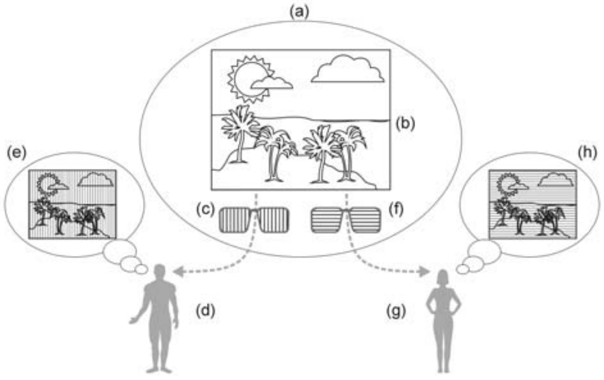


Figure 3: Simple metaphorical example of a filter that modifies user understanding in the communication experience provided by the VR system (see text for details).

To provide a simple metaphorical example, Fig. 3 shows a VR system (a) that holds the definition of a specific VE (b). As designers, we could define a “filter” within the system that would force the user to experience the environment in a particular way. In our example, we can represent this filter as a pair of virtual eyeglasses (c) that present the VE (b) to the user (d) with a vertical pattern of lines. This user would then have the impression that the VE has indeed this peculiar visual presentation (e). On the other hand, we could define another filter (f) that would force another user (g) to think that the VE (b) has a peculiar horizontal pattern associated to it (h). It must be stressed that the VR system (a) and, more importantly, the VE (b) are the same in both cases. This means that although the kernel definition of the VE may be unique, different users may have different experiences.

This example obviously oversimplifies the true situation. We need not specifically design a pair of virtual eyeglasses to give the user a filtered understanding of the experience. We could rather define a specific viewing range or a particular navigation speed and we would also be affecting the understanding of the VE by the user. Notably, notions such as “transparent interface” or “non mediated experience” have imbued VR designers, developers and scientists with the false idea that they can actually define “neutral” VR systems. By neutral we mean that these VR systems are merely simulating a “real” experience and hence, given the premise that they are indistinguishable from “reality”, these will not mediate the experience to the user. From a communication standpoint this has been proven false.

What we are trying to model in this paper is this mediation. This is the actual gap we have found and which we will call the virtual subjectiveness (VS) because it gives the notion of the user having a subjective view of the experience due to the mediating action described. Therefore, if we can formalize this mediation and we can understand how it works, we will better design and develop experiences for users.

3. The VR experience

To begin the definition of our model, we would like to propose a differentiation between the terms *Virtual Environment* and *Virtual Reality*. These terms have been historically used as synonyms; however, to our understanding, having so many terms that are used with little, if any, distinction in our field (VE, VR, telepresence, cyberspace, etc.) only impoverishes the field and causes confusion. Therefore, we propose a unique meaning for each term that we believe will enhance comprehension and concept clarification in our scientific community.

To aid us in this differentiation, we will use an analogy with two concepts from the field of simulation. According to Whicker and Sigelman [13] a *model* of a simulation is a representation of the structure to be simulated; i.e. a *static* definition that establishes structures, parameters and functions or algorithms. On the other hand, the *simulation* itself is a representation of such structure *in action*; i.e. when the “model of a simulation” is made to evolve over time, starting from an initial state, feeding it with input information and obtaining an output that is (hopefully) the desired result.

The analogy we propose is that a virtual environment (VE) be equivalent to the “model of a simulation”. Consequently, we propose that it only refers to *static* structures, i.e., a VE would include Ellis’ structural description of content (object definition), geometry (numerical database) and dynamics (static set of rules of the environment). On the other hand, we propose that virtual reality (VR) be the structures of a VE *put in action*. In other words, VR would be equivalent to *simulation* in the sense that it would refer to when the VE is made to evolve over time. Therefore, VR is the *real time experience* a user can have of a VE. In this sense, a VE could be used for a real time experience (interactive or not) or it could be used for a non-real time, off-line rendered CG animation, or for a CG single image render (e.g. for a poster). Hence, a VE is not, by itself, associated to the user (does not provide or generate anything) until it is put in action and interfaced to the user, so that it evolves over a period of time, during which the user perceives it and interacts with it.

We would like to stress here that with this analogy we are by no means trying to restrict VR to the area of simulation nor to simulation applications. On the contrary, this paper will hopefully widen the definition of VR to encompass all those applications of VR that lie outside the area of simulation and that are often left out of most existing definitions –e.g. most of the work developed by Myron Krueger and his Video Place system [14], which is widely acknowledged as being part of the realm of VR but is systematically left out by almost all definitions of VR–.

Going back to the definition of a VE given above, we have seen it fits exactly into what Ellis defined as: content, geometry and dynamics. However, we then encounter the key questions that this paper will try to answer: How can a user have an experience of this VE? How can the user understand and interact with the VE during a period of time? How is the VR experience generated and, more importantly, mediated to the user?

Of course we could take the approach of Latta and Oberg of studying how displays generate stimuli for the user, how sensors capture user actions and how the perceptual and muscle systems of the user react. However, as we have seen, this approach is too focused on the physical interfaces that link the user and the VR application; i.e. too focused on technology. We believe there is a need for a higher order conceptual model that relates more to the *semantics* of the experience rather than to the perception and the facilitators of the experience.

Another point we would like to clarify and stress in this paper is that our focus is on stimuli generated by a computer system. Therefore, we would like to differentiate our view of VR from any concept related to Telepresence - as opposed to Latta and Oberg or Steuer [7]. We understand Telepresence in its original sense defined by Marvin Minsky [15]. The rationale behind this being that on designing a telepresence application, one may have very low control over the mediation stated in the previous section. This is because it is an application that translates a physical world, the one on which the user will operate remotely, unto another physical world, that in which the user is found. Therefore we would like to define VR as: interaction with digital stimuli generated in real time. Although it may sound as a very open definition, we see three advantages in it. The first is that it does not restrict VR to 3D experiences and therefore, as stated previously, includes important 2D VR work such as that done by Krueger, the Vivid Group [16], etc. The second advantage is that it puts the accent on digital stimuli and therefore does not tie the definition to a specific VR technology. Lastly, it reinforces the idea of real time generation for interaction and this is an important property as we describe in the following section.

One final point related to the VR experience. The virtual subjectiveness we are trying to model has the goal to help us understand how to correctly define interesting and powerful experiences independent of any goal related to sense of presence. We believe this is important for a number of reasons. On the one hand, the sense of presence is not always desirable (Ellis [17] page 248) and no correlation has been yet proven to exist between task performance and sense of presence (e.g. Slater & Wilbur [18]). Also, presence is a term intimately linked to fidelity of stimuli (e.g. Sheridan [19]) and this suggests that it may only be applied to simulation applications (explained and supported by Ellis [17] page 248). Therefore, it seems like applications that lie outside the realm of simulation cannot yield a sense presence because there cannot be fidelity towards an imaginary world (in such applications it would probably be more appropriate to talk about a sense of agency[1]). Therefore, if our virtual subjectiveness model were linked to sense of presence we would not be providing a general-purpose model.

4. CHI for Multimedia & VR

We have briefly mentioned that CHI models must be different in VR to those for multimedia (MM) applications because they are different in essence. Therefore, before we go into our model, we would like to clarify how we think MM and VR differ.

MM is based on the integration of digitized media: images, sounds, videos, etc. This integration defines a situation in which all the options and data (or media) are pre-recorded; i.e. there is no real time generation of the material. On the other hand, VR is based on underlying mathematical models (not necessarily models of our physical reality) in the computer that require real time management to generate the desired final stimuli: images, sounds, etc. (Fig. 4).

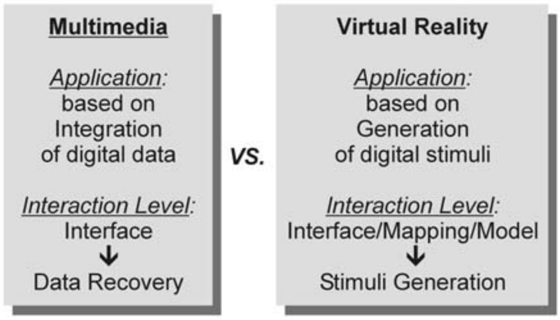


Figure 4: Essential differences between Multimedia and Virtual Reality applications from a CHI standpoint.

This is important, more than at a perceptual level, at an interaction level. In other words, at a perceptual level, the user can barely differentiate whether the application is MM or VR because we are only concentrating on how those stimuli are interfaced with our sensory-motor systems. Hence, the difference between MM and VR is important at an interaction level because a user of MM applications is confronted with a situation that can be generally described as *data recovery*, whereas the VR user is confronted with a situation of *real time stimuli generation* that is guided by three main components: interfaces, mappings and model. In other words, in MM applications the user searches for information and recovers it from different places within the application where the information is structured and predefined. In VR applications, the user experiences situations that are generated by her interaction and which allow her to explore, manipulate or contribute to the experience. Of course nowadays many hybrid applications may be found defining a continuum from MM to VR. However, for the sake of clarity we feel it is important to define a formal boundary between the two. Let us now define how this interaction in VR occurs.

5. The Interface

Before going into the kernel of our proposal, we would like to just briefly clarify one more point. This is the term and concept of *interface*. It is also a somewhat confusable term because each researcher uses it in slightly different ways and therefore, we would like to state what we understand by it. It is very common to find people referring to "the interface" of a system as mainly the physical devices that are related to input; e.g. "the interface of the application was a 3D mouse and an electromagnetic position and orientation sensor". This is one of the main problems we see with Latta and Oberg's model and proposal. Of course sometimes this is only an implicit way of speaking, but this tends to make people stop thinking of the interface as a two way system (of course, output devices are also interfaces) and, more importantly, it very often causes people to forget about the rest of the interface system: the *logical* (or software) interface and the *mappings*.

For example, we believe it is important to keep always in mind that when we interact with a PC nowadays in a windows environment (a WIMP system; *Windows, Icons, Menus and Pointing device*), we have the mouse as a physical interface, the cursor as the logical interface, a mapping between the two and a screen to visualize the results. It may sound obvious, but it must not be forgotten that without the cursor, the mouse is senseless. Also, without a proper mapping between the physical and logical units, the functionality of the interface may be useless (either too sluggish or too fast to control).

Therefore, this global idea of the interface must be seen as one of the essential parts of the entity that will allow us to define how the mediation of the user with the experience occurs. Bowman et al. [20] present useful guidelines for successful 3D user interface design. Now, on the one hand, we do not want to restrict our model to 3D applications. On the other, as they state, none of the techniques may be identified as the "best" for all situations because they are task- and environment-dependent. This is the reason why it is important to understand and describe the mediation of the experience to the user through a model such as the virtual subjectiveness.

6. The Virtual Subjectiveness: much more than an Avatar

Ellis gives a first hint of the relation "user-experience" on defining the *self* within his definition of *content*. Nevertheless, it is still an element that is clearly inside his definition of VE.

6.1. Interfaces & mappings

Let us sketch Ellis' elements in a first model that we will gradually complete. Fig. 5 shows a user confronted to a VE. In the VE we have the *logical interface*, which under

---

[1] The sense acquired when the user is conscious of being able to exert control over his surrounding environment.

Ellis' nomenclature would be the *self*. The logical interface is, according to Ellis, the definition of the "point of view from which the environment may be constructed" ([1] pp. 322). Understanding "point of view" in the widest possible sense, this is for us the first part of the definition of the mediation of the experience between the user and the VE; the logical interface may define the viewing direction, field of view, the type of projection, whether it has stereo view, the hearing capacities, force feedback properties, etc., and it may define its own appearance or representation (as in the case of an avatar). However, there is no explicit link to the user, or to the actual VE.

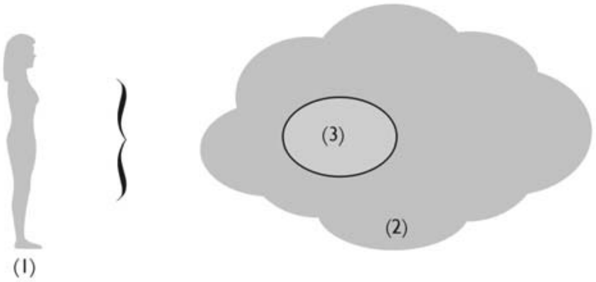


Figure 5: The user (1), the virtual environment (2) and the self (3), the logical interface.

Let us incorporate the *physical interfaces* (sensors and displays) to this model and the *mappings* that link them to the logical interface (Fig. 6). Now we are not only stating how the VE may be constructed from a point of view, but also we are explaining how this construction is linked to the user's perception (senses) and how the user's motor system is linked to the control of this point of view. This is close to what Latta and Oberg define, although here we would like to stress again the fact that when we say "how the user perceives" the environment, we are not referring to the physio-psychological processes, but rather to the cognitive semantic processes.

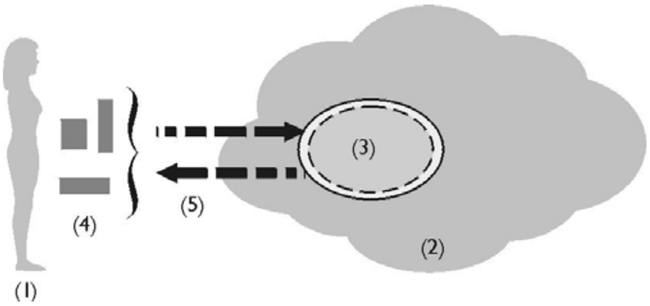


Figure 6: The user (1), linked to the logical interface (3) through the physical interfaces (4) according to the mappings (5).

For example, let us suppose an application where a user is confronted with a VE that defines a forest. Let us suppose that for the physical interface related to visual perception and point of view control, the user is provided with an HMD and a magnetic orientation sensor. Finally let us suppose a reasonable 1:1 mapping between the orientation data of the sensor and the orientation of the point of view of the logical interface; i.e. the user turns her head 90º to the right and she sees what is virtually 90º to the right of the logical interface. Now, this same forest would be understood by the user in a very different manner if the mapping between the magnetic sensor and the point of view were altered to a 1:2 mapping; i.e. the user turns her head 90º to the right and she sees what is virtually 180º to the right of the logical interface. This perception, for example, could allow a child to understand how an owl perceives its surrounding world. The owl has the physical capability to turn its head 180º, but the child does not. Of course the user could detect kinesthetic incongruence, however we can see how the definition of a non-standard mapping permits the possibility of experiencing a single VE in two extremely different ways through a VR experience (Fig. 7). It is not a property of the VE, the forest remains the same in both cases. It is rather a property of how the experience is put in action.

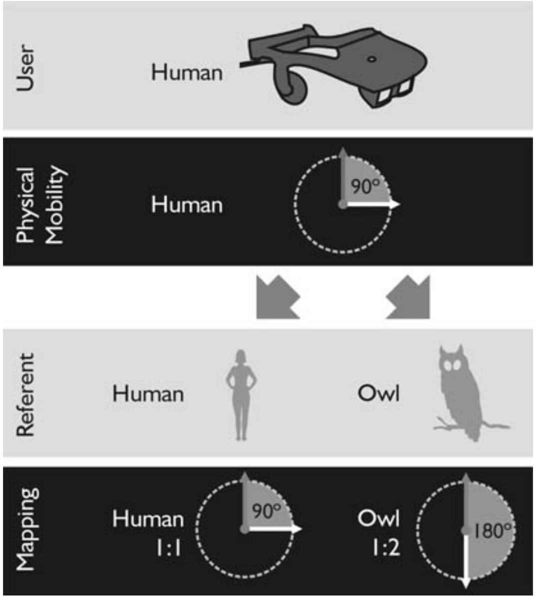


Figure 7: Two different "user to viewpoint" mappings for a single VE define different VR experiences.

Let us analyze another example. Imagine a user on a treadmill (physical interface) linked to the point of view (logical interface). If we define a mapping where, for every physically walked meter the point of view moves 10 units in the virtual environment, we are not giving any clues on what this really signifies to the user and we may be doing this definition arbitrarily. If we know the VE is 100 units in any direction, then we know the user needs to walk 10 meters to reach the opposite end of the VE. Suppose this makes any task in the experience very slow and cumbersome because it takes too long for the user (too many steps) to move from one place to the other. Now, apart from the idea of efficiency, cognitively, the user might think she is in a very large environment (Fig. 8.a). Hence the mapping does not only affect her performance and her perception of speed, but it also affects her perception of scale. Let us now modify the mapping such that now, for every 10 centimeters physically walked the point of view moves 10 units. The user might not only find that the tasks to be undertaken within the experience are now too difficult to control because of the huge speed, but also might feel the environment has suddenly shrunken because she now reaches all edges of the world with no effort (Fig. 8.b).

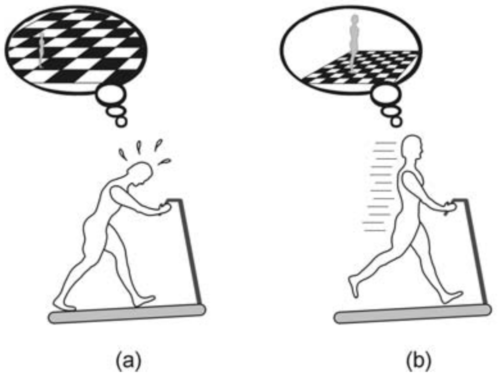


Figure 8: Different mappings make the user understand the VE in different ways (see text for details).

This relation defined by the mappings between the physical interfaces –directly linked to the user– and the logical interface –indirectly linked to the user– forming a unity of mediation, is extremely rich to work on to define the way we would like the user to perceive the experience. This is why we think this unity must be clearly identified and separated from the definition of the VE in order to help define the experience of the user.

6.2. Behaviors

Now that we have an explicit linkage of the user with the logical interface and hence have a first level of experience, we now need a complete linkage of the logical interface with the VE. This comes through the definition of the behaviors that may be associated to the logical interface with respect to the actors and other objects and elements of the VE (Fig. 9.6). These behaviors define explicitly how the user is allowed to interact with the experience through the logical interface. In other words, how the user may affect the VE.

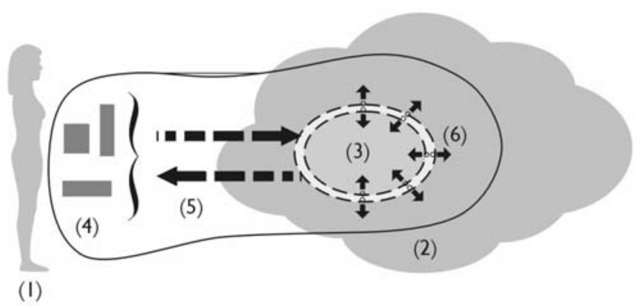


Figure 9: The Virtual Subjectiveness composed by the physical interfaces (4), the logical interface (3), the mappings (5) and the behaviors (6).

For example, the experience could be an architectural fly-through; i.e. it would be an explorative experience where the logical interface, the point of view, has no particular behavior to let the user interact with the VE (except for the real time construction of the images from a point of view). Now we could start defining behaviors for this logical interface, such as fixing height of eyesight to a certain distance above "the floor" and generating collisions with all encountered objects. The experience is still explorative however, the perception the user has is completely different; she cannot cross "walls" anymore nor can she have a bird's view of the VE. Now let us define the ability for this logical interface to have a manipulation element and be able to move objects in the VE. Again, the perception of the user changes radically when the VE still has the same definition.

6.3. The Virtual Subjectiveness

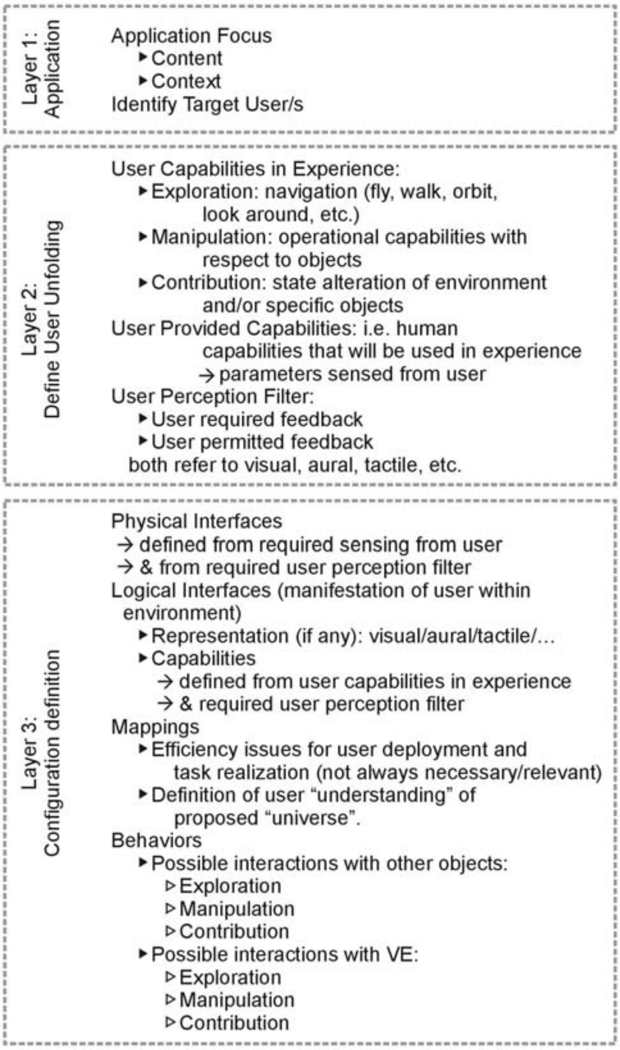
This is the final link we needed to define a whole user experience; i.e. what we call the *Virtual Subjectiveness* (VS). Specifically, the VS is the mediation element composed of the physical interfaces, the mappings, the logical interface and the behaviors (Fig. 9). The VS not only mediates the user's experience at a cognitive level defining for her a specific understanding of the VE, but also defines her complete *unfolding* within the experience. In other words, it does not only define how the user might understand the environment in which she is having the experience, but also what potentiality she has and how she can apply it in her activity and reactions within the experience. Hence, we may schematically represent the relation between the user and the VE through the VS as: *U{VS}* ←→ *VE* because the user does not directly interact with a VE to obtain an experience. The experience of the user is actually mediated by an intermediary entity that we call the VS.

This is the model we propose. A model where the accent is put on the user's relation with the VE and not on the technological interfaces, nor on the properties of the elements within the VE, nor on the physiological perception of the user. It is a relation based on the semantic information that is transmitted to the user through the VS in the interactive communication process with the VE.

7. The keystone of VR experience development

As we see it, the proposed model represents more than a theoretical advance. It provides a solid framework from which to design and analyze VR experiences. This framework may be summarized from the philosophy described above as a three-layered analysis of design requirements. Although we are still working on this top-down design scheme we present it in Fig. 10 as a preliminary reference.

In the case of simulation applications it helps in coherently defining all the properties that the user interaction must have. For example, if the application must define a virtual wind tunnel, our model helps in fully defining the role of the user within such a scientific application; i.e. the VS. Hence, it makes explicit the fact that the user might only need be an infinitely small point in space, that defines a point of view (direction and range), and a virtual smoke thread generator, another infinitely small point in space. It might sound obvious, however we often see scientific applications where the actual VS could be extremely simple in its definition, and nonetheless their authors insist on defining anthropomorphic representations and functionalities that limit the actual use of the application. By making it explicit, through our model, the scientist becomes aware of potentiality and limitations that might appear in a specific task that initially were not foreseen. Of course the user always remains a human and is hence limited by its physical and sensorial constraints, however the adequate definition of the VS, i.e. the adequate design of the mediation of the experience, may completely transform her experience.



Layer 1: Application

Application Focus
- Content
- Context

Identify Target User/s

Layer 2: Define User Unfolding

User Capabilities in Experience:
- Exploration: navigation (fly, walk, orbit, look around, etc.)
- Manipulation: operational capabilities with respect to objects
- Contribution: state alteration of environment and/or specific objects

User Provided Capabilities: i.e. human capabilities that will be used in experience
→ parameters sensed from user

User Perception Filter:
- User required feedback
- User permitted feedback

both refer to visual, aural, tactile, etc.

Layer 3: Configuration definition

Physical Interfaces
→ defined from required sensing from user
→ & from required user perception filter

Logical Interfaces (manifestation of user within environment)
- Representation (if any): visual/aural/tactile/…
- Capabilities
  → defined from user capabilities in experience
  → & required user perception filter

Mappings
- Efficiency issues for user deployment and task realization (not always necessary/relevant)
- Definition of user “understanding” of proposed “universe”.

Behaviors
- Possible interactions with other objects:
  - Exploration
  - Manipulation
  - Contribution
- Possible interactions with VE:
  - Exploration
  - Manipulation
  - Contribution

Figure 10: Preliminary framework definition for top-down VS design.

In the case of leisure or art applications the VS defines what could be considered as the only, or at least the most powerful, design tool that the “creator” might have during the design process. Let us contrast this with the creative possibilities of a cinema director who has the control of the frame to narrate, convey meaning, describe situations like mystery, etc. The director also has the control over time through cuts and editing. On the other hand, in VR an experience is, by definition, under the control of the user, because it is interactive and, very importantly, because the generation is in real time; i.e. the user chooses what to see, where to go and when to do so (unless the user is so limited and guided within the application that the experience then ceases to be a VR experience or even interactive). This apparently leaves the VR creator void of creative possibilities. However, the creator still has control over how the user is deployed within the VR application, i.e. how the user will unfold its participation, and this is done by designing the VS.

Many researchers have expressed the need to find those elements that relate the user and the experience to better understand what influences the perception, understanding, engagement, etc., of the user. For example, we believe this model comes to show that the “form”, as defined by Slater [21] and Waterworth & Waterworth [22], in which the stimuli are presented to the user is of much lesser importance than the way these stimuli are related to the semantics behind the mediation defined by the Virtual Subjectiveness; mainly because the same VE, with the same presentation form (i.e. sensor/display hardware configuration) may yield a completely different experience to the user by changing the properties of the VS.

8. Analysis of two applications

Let us now use our model to analyze two simple example applications to show the importance of a correctly designed VS. The two chosen applications are both from the leisure/entertainment area, specifically, both have been or are VR attractions of the DisneyQuest indoor amusement center at Disney World in Orlando, FL [23]. The first application will serve as an example of a faulty design of the VS, whereas the second will show a correct approach. It is a very interesting and useful situation to be able to have two real applications that are so close to one another because their analysis can then be very clearly and closely compared.

8.1. “Hercules in the Underworld” attraction

This interactive VR attraction is based on the Disney feature film “Hercules” [24]. The idea is that the users embody four characters of the film (Hercules, Phil, Meg and Pegasus) and fight the evil Hades by gathering thunderbolts from Zeus. Fig. 11 shows a schematic diagram of the attraction. In this attraction, up to four users can play (Fig. 11.a) by interacting each of them through a joystick (Fig. 11.b) and each wearing LCD shutter glasses to see the stereo images on three screens in front of them (Fig. 11.c).

Because each user embodies a different character, there is a mapping (Fig. 11.e) set from each physical interface (the joystick) to each logical interface (the character). The users see their characters in a third person view interaction scheme (Fig. 11.d). Hence, the user may explore the surrounding environment by moving his/her character around.

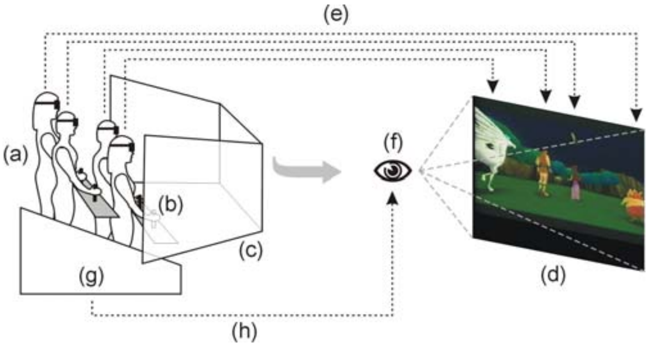


Figure 11: Schematic diagram of the "Hercules in the Underworld" attraction (see text for details).

Now, the issue in this set up is that the user in fact, when moving the character, finds himself limited in movement to a certain area around an imaginary central point of the group of characters. Apparently, many users did not understand this limitation and hence, many pushed the joystick strongly to try to move the character further away (causing robustness problems in the physical interface) [25].

Let us analyze the situation. On starting the experience, users get the impression that it is indeed a four multiuser game because they find four distinct joysticks and four characters on screen; i.e. a reasonable conceptual link between physical input interfaces and logical interfaces. When they make small movements of the character through their joystick, this scheme is reinforced, because of this 1:1 relation, and each user begins to want to explore more of the environment, possibly each in a different direction; i.e. a reasonable user understanding of logical interface behavior. However, this mental model [26] that the user applies is not correct because the third person view that the four users have of their character is in fact a common view of "the group" of characters; schematically shown in Fig. 11.f. What actually happens is that the four users move around the environment as a group and not as individuals. This group cannot divide itself because there is a single set of screens for the four users (there is a mismatch between the apparent amount of logical interfaces and the unique physical output interface). Therefore, when the game leads the users to move in one direction because of the occurring action, they have to move all together in that direction. In other words, the logical interface is actually divided in two from the vision of each user, namely: a unique representation of each character with a limited behavior and a group point of view that holds the basic navigation potential. Between the group of users (the whole set of joysticks; the physical interface)(Fig. 11.g) and the point of view from which the images are generated (the actual logical interface that gives user reference) (Fig. 11.g) a mapping (Fig. 11.h) is defined that actually reflects the activity VS. This is why the individual characters may only have a limited range of action around this invisible and abstract idea of "the group".

In terms of the VS we find it is not a unique entity that clearly defines the experience. The set of interfaces, mappings and behaviors belong to two clashing definitions or strategies, and hence confuse the user because they are not compatible. One gives the user the impression of being an individual entity that can freely explore and interact with the experience within a multiuser application. The other gives the sense of being a group of users moving together within an environment that may only be explored by the collaboration of the entire group (a group application).

8.2. "Pirates of the Caribbean: Battle for Buccaneer Gold" attraction

This interactive VR attraction is based on the famous Disney classical attraction "Pirates of the Caribbean" [27]. Here, the idea is that the users become pirate sailors in a ship that must attack and pillage other ships and towns in an environment composed of three islands and the sea areas between them [28]. Fig. 12 shows a schematic diagram of the attraction. This attraction has been conceived for up to four users acting as: the captain and three gunners (Fig. 12.a).

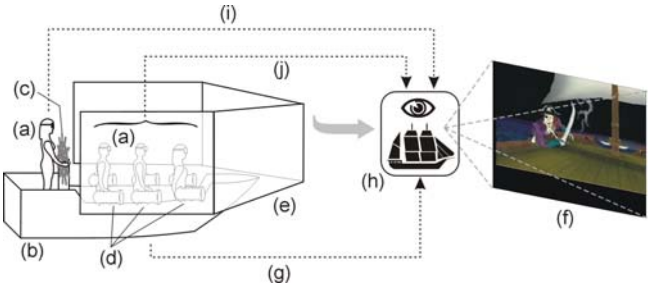


Figure 12: Schematic diagram of the "Pirates of the Caribbean: Battle for Buccaneer Gold" attraction (see text for details).

The attraction is based on a motion platform that defines part of the deck of the pirate ship (Fig. 12.b). The user that acts as the captain leads the ship with a physical helm (Fig. 12.c). The other three users, the gunners, may shoot at other ships or towns with six physical cannons that are placed three on each side of the ship (Fig. 12.d). The users see the environment through LCD shutter glasses on the stereo screens around them (Fig. 12.e). They see the images of the environment in a first person view scheme from the ship (Fig. 12.h) such that part of the ship is virtual within the images (Fig. 12.f).

The VS has a complex physical interface, the physical ship, which is composed by several elements (both input and output), namely: the motion platform, the helm, the cannons and the screens. It also has several elements that form the logical interface, the virtual ship (Fig. 12.h). Specifically: the hull of the ship, the cannonballs and the point of view centered on the ship. There are a set of mappings defined between the elements of the physical interface and those of the logical interface. For example, there is a mapping between the helm of the physical ship and the rudder of the virtual ship (Fig. 12.i) that relates equivalent rotations. There is also another mapping between the physical cannons and the invisible virtual cannons (Fig. 12.j), also related by rotations to orient the shots properly. Finally, there is a mapping between the hull of the virtual ship and the motion platform (Fig. 12.g) such that any movement of the virtual ship is translated to the physical ship. Apart from this, we must also consider the behaviors that are defined for the logical interface. Some of these behaviors are: the collisions of the hull of the virtual ship against the waves of the virtual sea; the shooting of the virtual cannonballs from the virtual cannons and how these cannonballs affect other ships; the action of the rudder against the virtual sea to change direction of the ship; etc.

The success of the design comes from the fact that, although the VS is a sophisticated structure, it is very clearly defined and identified: “the ship”. This is a crucial point because the users immediately understand that they all constitute a single unity within the game. Although they are four distinct users they are working together to be a single unity within the experience. The captain does not lead the ship independently from the gunners. Moreover, the change in direction of the ship affects the viewpoint of all the users. Any treasures won are counted on a common score, etc. Therefore, the success of this application is that the mental model that the users have is correctly matched to the VS of the experience.

We can clearly see how the “Hercules” application has not followed a consistent design of the VS for the users and therefore the experience is incoherent for them. On the other hand, the “Pirates” application is extremely solid in its conception and therefore mediates the experience very well to the users.

Conclusions

In this paper, we have presented a preliminary model to describe a VR experience with respect to: how the user is confronted to it, how she perceives it and understands it, and how this can help, not only in formalizing the properties of such experiences, but also in designing new experiences. This has been done through explicit differentiation of the terms virtual environment (VE), which is the static definition of structures, and virtual reality (VR), which is the actual experience of the VE when this latter is put in action and related to the user. The model then defines the key element of the relationship user-experience: the *Virtual Subjectiveness* (VS), a high level element that fully links the user and the VE in a VR experience: $U\{VS\} \leftarrow\rightarrow VE$. This VS is composed of:

- the *logical interface*,
- the *physical interface*,
- the *mappings* between them and
- the *behaviors* of the whole set (especially of the logical interface) with respect to the VE,

thus generating the experience.

We have briefly given a possible initial framework although it must still be elaborated and detailed in order to become a useful design and analysis tool. This ongoing research should also lead us to finally understand the underlying processes that control the mediation of the experience to the user. However, it already gives a clearer description of user experience, unlinking it from any specific VR technology and not restricting these experiences to the area of simulation applications, therefore leaving open the full range of possible experiences. It must also be analyzed theoretically to see how it may help in clarifying the specific properties of VR as a communication medium. This should lead us further in the process towards obtaining a solid model for a virtual reality experience.

Acknowledgements

We would like to thank the Audiovisual Institute of the Universitat Pompeu Fabra for providing the fruitful and rich environment that has made possible the research that has led to the present paper.

References

[1] S.R. Ellis. Nature and Origins of Virtual Environments: a Bibliographical Essay. In *Computing Systems in Engineering, Vol. 2, No. 4*, pp. 321-347, 1991.

[2] S.R. Ellis. Prologue. In *Pictorial Communication in Virtual and Real Environments, 2nd Edition* (pp. 3-11). London: Taylor & Francis Ltd., 1993

[3] J.N. Latta, D.J. Oberg. A Conceptual Virtual Reality Model. In *IEEE Computer Graphics & Applications, vol. 14*, p. 23-29, January 1994.

[4] F. Biocca, T. Kim, M.R. Levy. The Vision of Virtual Reality. In F. Biocca, M.R. Levy (ed.) *Communication in the Age of Virtual Reality*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc. pp. 3-14, 1995.

[5] N. Parés, R. Parés. Galeria Virtual. In *Proceedings First Eurographics Workshop on Virtual Environments*. Eurographics, Barcelona, 1993.

[6] N. Parés, R. Parés. Galeria Virtual: a platform for non-limitation in contemporary art. In R.A. Earnshaw, J.A. Vince, H. Jones. (ed.) *Virtual Reality Applications*. Academic Press Ltd., London, 1995.

[7] J. Steuer. Defining Virtual Reality: Dimensions Determining Telepresence. In *Journal of Communication. Austin TX: International Communication Association. Vol. 42*, pp. 73-93, 1992.

[8] V.E. Stone. Social Interaction and Social Development in Virtual Environments. In *Presence: Teleoperators and Virtual Environments, vol. 2, No. 2*, pp. 153-161, 1993.

[9] F. Biocca, M.R. Levy. Virtual Reality as a Communication System. In F. Biocca, M.R. Levy (ed.) *Communication in the Age of Virtual Reality*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc. pp. 15-31, 1995.

[10] P. Brown. The ethics and aesthetics of the image interface. In *ACM SIGGRAPH Computer Graphics, Vol. 28, Issue 1*, pp. 28 – 30, 1994.

[11] WorldToolKit. Sense8®. URL: http://sense8.sierraweb.net/ (last accessed May 2005).

[12] C. Carlsson and O. Hagsand, "DIVE - A Platform for Multi-User Virtual Environments, Computers and Graphics 17(6), 1993. URL: http://www.sics.se/dive/ (last accessed May 2005).

[13] M.L. Whicker, L. Sigelman. *Computer simulation applications an introduction*. Newbury Park, CA: Sage, 1991.

[14] M. Krueger. Videoplace - an artificial reality. In *ACM Conference on Human Factors in Computing Systems, April* 1985.

[15] M. Minsky. Telepresence. In *OMNI Magazine, 1980*. pp. 45-51, 1980.

[16] V.J. Vincent. The Mandala virtual reality system: the vivid group. In *Proceedings of the 3rd annual virtual reality conference and exhibition on VR becomes a business*. San Jose, CA. Meckler Corporation, pp. 167-170, 1993.

[17] S.R. Ellis. Presence of Mind. A reaction to Thomas Sheridan's 'Further Musings on the Psychophysics of Presence'. In *Presence: Teleoperators and Virtual Environments, Vol. 5, No. 2*, pp. 247-259, 1996.

[18] M. Slater, & S. Wilbur. A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. In *Presence: Teleoperators and Virtual Environments, Vol. 6, No. 6*, pp. 603 - 616, 1997.

[19] T. B. Sheridan. Further Musings on the Psychophysics of Presence. In *Presence: Teleoperators and Virtual Environments, Vol. 1, No. 1*, pp. 120-126, 1996.

[20] D. Bowman, E. Kruijff, J. LaViola, I. Poupyrev An introduction to 3-D user interface design. In *Presence: Teleoperators and Virtual Environments, Vol. 10, No. 1*, pp. 75-95, 2001

[21] M. Slater. A Note on Presence Terminology. In *Presence-Connect, Vol3(Jan)*, 2003.

[22] J. Waterworth, E. Waterworth. The Core of Presence: Presence as Perceptual Illusion. In *Presence-Connect, Vol 3(Jul)*, 2003.

[23] DisneyQuest. Walt Disney World, Orlando, FL. URL: http://www.disneyquest.com/ (last accessed May 2005), 1998.

[24] Hercules. Dir. R. Clememnts, J. Musker, writ. R. Clememnts, J. Musker, A. Dewey, K. Halland, N. Tobin. Walt Disney Pictures, 1997.

[25] M. Mine. Towards Virtual Reality for the Masses: 10 Years of Research at Disney's VR Studio. (Keynote) In *Proceedings of the workshop on Virtual environments 2002 & EGVR02*, ACM, Barcelona, 2002.

[26] D.A. Norman. *The Design of Everyday Things*. Basic Books, New York, 1988.

[27] Pirates of the Caribbean, Disneyland, Anaheim, CA. URL: http://www.disneyland.com/ (last accessed May 2005), 1967.

[28] J. Schell, J. Shochet. Designing Interactive Theme Park Rides. In *IEEE Computer Graphics and Applications,Vol. 21, N. 4 (Jul/Aug)*, pp. 11-13, 2001.

A Virtual Playground for the Study of the Role of Interactivity in Virtual Learning Environments

Maria Roussou and Mel Slater
Department of Computer Science
University College London
London, UK
m.roussou@cs.ucl.ac.uk, m.slater@cs.ucl.ac.uk

Abstract

Interactivity is regarded as one of the core components of a successful Virtual Reality experience, and is promoted widely for its effectiveness, motivational impact, and significance for learning. The research described in this paper sets out to explore learner interaction in immersive Virtual Environments, focusing on the role and the effect of interactivity on learning and conceptual change. In order to examine this relationship, different environments (immersive and interactive, immersive but passive, and non-virtual) have been designed to support a set of tasks for primary school students between 8 and 12 years old. The tasks are constructive by nature, including such things as the assembly of columns from parts or the re-design of a playground, and require performing mathematical calculations. A set of qualitative observations have been made on a case-by-case basis, while the analysis is continuing to look at the various elements that form the complex relationship between interactivity and learning.

Keywords--- **Virtual Reality and Education, Interactivity, Evaluation**.

1. Introduction

Interactivity is undoubtedly one of the defining components of Virtual Reality (VR). In the context of a Virtual Environment (VE), interactivity is regarded as the process with which users can have a first-person experience, in other words, explore, act upon, control, and even modify the environment. Interactivity is also largely regarded as one of the fundamental requirements for presence within virtual reality [1] [2] , though specific studies on this are hard to find, other than studies that have been looking at the effect of body movement [3] .

In any case, the plethoric development of interactive virtual environments for rapid prototyping, industrial design, and training, to name just a few domains, and the evolution of the interfaces, emphasize the appeal of interactivity. Moreover, the proliferation of immersive systems in public spaces, such as museums and entertainment settings, and the growing sophistication of home gaming systems, advertise interactivity as a core attraction of the virtual experience. In all these contexts interactivity is being promoted widely for its effectiveness, motivational impact, and significance for learning.

Virtual environments, in general, have been valued as being extremely motivating for learners [4] , especially for those with non-traditional learning styles. Ongoing efforts at studying the other essential properties of VR, such as immersion and presence, are beginning to clarify their educational effect [5] . However, when it comes to interactivity, there is a common belief that the effectiveness of a VE that provides a high degree of interactivity is substantially more than the effectiveness of a VE where interactivity is limited. Little systematic research is available to substantiate this assumption and, to date, no clear evidence exists that interactive VR applications can bring “added value” to learning, especially for children. Furthermore, it is not certain if interactivity alone, as an essential property of the virtual reality medium, can provide a strong effect upon learning. This problem is particularly acute where deep understanding, not behavior, is of concern. Hence, a central question emerges: does interactivity enable learners to construct meaning? This research is interested in examining the dimension of interactivity in a VR experience and, in particular, its potential and limitations for learning.

Defining learning is notoriously difficult. There are a range of different perspectives on learning and a great number of theories on how learning takes place. Moreover, the notion of what constitutes learning has evolved throughout the years from a behaviorist [6] to a constructivist and social constructivist approach [7] . We are interested in examining the effect of interactivity on conceptual learning, as opposed to factual learning. Conceptual learning is identified with deeper, transferable understandings of abstract knowledge; it has do to with logical thinking, the formation of scripts, stories, cases, mental models or constructs, concepts, associations, perspectives, strategies [8] [9] .

Similarly, the different definitions of interactivity, as encountered within different contexts (socially-based contexts, distance education, museum education, etc.), illustrate the fact that interactivity remains a vaguely defined concept, despite its implicit “hands-on” or “physical” nature [10] . Nevertheless, there have been a number of attempts to provide a structure by identifying types, levels, varieties, or degrees of interactivity in an effort to better define the role of interaction and interactivity within computer-mediated learning environments. At a minimal level, most of these attempts recognize gradations of interactivity, with some actions being more or less interactive than others and the underlying assumption being that the higher the level of interactivity, the better the outcome. For this research, a working definition of interactivity which defines it as the process that actively involves the learner physically (i.e. kinesthetically) and intellectually, is adopted. This refers to more than a one-to-one call-and-response and instead implies multiple decisions and components on different levels: on one end, spatial navigation, considered to be the lowest possible form of interactive activity, manipulation of the environment or parameters of the environment as the basic middle level of interactive activity, and, on the top end, the ability to alter the system of operation itself as the highest form of interactivity. Similarly, Pares and Pares [12] have defined interactivity as *explorative*, *manipulative*, and *contributive*, categories which essentially correspond to the definition that we have adopted.

2. Previous research on VR and education

A number of educational VR research projects have been developed throughout the years, mostly in academic contexts, with a goal to apply and test the potential of virtual reality as a medium for educating students [13] . In some projects, very specific applications of VR have been developed (i.e. in chemistry, physics, etc) that examine how students react to these and if they achieve the learning goal [14] [15] . Although many interesting evaluation studies have been carried out as part of the various research efforts, these, unavoidably, produced limited or questionable results due to the fact that the complex nature of the medium was not taken into account and the evaluations isolated parameters neglecting important, in our view, contextual information. In other cases, the opposite holds, with exploratory studies that looked at general aspects rather than specific processes through which the systems cause learning [16] . Nevertheless, despite it being a very young field, virtual reality research in education has already produced a significant body of work that is also considering the longitudinal effects [17] .

However, very few studies single out and explore the influence of interactivity on conceptual learning or approach critically or even question the significance of interactivity as a facilitator of the learning process in VR. Even fewer go further to consider which forms of interactivity, if any, are effective. A study which has tackled this question in the context of geometry teaching with diagrammatic representations, focused on the comparison between different graphical representations of the concept of stereographic projection and the effect that the addition of various interactive properties might have on the learning goal [18] . The results led to the conclusion that just adding interactivity did not seem to increase the efficiency of the learning environment since the interactive 3D environment did not seem to provide the expected learning gains. However, it was noted that the study was exploratory and additional investigation was required, since learning seemed to be affected by a complex interaction of representation properties, task demands, and within-subject factors.

To summarize, VR projects developed for informal education or for other, research-based educational VR studies, have either not provided the analytical evidence to demonstrate learning as a result of interaction with the environment or, where an educational impact was perceived, there is no explanation of how and why. More importantly, the role of interactivity within learning has not been the focus of any of the evaluations carried out as such. Hence, the research question that emerges is *how* interactivity in a virtual learning environment can influence learning. To answer this question, we first need to address how this can be studied, how we can provide evidence that interactivity in a virtual environment influences learning. In the next sections, we describe the design of our studies and the virtual environments created to support the studies, in an effort to provide some answers to the above methodological question.

3. Studying interactivity in VR

3.1. Exploratory pilot studies

Since what is sought is to study learning as a result of the learner’s interaction with a virtual environment, a learning task had to be specified and an interactive virtual environment built with enough features as to invoke the aforementioned multiple levels of interactivity found in VR applications [12] . Our first idea, which was developed with consultation from supportive math and science teachers, was to create a task where the user had to build a temple by identifying and assembling its various parts. As an idea, the construction of a temple is advantageous because it encompasses an inherently activity-rich process, so it formed the basis for our exploratory studies.

A set of exploratory studies was carried out with three children between 8 and 12 years old. The children were asked to complete tasks involving the assembly of ancient columns from parts in an immersive stereoscopic VR system (a CAVE®-like display) using a 3D joystick device with buttons for interaction. The learning goal was to understand the differences between columns of different order (Doric and Ionian) and symmetry. The tasks included selection, comparison, and resizing of the column parts in order to fit them to their correct bases. Since these studies were exploratory, we followed a qualitative approach based on observation (aided by a think-aloud protocol) and informal interviews with the children. We observed the children’s activity in the VE and looked for the following different occurrences of learning for the purpose of analyzing our data:

- Conceptual change, where participants revise their conceptions or change their interpretation of something.
- Additive knowledge, where participants have added to what they have already experienced, as long as this involves some kind of reinterpretation of previous action rather than just the accumulation of information.
- Changes in behavior. Despite the constructivist focus of our study, changes in behavior were considered an important indication of learning simply because they were more likely to occur in the observational data of such a small study, than strong evidence of some internal understanding.

Similarly to [19] , our method of analysis was based on supporting or refuting emerging hypotheses; we reviewed the video of all sessions and identified various points where interesting interactions seemed to occur. We then proposed a hypothesis concerning what we saw, explaining this in terms of learning. We chose to focus on points where participants made a statement that indicated they had changed their conception or where we could conclude things from our observation of the participant's behavior in the environment. The organizational framework of Activity Theory [20] provided us with the conceptual vocabulary to help interpret these points qualitatively. Our findings indicated three kinds of instances where learning seemed to take place: learning about the system as a result of technical problems, learning caused by (unintentional) observer intervention and, to a lesser extent, learning arising from system feedback. The latter case of instances is what we are most interested in, since it involves interaction between the learner and the digital environment without human mediation. We thus focused on excerpts where such instances provoking internal contradictions leading to conceptual change seemed to occur. These caused the participants to change their behavior as well as revise their rules and conceptions, triggered by the rules set out by the system. The participants' observation of the system's rules guided them in evaluating their actions, assessing for themselves the contradiction within the system and resolving it in order to achieve the objective.

To make the analytical methodology clearer, let us look at the example of 10 year-old John. John had started constructing a column from the capital (the top part), which he placed in the air and then begun building downwards by placing each one of the drums underneath. He had managed to squeeze the last drum under the others and attempted to pick up the column base. The VE was not programmed to provide any explicit feedback; however, it was designed with certain features that provided intrinsic feedback, such as the fact that the column bases could not be moved. This was the only type of feedback that represented the system's interactive capabilities and which implicitly aided John in changing his course of action.

Observer: How do you see that this piece goes at the bottom rather than the top?

John: It's the last piece.

Observer: How do you know that it is the last piece?

John: Because I put that one [showing the bottom last column drum] and saw that there is no other one that fits below it... Anyway, you can tell it's the last piece.

John: [trying to pick up the last piece and realizing that it doesn't move] It is glued on the floor...

Observer: Why would it be glued on the floor?

John: [thinks for a moment] …Oh! So that I can put the other pieces here.

He then took apart the column he had constructed in the air and began constructing it piece by piece on top of the base by reversing the sequence in which he was placing the column drums until he reached the capital. The "Oh!" is the "Eureka" moment that both triggers his change in behavior and indicates a change in his conceptions. Furthermore, in the tasks that followed, John identified the bases immediately, having remembered from this first task that the bases do not move, and started constructing the columns from the bottom working up. For a detailed analysis of the exploratory studies using the Activity Theory framework, see [21] .

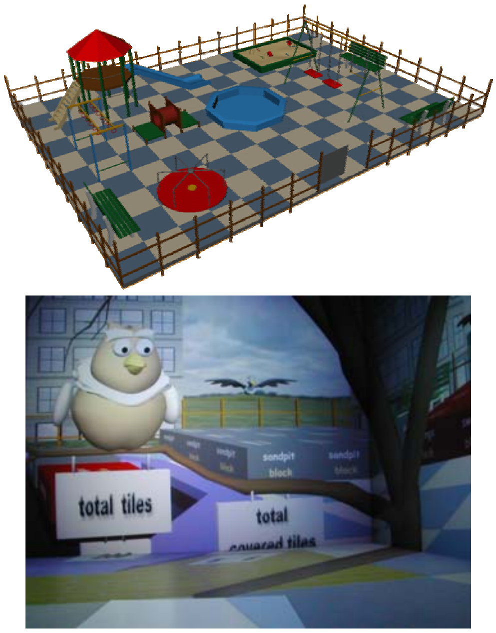


Figure 1. The layout of the virtual playground (top) and a view of the virtual environment as displayed in a cubic immersive display (bottom).

Overall, the exploratory case studies set out to explore the research question (how to provide evidence that interactivity influences learning) and helped in clarifying issues concerning the methodology for working with children for this problem, while acting as a test bed for the application of the analytical framework. They also allowed shortcomings of the task to be identified; the observed learning outcomes indicated that the learning goal of the tasks, to learn about the order and symmetry of ancient columns, was not easily quantifiable and did not provide enough opportunities for conceptual learning to occur and, consequently, to be assessed. This led to a re-design of the study, which required the design of a different virtual environment, as discussed in the following section.

3.2. The Virtual Playground

Since what is sought is evidence of conceptual change arising from a process of scaffolding and feedback generated by the system, the experiment tasks had to be re-designed in order to foster such change and minimise the other kinds of learning, such as technical learning (i.e. learning how to use a system and how to perform a task) or learning as a result of external aid from the observer.

It became apparent that the column construction activity did not provide enough opportunities for conceptual challenge and could not be easily linked to the everyday life and interests of today's children between 8 and 12 years old. Therefore, a different learning domain was chosen that would allow us to exploit the capabilities of the VR medium in visualizing abstract and difficult conceptual learning problems and providing feedback. In order to examine "interactivity", it was decided that varied levels of control over the parameters of the system should be provided through an experimental VE in which children will be asked to complete constructivist tasks that are designed as mathematical ***fraction*** problems. Fractions were chosen as the learning topic due to the difficulty that primary school students have in understanding and connecting them to real-world situations [22] . In other words, fractions lend themselves to designing learning tasks that are, at the same time, conceptually difficult, abstract enough to justify representation via a VR simulation of a real-world situation, and can allow for a kind of varied and incremental interactive treatment.

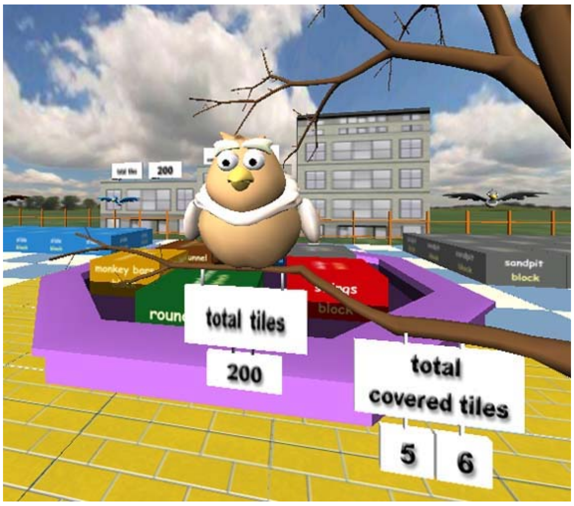


Figure 2. View of the virtual environment used for the main studies, in which children interactively design a playground based on the rules that are provided by expressive virtual characters. The owl is the main character that greets each participant and provides the general rules before the start of the design.

We decided to incorporate learning problems based on fractions into an engaging virtual reality application with a game-like scenario. Consequently, the idea of designing a playground emerged. We created both a Virtual Playground (Figure 1) and a physical model using LEGO™ bricks (Figure 4). The tasks designed for the virtual playground application involve modifying (resizing and placing) the various elements of the playground (swings, monkey bars, a slide, a roundabout, a crawl tunnel, and a sandpit). Each element covers an area which is color-coded and represented by blocks. The area representing each playground element is initially incorrect (either too big or too small) and must be redesigned, according to rules that require fractions calculations. The swings, for example, initially cover a 3 x 4 area, that is twelve blocks. The children are told to increase the area by comparing two fractions (the fractions 1/3 and 1/4) and choosing the number that represents the larger amount. In this case, the fraction 1/3 which results in 4 blocks must be chosen and the 4 blocks must be added to the swings area, by picking blocks from the central pool and placing them on the 4 tiles that need to be covered.

Figure 3. Different coloured birds represent each area that needs to be changed by the participant. When approached, the bird speaks out the rule, which requires performing fractions calculations, for its area.

The system provides both visual and audio feedback to respond to the children's activity, including feedback on the rules of the task provided by virtual characters, such as an owl (Figure 2) and six birds (Figure 3). When the correct area is formed, the user can press a button to switch to "playground mode", and immediately see the playground element appear correctly. If the area is not formed correctly, then the playground element will not appear and the user will be prompted to reconsider her actions. In addition to the switch between block mode (in which construction takes place) and playground mode, the system provides a number of other tools to facilitate the user's activity, such as the ability to switch between multiple views (ground view or top-down view).

It is important to note here that the Virtual Playground is not designed as an instructional environment following specific pedagogical models for teaching fractions, but as an evaluation environment. Hence, the characters (owl and birds) are neither avatars nor intelligent agents that respond to the user's actions and questions. They are merely "rule providers", meaning that they simply state the rules of the tasks that must be performed (in place of a written instruction sheet, for example).

4. Main Experimental Study

As already mentioned, the purpose of this research is to evaluate the value of user interaction in interactive virtual learning environments. Specifically, the goal is to evaluate if children learn better by interacting in (i.e. exploring, reacting to, and acting upon) an immersive virtual environment, or, if their interaction enhances conceptual learning of a subject matter. The Virtual Playground environment was designed as the vehicle for the evaluation of our research question. Centered on this environment, an evaluation study was planned, which started in late 2004 and continues to run. At the time of writing, approximately 30 children, between 8 and 12 years of age, have participated in two of the three conditions of the study and another 15 have been planned to take part in the third condition (Figures 5-8).

Prior to the main study, a set of pilot studies were carried out, aiming at improving the usability of the VE and allowing us to reflect on the overall process of the evaluation, so as to better prepare for the main study.

4.1. Experimental procedure

The study is being conducted with one participant at a time. The duration of the study is approximately 2 hours for each child. The nature of the study is such that the child is free to act or interact for as long as she wishes with the playground, be it the virtual or the LEGO playground.

In the first part of the study, the participant is asked to fill out a questionnaire with math questions that are based on the fractions questions found in standardized tests. A user profiling questionnaire is also given at this time. This includes questions that attempt to draw a picture of the child's familiarity with computers, frequency of computer game play, and understanding of or prior experience with virtual reality.

After the questionnaires have been collected, each child is assigned to one of three experimental conditions; either the control condition or one of two experimental conditions, in an even spread according to aptitude and gender (Table 1).

Table 1

condition	form of activity	interactivity	immersion
control	active	no	no
interactive VR	active	yes	yes
passive VR	yes*	no	yes

(*) in the case of the passive VR condition, interactivity is not directly experienced by the participant, but "through the eyes" of an invisible person who interacts with the VE while the participant watches.

If assigned to the control condition, the participant will take part in an activity using LEGO bricks. The activity will involve the design of a playground on a grid-like floor plan, similar to the top-down view of the virtual reality environment. As in the Virtual Playground, the differently coloured bricks represent the swings, slides, etc., which the participant must position according to the requirements and specifications provided. This condition does not take place in a digital environment. Thus, although, each participant is actively involved in designing the playground no interactivity (system feedback) exists.

If assigned to the interactive VR experimental group, the participant takes part in a similar activity, in a typical CAVE-like system consisting of four projection surfaces (three walls and the floor). The participant views the projected stereoscopic images by wearing a pair of active stereo glasses and can move around freely to interact with the environment by using a wireless wand which contains a joystick and buttons. The wand is used to navigate around the virtual world, and to select and manipulate virtual objects within that world. A wireless head tracker is specially adjusted on a cap that is worn by the participant, thus relaying the head position and orientation to the computer.

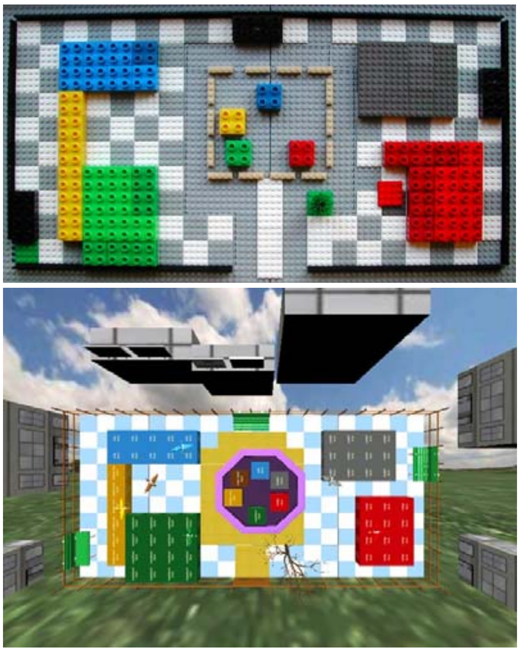


Figure 4. In the control condition, children engage in a hands-on playground design activity using actual (physical) LEGO bricks (top photograph). A similar top-down view of the playground is also provided in the virtual environment (simulator image on bottom).

In other words, the participant is immersed in the 3D re-construction of the playground in virtual reality and is asked to design the playground in this 3D space. In this case, the participant actively designs the playground, having full control over the interactive features of the system. The experience requires that the child actively explores the virtual surroundings and explains her/his actions to the observer. The task is similar to playing with a computerized construction kit or a computer game. Before starting, the task is explained to the participant who has a chance to practice moving objects around in the virtual space of a training environment.

Finally, the third condition is that of a "passive VR" experience, where the re-design of the playground is played out as in a video sequence without allowing the participant to act.

In all cases (activity with LEGO bricks, activity in the interactive VR scenario, and participation in the passive VR scenario), the participant is asked to complete a post-test with questions related to fractions, similar to the pre-test. Finally, every participant is interviewed about his/her experience by an observer who has noted the specific actions in which the participant has had problems with, and can direct the participant to reflect on these accordingly.

5. Preliminary observations from the study

Although the study has not been completed, a number of interesting observations have been made on a conceptual level and can be reported at this stage.

5.1. The problem of comparing fractions

The chief finding from the study thus far has been the confirmation of the difficulty that children have when asked to compare fractions. This was a consistent finding across most participants. Jack, for example, was able to solve almost all of the simpler exercises with relatively minimum help from the observer. When he got to the last exercise, which involved increasing the area of the swings (currently a 3 x 4 area of twelve blocks) by comparing two fractions (the fractions 1/3 and 1/4) and choosing the number that represents the larger amount, he immediately replied that he would increase the area by 1/3. However, when asked by the observer how he came up with that result, in other words, how many blocks he believed that 1/3 represented, he replied that 1/4 is four blocks and 1/3 is five blocks. This explained why he chose 1/3. The observer let him continue with his decision to add five more blocks to the swings area. When he completed the placement of the blocks (inevitably creating a non-rectangular area), he clicked on the red button to switch to "playground mode" and see if his decision was correct. When he saw that it was not, he understood that the area "did not have the right shape", but required help from the observer in order to correct it.

Mark, on the other hand, is a 9 year old boy who was very good in solving the individual fractions exercises in the pre-test. When he got to the swings, he immediately responded that 1/3 would make the swings area bigger. However, when the observer asked him how he came up with that response so quickly, he had difficulty in explaining his thought process. He eventually was able to explain that 1/3 of twelve is four, but it did not seem that he had consciously made his decision after performing the calculation; rather his decision was intuitive and seemed to be triggered by the shape of the swings area and what would look more correct.

Figure 5. A 12 year-old boy in the Virtual Playground.

It was later revealed, when talking with the parents and teachers, that both Jack and Mark had not been explicitly taught how to compare fractions in school yet, so their responses were, in some cases, random. This reinforces our observation that some decisions were made intuitively, supported also by the cues provided by the environment (the shape of each area and the surrounding space). It is possible that this intuitive action is closely linked to the form of the representation of the problem and, consequently, the value of VR over formal, abstract instruction as a way of supporting learning. Our goal in the analysis of the remaining cases will be to capture and isolate activity that seems to be a result of intuition, and carefully juxtapose it to the results of the pre- and post-tests.

Figure 6. An 11 year-old girl exploring the Virtual Playground.

Similarly, Lisa, a 10 year old girl who has been taught most fractions in school, made some decisions based on what "looked right". These decisions were evident in two cases, in which she made mistakes with her fractions. In the case of comparison between 1/3 and 1/4, she decided to increase the swings area by 1/4. When asked why, she replied: “because I counted them and they are twelve, so divided by three they will not be enough... so... [I decided that it will be] four”.

Observer: So you decided to increase by 1/4...
Lisa: yeah.

Observer: And how many blocks is that?

Lisa: uhm... [distracted by what she was doing], four.

Lisa made the common mistake (identified by [23]) of choosing 1/4 as the fraction that results in the larger number. However, she correctly added four blocks (the result of 1/3, not 1/4) to the swings area. This correct action seems, in part at least, to be attributed to her intuition rather than her calculations.

Figure 7. An 11 year-old boy placing a block in the Virtual Playground.

5.2. The power of the real world

Another interesting situation occurred with the monkey bars. In their incorrect version the monkey bars occupy an area of six blocks, placed in a long strip. The rule communicated to the participant states that the current area is too long and that it must be decreased by 1/6 of the area of the sandpit. David, an 11 year-old, immediately went to the sandpit (which occupies twelve blocks) and decided that the answer is six (another common mistake made by more than half of the participants in the study).

David: ...it’s too long [the monkey bars].

Observer: What did the bird tell you?

David: That they have to be 1/6 of the area of the sandpit...

Observer: How much is that?

David: Six.

He was certain that six was 1/6 of twelve. However, the playground confused him, since the monkey bars were already six blocks long, so if he took out six this would leave no blocks on the ground. He was stuck so the observer suggested that he try removing some blocks to see what happens. He then removed two blocks, and then another two, at which point he got it right and exclaimed that he had known all along that the correct answer was two but hadn’t thought of it from the start. When asked later why he was confused even though he knew that 1/6 of 12 is two, he responded that the correct result (two blocks) did not make sense to him, because “in real life the area for the monkey bars could not have been so short”. In this sense, it could be argued that the realistic representation of the learning task provoked “common sense”, which stood as an obstacle to conceptual change.

5.3. The choice of different views

Another interesting observation concerns the choice of views within the virtual environment (ground view or top-down view), which are provided by the design. No participant, except for one who is an avid computer game player, chose to use the top-down view of the playground (which resembles an architectural plan), even when counting the blocks in an area. Many different explanations may be given to this, either because they simply forgot about it, or because they are not used to using alternative tools that may simplify their task when a task can be performed in one way. Nevertheless, this may be interesting to follow up in the main studies, where we are considering including a reminder that will prompt the children to use the top-down view.

In summary, some generalizations have emerged from the preliminary informal analysis of the different cases, especially when examining each child’s activity and reaction to individual problems. Although we have not yet proceeded in examining where added learning value or conceptual change may have occurred, we have identified the individual sections where interesting contradictions seem to have occurred.

Figure 8. Children construct the LEGO playground as part of the control condition.

Conclusions

During the exploratory studies (constructing columns) and the pilot studies with the Virtual Playground, a number of methodological and practical issues emerged related to the challenges of designing and evaluating technology for and with children. For the main studies, the focus has been to capture behavioral and conceptual change, which can lead to indications of learning triggered by interactive activity in the virtual environment. To identify this change a number of measures have been taken. Different conditions result in a between-groups design, attempting to cover the different combinations of activity, interactivity and immersion. Then, multiple different methods of testing have been designed, ranging from the quantifiable pre- and post- questionnaires to the more qualitative observations and interviews. This is to ensure that the data collected will result in a wealth of information, which we can meaningfully combine and analyze. On the other hand, this wealth of information is a double-edged sword, as one can easily become distracted in a labyrinth of qualitative and anecdotal data of uncertain value. The use of an analytical framework such as Activity Theory, as used for the exploratory study, can help us identify the critical incidents and thus focus the analysis on these.

At the same time, the studies have so far highlighted some of the inadequacies of the methods used to collect and interpret the data. The participants, being young children, have difficulty in explaining their actions and, most of all, externalizing their thought process, while direct observation alone is unable to provide adequate insights into these internal thought processes. The think-aloud protocol that we used to obtain verbalization data can be somewhat effective, but this largely depends on the participant's learning style, capacity to verbalize, level of extroversion, or even gender [25] . Also, we hope to be as unobtrusive as possible during observation of each child's experience but it proves difficult given that the participant has to be asked questions while interacting with the virtual environment. This is a particularly common problem, especially in VR where achieving presence is paramount to the success of an experience and any direct method of eliciting information from the user during the experience can cause breaks in the user's sense of presence [26] . Nevertheless, our observations so far with the children that have interacted in the Virtual Playground indicate that not only do children feel comfortable and interact naturally with the environment after only about 2 minutes of training, but they also display a high level of presence throughout, illustrated by their movement (trying to touch the birds or sit on the swings) and comments such as "oh I keep on forgetting that I am not in a real playground!" (Figure 9).

Overall, we hope that the main studies will enlighten our understanding of children's activity and, through this, our understanding of their emerging knowledge of fractions. However, to be realistic, a short experience in a virtual environment which incorporates an alternative representation of a difficult problem is unlikely to provide us with groundbreaking evidence of conceptual learning. What we hope to achieve is to gain an insight that will help us draw some conclusions about the effect of the interactive features of an immersive environment on something so broad, deep and undefined, as learning is.

Figure 9. Children attempt to ride on the roundabout, following their successful re-design of the Virtual Playground.

In this sense, this research is expected to contribute to the understanding of the complex relationship between interactivity in advanced technological environments and learning. The experiments designed and carried out, should provide insights as to how people interact and learn in virtual environments and lead to recommendations on how interactivity should be designed in order to achieve meaningful learning experiences. The understanding of how humans interact in immersive digital environments can aid the broader community and practitioners in designing and engineering interactivity for training as well as formal or informal educational systems and contexts. This is increasingly important in a world where VR systems are becoming commonplace, especially in learning and leisure-based contexts. It is believed that VR research, an inherently interdisciplinary domain, will encompass even more and diverse research strands in the future. This work aims at advancing the study of future virtual reality systems by bringing together a number of separate yet intertwined areas that should be explored, synthesized, and translated into practice.

Acknowledgements

The authors wish to thank Dr. Martin Oliver of the London Knowledge Lab, Institute of Education, for his close involvement and valuable contributions to this research. Many thanks are due to the children that participated in the studies, and their parents and teachers. Special thanks are due to Maria Mplouna, mathematics teacher, for the time, enthusiasm, and expertise that she has offered to this project and to Alexandre Mangon-Olivier, Dimitris Christopoulos, and George Drettakis for modeling, animation and technical guidance concerning the implementation of the virtual environment.

The studies for this research have been approved by the UCL Committee on the Ethics of Non-NHS Human Research, Study No. 0171/001.

References

[1] W. IJsselsteijn. Elements of a multi-level theory of presence: Phenomenology, mental processing and neural correlates. In *Proceedings PRESENCE 2002, 5th Annual International Workshop on Presence*, Porto, Portugal, 2002, pp. 245-259.

[2] M. V. Sanchez-Vives and M. Slater. From presence to consciousness. In *Nature*, vol. 6, April 2005, pp. 8-15.

[3] M. Slater, A. Steed, J. McCarthy, and F. Maringelli. The Influence of Body Movement on Subjective Presence in Virtual Environment. *Human Factors*, vol. 40, pp. 469-477, 1998.

[4] M. Bricken. Virtual Reality Learning Environments: Potentials and Challenges. *Computer Graphics*, vol. 25, pp. 178-184, 1991.

[5] W. Winn. A conceptual basis for educational applications of virtual reality. Human Interface Technology Laboratory, University of Washington, Seattle, Washington, Technical Report TR-93-9, 1993.

[6] C. M. Reigeluth. *Instructional Design: What is it and why is it?* Instructional Design Theories and Models. Lawrence Erlbaum Associates, Hillsdale, NJ, 1983.

[7] D. A. Kolb. *Experiential Learning: Experience as the Source of Learning and Development*. Prentice-Hall, Inc., Englewood Cliffs, N.J., 1984.

[8] E. H. Wiig and K. M. Wiig. On Conceptual Learning. Knowledge Research Institute, Inc., Working Paper 1999-1 KRI WP 1999-1, 1999.

[9] J. A. Waterworth and E. L. Waterworth. Presence and Absence in Educational VR: The Role of Perceptual Seduction in Conceptual Learning. *Themes in Education*, 1(1):7–38, 2000.

[10] R. Sims. Interactivity: A Forgotten Art? *Instructional Technology Research Online*, http://intro.base.org/docs/interact/, 1997.

[11] J. Steuer. Defining virtual reality: Dimensions determining telepresence. *Journal of Communication*, 42(2):73–93, 1992.

[12] N. Pares and R. Pares. Interaction-Driven Virtual Reality Application Design. A Particular Case: El Ball del Fanalet or Lightpools. *PRESENCE: Teleoperators and Virtual Environments*, vol. 10, pp. 236-245, 2001.

[13] C. Youngblut. *Educational Uses of Virtual Reality Technology*. Institute for Defense Analyses, Alexandria, VA., Technical Report IDA Document D-2128, 1998.

[14] C. M. Byrne. Water on Tap: The Use of Virtual Reality as an Educational Tool. PhD Dissertation, University of Washington, 1996.

[15] C. Dede, M. C. Salzman, and B. R. Loftin. MaxwellWorld: Learning Complex Scientific Concepts Via Immersion in Virtual Reality. In *Proceedings Second International Conference of the Learning Sciences*, 1996, pp. 22-29.

[16] M. Roussos, A. E. Johnson, T. G. Moher, J. Leigh, C. Vasilakis, and C. Barnes. Learning and Building Together in an Immersive Virtual World. *PRESENCE: Teleoperators and Virtual Environments*, vol. 8, pp. 247-263, 1999.

[17] T. Moher, A. E. Johnson, S. Ohlsson, and M. Gillingham. Bridging Strategies for VR-Based Learning. In *Proceedings ACM SIGCHI 1999 (CHI '99: Conference on Human Factors in Computing Systems)*, 1999, pp. 536-543.

[18] N. Otero, Y. Rogers, and B. du Boulay. Is Interactivity a Good Thing? Assessing its benefits for learning. In *Proceedings 9th International Conference on HCI*, New Orleans, 2001, pp. 790-794.

[19] S. A. Barab, K. E. Hay, and M. G. Barnett, Virtual Solar System project: Building Understanding through model building. In *Proceedings Annual Meeting of the American Educational Research Association*, Montreal, Canada, 1999.

[20] B. A. Nardi. Context and Consciousness: Activity Theory and Human-Computer Interaction. MIT Press, 1996.

[21] M. Roussou. Examining Young Learners' Activity within Interactive Virtual Environments: Exploratory Studies. University College London, Department of Computer Science, London, UK, Technical Report RN/04/08, 2003.

[22] I. Harel. Children Designers. Interdisciplinary Constructions for Learning and Knowing Mathematics in a Computer-Rich School. Ablex Publishing Corporation, 1991.

[23] N. K. Mack, Learning fractions with understanding: Building on informal knowledge. *Journal for Research in Mathematics Education*, vol. 21, pp. 16-32, 1990.

[24] J. Piaget, To Understand is to Invent: The Future of Education. Grossman, 1973.

[25] P. Markopoulos and M. Bekker. On assessing usability testing methods for children. *Interacting with Computers*, vol. 15, pp. 141-150, 2003.

[26] A. Brogni, M. Slater, and A. Steed. More Breaks Less Presence. In *Proceedings PRESENCE 2002, 6th Annual International Workshop on Presence*, 2003.

Analysis of Subject Behavior in a Virtual Reality User Study

Jürgen P. Schulze[1], Andrew S. Forsberg[1], Mel Slater[2]
[1] Department of Computer Science, Brown University, USA
[2] Department of Computer Science, University College London, UK
[1] {schulze,asf}@cs.brown.edu, [2] m.slater@cs.ucl.ac.uk

Abstract

This paper presents the findings of new analysis of data collected from a prior user study comparing a CAVE-like environment and a Fish Tank VR setup. In particular, our earlier study focused on subject performance across multiple conditions and this study focuses on demographics and subject behavior during trials. We found some unexpected relationships between subject background and performance, as well as interesting details of the marking process with regard to timing and posture. We present novel ways to analyze the large amounts of data gathered in virtual reality user studies.

1 Introduction

User studies are an important part of virtual reality research because they provide insight into how well a subject pool can use systems and techniques to solve real problems. We designed and ran a user study in order to examine differences in performance of a task associated with different levels of immersive system. In this study the task was derived from a biological one, where scientists visualize a complex structure normally rendered as 3D volume data, and have to count the occurrence of specific features. In order to avoid the problem of the 3D volume data rendering at different speeds on different machines, we abstracted out of this a generic task in which subjects were asked to place the tip of 3D cone-shaped markers inside spheres that represented cells from a biological application. Subjects did this task in five variations of a four-walled CAVE-like environment and a 'Fish Tank' Virtual Reality setup (a CRT screen with active stereo and six degree of freedom head- and hand-tracking). Our initial analysis was aimed at discovering which environment subjects marked spheres fastest and most accurately in, as well as which environment subjects preferred.

In an earlier work [1], data we collected led to four significant findings: a) in the Cave the subjects preferred medium sized or large spheres over small spheres, b) when only a few of the targets had been pre-marked, larger spheres were marked faster than small spheres, c) large spheres are marked most accurately, and d) the single Cave-wall display is not comparable to the fish tank virtual reality environment when the spheres are small. Additionally, occlusion and larger field of view inhibited performance in the Cave more than at the fish tank when the task was dominated by visual search.

Below we present new findings after further analysis of the data collected previously.

2 The Experiment

We have implemented a software application named VOX (for VOlume eXplorer) that can allow users to mark 3D spheres in virtual reality. Our goal was to find out how well this task could be performed in different types of virtual environments and at different dataset scales. This system was motivated by our collaboration with developmental biologists at Brown University with whom we built VOX for immersively viewing data sets from confocal microscopes.

Counting cells or cellular components is standard practice in studying many biological processes, such as assessing the proliferation of a tissue and determining the size of cells. The density of certain components within a particular volume is often compared in control and experimental samples. Immunohistochemical techniques that use antibodies, tagged (or labeled) with a fluorochrome or other molecule that fluoresces under particular wavelengths of light, allow biologists to highlight (or recognize) the structures of interest within a tissue preparation. The digitized data is collected with a laser scanning confocal microscope that generates a static volume data set.

Counting cells requires uniquely identifying and tallying the cells within a volume; here the main challenges are isolating individual cells and (due to the large number of cells) avoiding double-counting. We selected this task because it did not require specialized knowledge of biology but still involved visualizing and interacting with the biological data.

In immersive VR, the user is fully immersed in a three-dimensional world that is completely computer-generated. The user sees a stereo image of the dataset through stereo glasses, and the user's head is tracked so that the image can be interactively rendered for the user's viewpoint. Direct 3D interaction with objects is also possible. Thus, we expected immersive VR would be a good technology for the cell-counting task because head-tracked stereo viewing would ease both seeing the dense cluster of cells and annotating already counted cells. However, there are many variations of immersive VR systems and we could not predict which combination of system attributes (e.g., field of view, scale of data) would be "best". This study was aimed at investigating these issues.

Our visualization system runs in a Fish Tank VR environment and also in our Cave, which is a CAVE-like virtual reality system. The Cave can be configured to show images only on its front wall, making it a single-wall display.

In our study users used cone-shaped icons to mark spheres while the computer tracked the total number of markers placed. In the Cave, we scaled the data set to be 0.30m^3, 0.91m^3, and 2.13m^3; at the single wall and the fish tank we used only the 0.30m^3 data set because only that size would completely fit on the smaller screen. Below we refer to the five conditions as follows: CS = "Cave small" 0.30m^3 condition, CM = "Cave medium" 0.91m^3 condition, CL = "Cave large" 2.13m^3 condition, SW = "Single wall" Cave condition, and FT = Fish Tank condition.

Because we were interested in the rate of marking, it was not necessary for subjects to mark the entire data set. However, we expected that marking rates might vary over time—in particular, as more cells were marked the task might become dominated by visual search for the next unmarked cell. Time constraints also limited the number and length of conditions we could use. Consequently, instead of having the subjects mark all 250 spheres in a data set, we divided the marking task in two halves: we started with an unmarked data set (UM) and gave the subjects two minutes to mark as many spheres as they could. Then we interrupted the experiment, removed the markers the subject had placed, and loaded spheres in the same configuration, but with 210 of them pre-marked (PM). The subject had another two minutes to mark the remaining 40 spheres.

3 Analysis

Below we present our additional analysis of the data collected during the user study. In particular, we report our findings on a demographic analysis, an analysis of tracker data, an analysis of the marking process, and further insights into why the 0.30m^3 data set was marked slower than the larger ones.

3.1 Description of variables

Response Variables (dependent variables). The dependent variable is the count of the number of correctly marked spheres.

Independent Variables. The condition was the only independent variable (CS, CM, CL, SW, FT).

Explanatory Variables. Age, type of degree, and gender were not controlled for. Additionally, in our analysis we measured the number of seconds each user spent, respectively, tumbling and moving the data set in UM and PM.

3.2 Marking accuracy and rate

The data shows several interesting results with respect to marking accuracy and rate. A marker was "accurately placed" if its tip was inside a sphere and it was considered "inaccurately placed" if its tip was outside of the sphere. We expected that if a subject placed markers slowly then they would place them accurately. However, the data shows that there was no correlation between marker accuracy and rate. Before the experiment we had instructed all subjects to focus both on marker accuracy and rate, so perhaps they actually followed this guideline, and the differences are caused by their general ability to use our system, which might be influenced by experience with other computer programs, especially games.

3.3 Relationships between marking rate and independent and explanatory variables

Our analysis also showed a relationship between the response variable (the number of markers placed) and the independent (condition) and explanatory variables (age, degree type, gender, time spent rotating data set). Since our response variable is a count, it would have a Poisson distribution. The null hypothesis is that it is unrelated to any of the independent or explanatory variables. We use log-linear regression analysis from the generalized linear model [2].

Many user study analyses of virtual reality applications have been published in the past [3-10], but few of them went beyond the direct and exclusive analysis of the response variables.

3.3.1 Unmarked condition

Condition is significant. When no spheres had been pre-marked we found the condition is significant (Chi-squared to delete from the model is 89.17 on 4 d.f.). Subjects counted a significantly higher number of spheres in CM and CL than CS, but they were not significantly different from each other. SW is not significantly different from CS. Subjects counted significantly more spheres in FT than CS, but less than CM or CL (these statements are taking into account the other terms in the regression analysis).

We think that the fact that the spheres were bigger made them easier to mark, despite the greater distance between the spheres. It is our hypothesis that at some level of scale larger than CL, the marking rate will start to drop because the user spends significantly more time navigating between spheres. More sample points in the condition domain would be needed to determine the "optimum" level of scale in the Cave.

Degree is significant. Non-Computer Science concentrators marked a significantly lower number of spheres than Computer Science concentrators. (Chi-squared for deletion from model is 29.78 on 1 d.f.).

We think that this result may be influenced by the subjects' amount of experience with similar, game-like tasks, but only a further user study that collects that information can prove this hypothesis.

Time spent rotating model. Subjects that spent more time tumbling the data set with the trackball marked significantly fewer spheres than those who spent less time tumbling the data set with the trackball. (Chi-squared for deletion from model is 7.411 on 1 d.f.).

The explanation for this result could be that those subjects that spent less time tumbling the data set had more time to place markers.

3.3.2 Pre-marked condition

Condition is significant. Subjects did not mark significantly different numbers of spheres in CS, CL, and SW. CM and FT were associated with higher sphere counts, but were not significantly different from one another. Chi-squared for removal is 14.4 on 4 d.f.

Pre-marked trackball tumbling is significant. Subjects that spent more time tumbling the data set with the trackball marked significantly fewer spheres than those who spent less time tumbling the dataset. Chi-squared for removal is 13.29 on 1 d.f. (very highly significant). Notice that this result is aligned with the corresponding result for the unmarked condition.

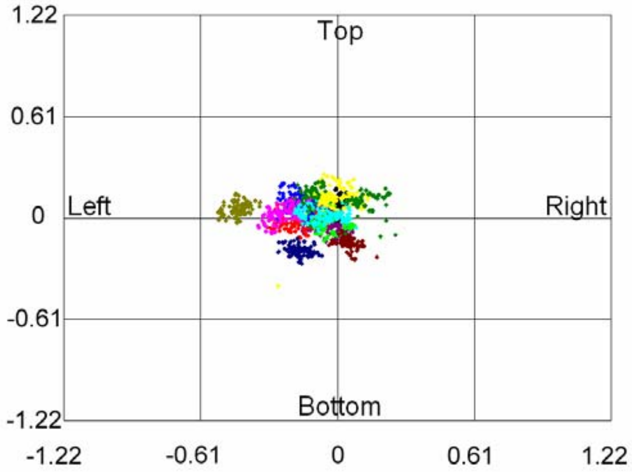


Figure 1: Wand position in CS-UM as seen from the rear of the Cave. The Cave is 2.44m wide and 2.44m high, the origin of the coordinate system is in its center. The coordinate axes indicate position in meters.

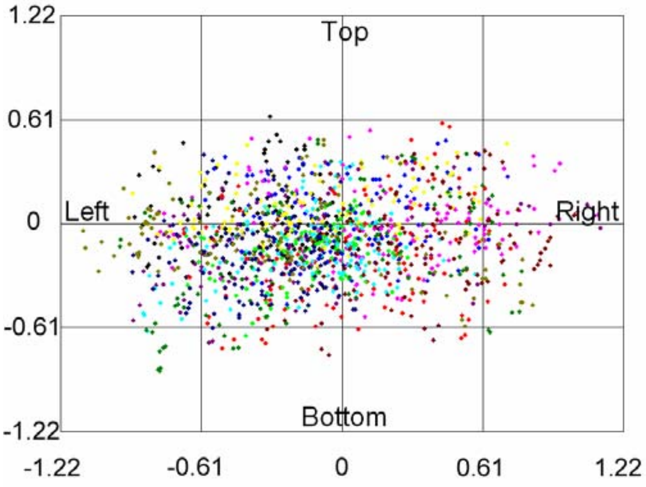


Figure 2: Wand position in CL-UM, as seen from the rear of the Cave.

3.4 Analysis of tracker data

In each experiment we stored position and orientation of head and hand once every second. Then we graphed the results with scatter plots where each dot represents a tracker sample. We use colors to distinguish subjects.

Figure 1 shows the result for condition CS-UM. The dots are projected on a plane parallel to the front wall of our four-walled Cave. Most users moved the wand relatively little.

Figure 2 shows the corresponding graph for CL. It is obvious that the users moved around a lot to reach spheres. Instead of reaching out for spheres they could have navigated the data set and bring the spheres to them, but they preferred to take advantage of the space in the Cave.

Figure 3 shows the head tracker data for CL-UM. The head moved much less than the hand. This graph is the one with the most head movement in our study, and thus it indicates that for our marking task most head tracking happened between about 1.2m and 1.8m from the ground.

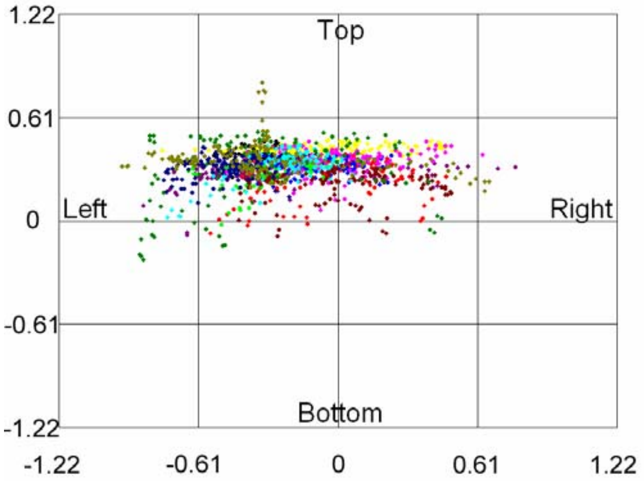


Figure 3: Head position in CL-UM, as seen from the rear of the Cave.

3.5 Marking rate development

In Section 3.3 we analyzed how the marking rate as a whole (i.e., total number of markers placed in each trial) is related to other variables of the experiment. In this section we focus on the marking process itself, looking at the times when individual markers were placed.

In our studies subjects marked spheres with two initial conditions: UM and PM. We chose these two tasks because they simulated the beginning of the marking process: when it is easy to pick out unmarked spheres and the marking rate is determined by the speed the person can move the hand to place a marker. Towards the end of a real marking process, it gets harder to find non-marked spheres and placing a marker is not as critical any more. We expected the curve to be exponential because we assumed that it is more and more difficult to place markers the more are placed.

Figure 4 shows the result for the large scale data set (CL) in the Cave. The graph consists of two parts: the left shows the marking process in the UM condition, the right shows PM. As in the scatter plots, colors distinguish users. The graph shows that all users' marking rates decreased over time. There is a significant difference in the slope of the curves in UM and PM, indicating that we are missing data between UM and PM to know how the complete graph looks, one that we would get if subjects were given enough time to mark all spheres. However, even with the existing data we can see that the curve is not strictly exponential. We hypothesize that this is because subjects typically tumble the data set and then work on a previously unmarked cluster of spheres, in which marking is fast at the beginning and slower towards the end, before the user tumbles the data set again. In the graph, when there is a gap between two marker placement events, it can be assumed that users were navigating by tumbling or moving the data set.

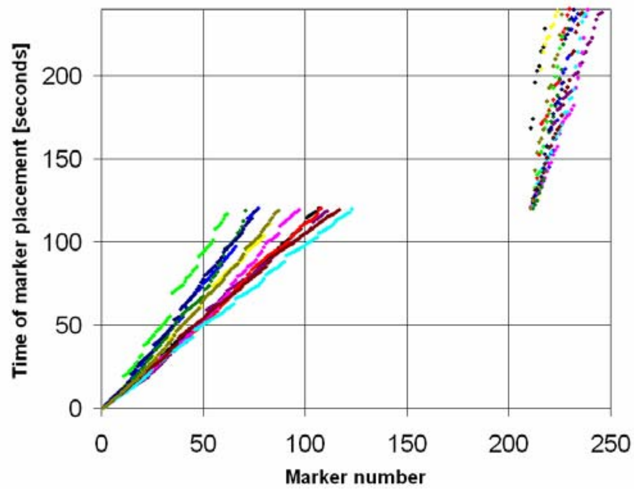


Figure 4: Marker placement times for CL.

To learn more about the rate at which subjects marked, we looked at the time that passed between any two consecutive marker placements. Figure 5 shows the corresponding graph for CL. The middle 90% of these times are between 0.7 and 2.7 seconds for UM and between 0.9 and 13.3 seconds for PM (for this analysis we ignored the slowest and fastest 5% of the times to reduce the number of outliers). The 90% boundaries for the other conditions are listed in Table 1.

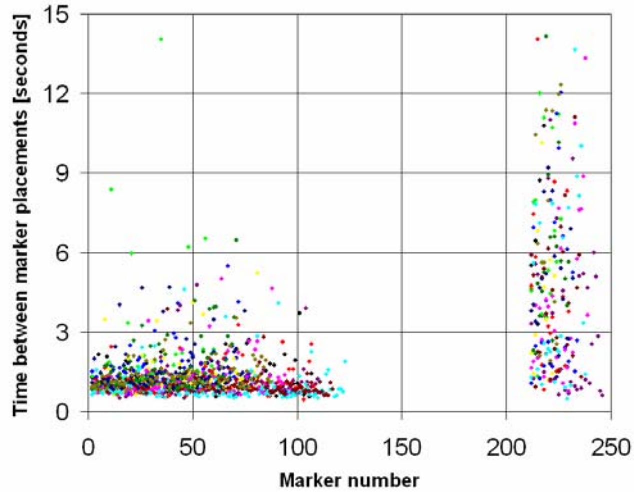


Figure 5: Marker placement rate for CL.

Table 1: 90% of the times between marker placements were between these numbers of seconds.

	CS	CM	CL	SW	FT
UM	0.8..3.1	0.7..2.3	0.7..2.7	0.8..3.0	0.7..3.4
PM	1.2..10.5	1.1..10.9	0.9..13.3	1.3..11.5	1.1..11.4

Condition CL-PM has the greatest range of times between marker placements. This may indicate that large spheres are easy to mark when they are in reach, but when none are in reach navigation takes longer than under other conditions.

3.6 Head and hand posture

A hypothesis we had on user posture was that subjects would hold the wand at the same average distance from the head, independent from the data set's level of scale, because each subject had their own "natural" distance. We were surprised that the data suggests the opposite. Table 2 lists the average distance of the wand from the midpoint of the eyes in the three Cave conditions. The table shows that with larger scale data sets the average distance from eyes to wand increased. The difference between UM and PM within each condition is not significant. We do not list SW because it is comparable with CS, and we do not list FT because it is not comparable since the user sits in front of the computer as compared to standing in the Cave.

Table 2: Average (and standard deviations) distance of wand from midpoint between eyes, given in meter.

	CS	CM	CL
UM	0.38 (0.05)	0.53 (0.10)	0.59 (0.10)
PM	0.39 (0.06)	0.53 (0.10)	0.58 (0.08)

We analyzed the same tracker data as above, wand position relative to head, for the angle at which the wand was held relative to the head direction (see Table 3). We found that there is no significant difference between conditions CS, CM, and CL. However, there is a difference between UM and PM. In PM the average angle is consistently higher than in UM, and it has a higher standard deviation. We think this result can be explained by the subjects having to look around more in PM compared to UM, because they have to search for unmarked spheres.

Table 3: Average (and standard deviations) angle between vector from eyes to wand and head direction, given in degrees.

	CS	CM	CL
UM	29.2 (7.3)	23.2 (9.6)	23.1 (10.6)
PM	32.7 (10.3)	34.4 (15.6)	38.4 (19.1)

4 Discussion

We think that one of the most important results of our user study is that the working volume plays a significant role in a highly interactive task in virtual environments. The more the working volume resembles the natural, comfortable range of human motion, the more efficient a task can be performed. In our user study those levels of scale of the data set which took advantage of the bigger working volume in the Cave, compared to the Fish Tank, resulted in higher marking rates. We think that for even larger spheres than we tested, at some point the marking rate will become smaller again, and it would be interesting to find out when this happens. A trivial "worst case" scenario to support this hypothesis is that if each sphere is about the size of the Cave, then it would take a considerably larger amount of interactions to navigate to unmarked spheres and increase occlusion, which would outweigh the benefit of visually easy-to-target spheres. Our result in Section 3.4 supports this hypothesis, because it shows that the subjects took advantage of most of the available space in the Cave to mark the spheres in condition CL. Bigger spheres than in CL would mean that more navigation will be required, which, according to our interpretation of the results in Section 3.3, would result in a smaller marking rate.

5 Conclusions

We presented novel ways to analyze the large and detailed amounts of data that can be gathered in user studies involving virtual environments. Our analysis is not meant to be complete, but we think it is likely that other interesting and meaningful relationships can be found in our data.

In the future we want to refine our data recording mechanisms to store information about the user performance in virtual environments on top of low-level tracker events, for instance the relationship between features in the data set and the viewing direction. We would also like to record user behavior in greater detail than just tracking head and hand. Furthermore, we would like to test more levels of scale to refine the number of samples we have for this condition.

References

[1] Schulze, J.P., Forsberg, A.S., Kleppe, A., Zeleznik, R.C., Laidlaw, D.H. (2005). Characterizing the Effect of Level of Immersion on a 3D Marking Task. *Proceedings of HCI International*, Las Vegas.

[2] McCullagh, P., and J. A. Nelder. 1989. Generalized linear models. London; New York: Chapman and Hall.

[3] Arns, L., Cook, D., & Cruz-Neira, C. (1999). The Benefits of Statistical Visualization in an Immersive Environment. *Proceedings of the IEEE Virtual Reality Conference*, Houston, TX, pp. 88-95.

[4] Bowman, D.A., Datey, A., Ryu, Y.S., Farooq, U., & Vasnaik, O. (2002). Empirical Comparison of Human Behavior and Performance with Different Display Devices for Virtual Environments. *Proceedings of Human Factors and Ergonomics Society Annual Meeting*, Baltimore, pp. 2134–2138.

[5] Demiralp, C., Laidlaw, D.H., Jackson, C., Keefe, D., & Zhang, S. (2003). Subjective Usefulness of CAVE and Fish Tank VR Display Systems for a Scientific Visualization Application. *IEEE Visualization '03 Poster Abstract*, Seattle, WA.

[6] Gruchalla, K. (2004). Immersive Well-Path Editing: Investigating the Added Value of Immersion. *Proceedings of the IEEE Virtual Reality Conference*, Chicago, IL, pp. 157-164.

[7] Pausch, R., Proffitt, D., & Williams, G. (1997). Quantifying Immersion in Virtual Reality. *ACM SIGGRAPH 97 Proceedings*, pp. 13–18.

[8] Robertson, G., Czerwinski, M., & van Dantzich, M. (1997). Immersion in Desktop Virtual Reality. *Proceedings of the Symposium on User Interface Software and Technology (UIST)*, Banff, Alberta, Canada, pp. 11-19.

[9] Slater, M., Usoh, M., Linakis, V., & Kooper, R. (1996). Immersion, Presence and Performance in Virtual Environments: An Experiment with Tri-Dimensional Chess. *Proceedings of ACM VRST*, pp. 163–172.

[10] Swan II, J.E., Gabbard, J.L., Hix, D., Schulman, R.S., & Kim, K.P. (2003). A Comparative Study of User Performance in a Map-Based Virtual Environment. *IEEE Virtual Reality '03 Proceedings*, Los Angeles, CA, pp. 259–266.

[11] Swartz, K., Thakkar, U., Hix, D., & Brady, R. (1999). Evaluating the Usability of Crumbs: a Case Study of VE Usability Engineering. *Proceedings of the 3rd International Immersive Projection Technologies Workshop*, May '99, Springer-Verlag, pp. 243–252.

Exploring the relationships between the usability of a medium and the sense of Spatial Presence perceived by the user

Andreas Nosper[1], Katharina-Maria Behr[1], Tilo Hartmann[1] & Peter Vorderer[2]
[1]Department of Journalism & Communication Research, Hanover University of Music and Drama, Germany
[2]Annenberg School for Communication, University of Southern California, Los Angeles, USA

Abstract

In this paper the relationships between the usability of Virtual Reality inducing technology and the sense of Spatial Presence perceived by the user are discussed from a theoretical perspective. After an explication of both constructs, joint prerequisites are identified and discussed: media factors, user factors and features of the context of use. Furthermore, some mutual effects between the two constructs are hypothesized. The paper aims to provide inspiration for further empirical work on the relationships between usability and Spatial Presence.

Keywords: Presence, Spatial Presence, Usability, Virtual Reality

1. Introduction

With the advent of enhanced interactive communication interfaces the immersive potential of electronic media has increased enormously [1]. Today, even common interactive media technologies like computer games or home cinemas have the potential to ‘catch’ the user in the virtual worlds they are providing, creating Presence experiences. In general, Presence research deals with the user’s “perceptual illusion of non-mediation” [4], which occurs when users of media systems temporarily become unaware of the fact that they are using a medium. One important facet of this phenomenon is the state of Spatial Presence that relates to the psychological sensation of the user to be located in a mediated environment, despite the fact that the environment is only an illusion affected by the medium [cf. reviews 5; 6; 7; 8; 9].

In future, the diffusion of new immersive technology might lead to a wide-spread accessibility of sophisticated Virtual Reality (VR) systems which today still are subject to laboratories of researchers. Whenever a technology is intended to be distributed to a large market, the usability of the product becomes an important issue. Accordingly, keywords like “usability engineering” [2] and “user-centered design” [3] reflect the efforts to design hardware and software in order to meet the needs of the user and enable a comfortable and effective human-technology-interaction.

In this paper, relationships between the usability of a medium and the sense of Spatial Presence perceived by its user are discussed. First, the central terms are explicated by introducing conceptualizations of usability and Spatial Presence. Building on these assumptions, the joint prerequisites of both constructs are identified and hypothetical mutual effects are discussed. The paper concludes with an outline of scenarios in which the interplay of usability and Spatial Presence becomes practically relevant.

2. What is Usability?

Early conceptualizations of usability were inspired by graphical user interfaces of personal computers which were primarily used for office applications like word processing and data-base calculations. In this context, several definitions of usability were introduced of which the one provided by the international standard ISO 9241-11 [11] has received the greatest acceptance [cf. 3]: *Usability is the “extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use”* [11 p. 2].

According to this definition, usability is a construct consisting of three dimensions: effectiveness, efficiency and satisfaction. These dimensions are framed by a specific context of use which reflects the particularities of users, tasks, equipment and the physical and social environments in which a product is used [cf. 11 p. 2].

Effectiveness is defined as the accuracy and completeness with which users achieve specified goals [11 p. 2]. Accuracy refers to the extent to which the quality of the output corresponds to specified criteria, whereas completeness stands for the proportion of the target quantity which has been achieved [11 p. 12]. Thus, the interaction with a media technology can be considered as effective when it enables the user to achieve his/her intended goals. For example, a training session in a flight simulator is effective when it enables the user to successfully replicate the actions performed in the VR in reality.

Efficiency refers to the relation between the expended resources (e.g. cognitive effort or time) and the effectiveness of the interaction [11 p. 2]. Thus, the interaction with a media technology can be considered as efficient when it enables the user to achieve his/her intended goals with reasonable effort. An example for an efficient interaction with a media offering providing VR experiences would be a VR system that enables the user to successfully navigate through the depicted surrounding without taking up all mental resources, i.e. keeping mental resources free for the completion of additional tasks.

Satisfaction stands for the user’s freedom from discomfort and positive attitudes towards the use of the product [11 p. 2]. According to this definition, the interaction with a media technology can be considered as satisfactory if the user feels free from discomfort or even enjoys the experience. For example, users of a highly immersive computer game feel satisfied when they can successfully control the avatar representing themselves in the virtual environment without distracting malfunctions of the interface.

Recently, the concept of the traditional graphical user interface usability has been extended to fit the particularities of VR technology. According to Biocca and Delaney [1], VR technology can be considered as an “extension of body and mind”: it features sophisticated input devices (e.g. data gloves, eye movement, audio input, psycho-physiological input) as well as multi-modal output devices like head-mounted displays, aural displays, haptic output, force feedback devices, whole body movement displays or even nasal displays, providing the user with continuous streams of highly realistic mediated information as well as the ability to interact with the virtual environment.

Research on the usability of VR technology has already covered various aspects like the usability of virtual environments in general [12; 13; 14], interactive systems [15], shared space virtual conferencing [16] and collaborative virtual environments [12]. The common ground of all these approaches is that the specificity of VR technology prompts specific criteria for the assessment of usability which go beyond traditional understandings.

Stanney et al. [12] for example argue that common tasks in virtual environments are wayfinding, navigation, object selection and object manipulation – tasks that are not typical for non-immersive technologies. Furthermore, the integration of visual as well as auditory and haptic system outputs confine VR systems from traditional graphical user interfaces, which are generally limited to audio-visual output.

3. What is Spatial Presence?

Simply spoken, Presence refers to a “perceptual illusion of nonmediation” [4]. Presence has been differentiated into various subtypes like “Social Presence” and “Co-Presence” [cf. 18]. A more specific concept is “Spatial Presence”, which has been described as “a sense of being there” occurring “when part or all of a person’s perception fails to accurately acknowledge the role of technology that makes it appear that s/he is in a physical location and environment different from her/his actual location and environment in the physical world” [19].

Spatial Presence is closely related to the phenomenon “situation awareness” as explicated by Endsley [20; 21]. Endsley [21] defined situation awareness as “the perception of the elements of the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future” (p.97). Situation awareness refers to the perception of situations of any kind, thus we understand Spatial Presence as a kind of “mediated situation awareness”.

Wirth et al. [22; see also 23] conceptualize Spatial Presence as a two-dimensional construct: The core dimension is the sensation of being physically located in a mediated environment, the second dimension is the perception of actions possible in this environment. The model suggested by Wirth et al. [22; 23] is designed to explain the occurrence of Spatial Presence in different media settings, and of course it covers VR systems. It explains the formation of Spatial Presence as a process affected by the interaction of personal, situational and media factors. The model is organized in two levels of processes that are involved in the emergence of Presence: The first level refers to the formation of a spatial situation model (SSM), which is a mental model of the perceived spatial (media) environment. On the model’s second level, the transition from the mere existence of an SSM towards the actual state of experiencing Spatial Presence is facilitated through the (successful) test of a perceptual hypothesis. The hypothesis states that the mediated environment is the user´s “primary ego-reference frame” (PERF). If it is confirmed, users assign their self-location and perceived possible actions to the mediated environment (i.e., they feel spatially present).

Both levels are influenced by media and user factors. Media factors attracting the user´s attention or supporting the construction of an SSM, for example spatial cues like motion parallax, contribute to the first level of the model. User factors influencing the construction of an SSM can be divided into user traits and user states and actions. Traits influencing the construction of an SSM are the domain specific interest and the user's spatial ability. States and actions are the situational motivation and the activity of spatial thinking which are corresponding to the user's traits. The allocation of attention to the presentation as a prerequisite of constructing an SSM depends on those media and user factors. Having constructed an SSM, the user may take the step to the experience of Presence, which is again influenced by media and user factors. Media factors related to the perceptual hypothesis testing like realism and meaning of the presented environment contribute to the formation of Spatial Presence. The most important user trait influencing the formation of Spatial Presence is absorption, which refers to the individual’s motivation and skill in dealing with an object in an elaborate manner. The corresponding user states and actions are involvement, defined as active and intense processing of the mediated world, and the suspension of disbelief, defined as the intentional elimination of external stimuli and internal cognitions that (might) contradict the medium-as-PERF-hypothesis.

The model of the formation of Spatial Presence experiences is visualized in figure 1.

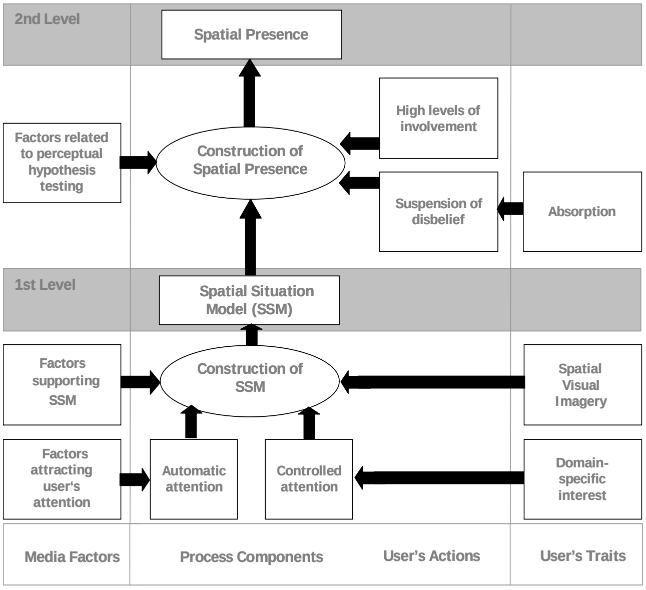


Figure 1. Visualization of the two-level-model of Spatial Presence experiences by Wirth et al. [22]

4. Joint prerequisites of usability and Spatial Presence in human-media-interactions

As explicated above, Spatial Presence is a psychological state that occurs in interaction with a medium when certain media and user factors are given. Usability, however, describes the quality of such a human-media-interaction in terms of its effectiveness, efficiency and satisfaction. Thus, both phenomena occur during human-media-interactions featured by specific conditions. These conditions can be differentiated into three categories: media factors, user factors and the context of use.

Media factors include all features of the technology (i.e. input- and output-devices and their specifications). User factors relate to the traits and states of the media user (e.g. absorption or domain-specific interest). The context of use, however, consists of the task which has to be solved in the VR (e.g. navigation through a building) and factors describing the physical and social environment in which the medium is used (e.g. external noise, room temperature or the attendance of other people). This conceptualization of human-media-interactions is visualized in figure 2.

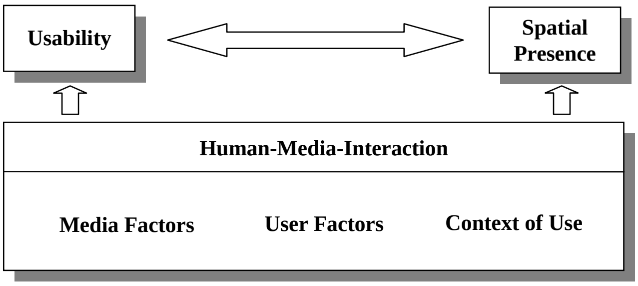


Figure 2: Conceptualization of human-media-interactions

We suggest that – when using VR technology – there are several prerequisites that foster the experience of Spatial Presence and enable high levels of usability at the same time. These joint prerequisites will be explicated in the following.

4.1 Media factors

One important factor fostering usability as well as Spatial Presence is the allocation of attention to the virtual environment. If the user is not paying attention to the mediated environment, s/he will neither be able to solve tasks in an effective, efficient and satisfactory manner nor create an SSM, which is a necessary prerequisite for the experience of Spatial Presence. Media factors can foster processes of attention allocation and thus contribute to Spatial Presence as well as to usability. Terms like “vividness” [26] refer to media factors contributing to the allocation of attention. The more sensory information a medium emits, the more likely it is that a user will persistently allocate his/her attention to the media offering. Media content is also an important factor that can affect the persistence of attention allocation: Narration, drama, and plot have been reported to increase the user’s interest in media products [35; 36].

Media factors are also important for the way the mediated information is processed. The usability of a VE as well as the occurrence of Spatial Presence depend on the user’s perception of spatial arrangements. If the user does not have the sensation of a spatial environment, s/he will not be able to form an SSM and thus will not experience Spatial Presence. On the other hand, the performance of tasks including navigation and wayfinding in a spatial environment will not be solved effectively and efficiently if the user does not experience the environment displayed by the medium as spatial. A broad range of so-called “spatial cues” fostering the perception of space within a media offering has been introduced in the literature [e.g. 28; 6]. Many spatial cues are linked to the visual modality. Static monocular cues include occlusion, height in the visual field, relative size, relative density and texture effects, aerial perspective and relative brightness, foreshortening and accommodation [28]. Binocular cues include for example stereopsis [29]. Spatial information can also be mediated by spatial audio [30], haptic cues [31] or vestibular cues [32]. The spatial information has to be displayed concise and consistent in order to enable the user to discriminate different elements and to assemble these elements into meaningful objects [“sensory integration”; 33].

In sum, media offerings that display a variety of concise spatial cues within different perceptual channels, which are linked in a consistent and plausible manner are more likely to evoke precise and coherent SSMs than those presenting only a few, diffuse or inconsistent cues. If users of VR systems are confronted with limited, constrained, or incoherent sensory inputs, they may not only have problems in forming an SSM, but may also experience some sort of confusion or impairment. The most prominent phenomenon of this kind is motion sickness, which has also been labelled cyber sickness [34], which is usually experienced as very unpleasant. VR systems fostering motion sickness have a limited usability because of poor satisfaction.

Media factors can even affect the formation of Spatial Presence directly, because they contribute to the confirmation of the medium-as-PERF-hypothesis. The mediated place should display many similarities with real places and be coherent in terms of structure, components, and dynamics, which increases the likelihood of a confirmation of the medium-as-PERF-hypothesis [22; 23]. Interactivity can also provide strong support for this hypothesis. The degree of interactivity depends on the adequacy of the given feedback and on breadth of possible interactive actions [33; 26]. Inappropriate configurations like, for example, high response latencies, will remind the user of the artificial nature of the VR system and the medium-as-PERF-hypothesis will be weakened. Interactivity does also have an impact on usability. Improper configurations will hamper effectiveness. If acting in the VR is complicated, cumbersome, and not appropriate to the task, the user will have to spend a lot of effort in order to perform the task with high quality and in an suitable time period. This task performance will not be efficient due to the additional efforts and usability can be considered as rather poor.

4.2 User factors

Several user characteristics have been identified to foster Spatial Presence experiences [cf. 37]. Some of these are likely to promote the usability of VR technology simultaneously. Wirth et al. [22; 23] propose four user traits to influence Spatial Presence experiences positively: spatial ability, absorption, domain-specific interest and the willingness to suspend disbelief.

Spatial ability, defined as "the capability to produce vivid spatial imaginations" [22 p.9], fosters the construction of an SSM and therefore supports Spatial Presence experiences. Simultaneously, a user's spatial ability can enhance the usability of a VR system, as this trait is likely to support the user completing wayfinding and navigation tasks in an effective and efficient manner.

Several studies showed that absorption fosters Spatial Presence experiences [cf. 37]. One possible explanation for this effect is that users with a high tendency to get absorbed allocate significant amounts of attention on the media stimulus. The allocation of attention is a prerequisite for high usability, too, as it enables users to perform tasks effectively.

The same argument can be applied to the role of the user's domain-specific interest in the mediated environment: Such an interest is likely to foster Spatial Presence experiences, as it supports the allocation of attention to the VR content. Simultaneously, it can be assumed that users who are interested in the mediated environment perform tasks more effectively and with greater satisfaction.

Suspension of Disbelief, defined as the intentional elimination of external stimuli and internal cognitions that (might) contradict the medium-as-PERF-hypothesis [22], fosters Spatial Presence experiences. It can also enhance usability, as users with a high willingness to suspend disbelief may be able to compensate disturbances of the human-media-interaction (e.g. response latencies of input devices) cognitively, fostering the efficiency and satisfaction with which tasks are solved.

4.3 Features of the context of use

In addition to media and user factors, human-media interactions are characterized by the context of use, which includes the task which has to be solved as well as the physical and social environment in which the medium is used.

We assume that two specific task characteristics can be joint prerequisites for Spatial Presence experiences and usability: The relevance of spatial cues for task completion as well as the relevance of the task for the user.

In order to be solved effectively, many tasks require information about the spatial structure of the virtual environment. This applies for all tasks that demand orientation, e.g. navigation and wayfinding but also for complex tasks like the accomplishment of tele-operated surgeries. The presentation of spatial cues fosters Spatial Presence experiences, as they help the user to create an SSM. Simultaneously, the presentation of spatial cues is a necessary condition for the effective completion of tasks demanding orientation in the virtual environment.

A high relevance of the task for the user is another joint prerequisite. Some tasks have a high relevance for the user, e.g. the realization of tele-operated surgeries, where the success of the task completion judges over the life or death of the patient. In cases of high relevance, it can be assumed that media users allocate striking amounts of their cognitive resources on the task. As explicated above, the allocation of attention is a major precondition for the experience of Spatial Presence. Simultaneously, it can be hypothesized that the allocation of attention on the task will enhance the usability of the medium, as focused users are likely to solve tasks more efficiently.

As far as the physical and social environment in which the medium is used is concerned, the absence of disturbing stimuli can be considered as a joint prerequisite of Spatial Presence experiences and usability. Disturbing stimuli, defined as stimuli not generated by the VR technology, can disturb or even impede Spatial Presence experiences, as they distract the user's allocation of attention on the mediated environment. Coevally, a lack of attention lowers the efficiency with which tasks are solved and therefore confines the usability of VR technologies.

5. How usability can have a particular effect on Spatial Presence

In addition to the identification of joint prerequisites of the two constructs, we will also 'dare' a brief closer look on possible effects of the usability of VR systems on the experience of Spatial Presence. We assume that all three core aspects of usability (i.e. a high satisfaction, a high efficiency, as well as a high effectiveness) can positively influence the psychological processes involved in the formation of Spatial Presence.

First, a high satisfaction should motivate the user to extend the exposure to a stimulus or environment [cf. 38]. Thus, the attention allocation onto a depicted spatial scenery could be prolonged, which in turn should increase the likelihood of Spatial Presence experiences (see section 3).

Second, a low efficiency implies that many resources (i.e. cognitive workload, time, etc.) need to be invested by the user in order to achieve a given task. It might be argued, in turn, that in special cases a lack of cognitive resources can impede the formation of Spatial Presence. For example, this could be true if the user encounters a poor media offer with insufficient or ill-defined spatial cues and thus has to invest a relatively high amount of cognitive resources (in terms of visual spatial imagery or suspension of disbelief; see section 3) if he wants to feel spatially present. However, with much resources being bound to task achievement, the user might abandon processes to advance the 'spatial illusion' in order to avoid a cognitive 'overload'. A low efficiency, then, would hinder or impede Spatial Presence experiences. This means, in turn, that a high efficiency enables the user to invest additional cognitive workload to the spatial illusion, if necessary.

Third, it can be argued that effectively solved tasks lead to mastery experiences or feelings of self-efficacy [39].

Given that a typical task conducted in a VR (e.g. navigating a robot through a labyrinth to an exit) builds on a chain of subtasks (e.g. moving the robot, turning the robot, reaching a room with the robot, etc…), it might be argued that effectively solved tasks lead to persistent feelings of mastery, which closely resemble a specific meta-experience that has been described as flow [40]. One typical aspect of flow is the absorption of the user in the experience, causing that he or she temporarily forgets about the outer world, space and time. Given the general similarity to Presence experiences and with specific regard to the role of absorption in the formation of Spatial Presence (see section 3), it seems reasonable to think of flow experiences as promoters of Spatial Presence [41]. In sum it can be argued, then, that a chain of effectively solved tasks result in flow experiences, which in turn could contribute to the sensation of Spatial Presence [for empirical evidences for a positive contribution of task performance on Presence cf. 42 - 43].

6. How Spatial Presence can have a particular effect on usability

We assume that one can also distinguish a particular effect Spatial Presence might exert on usability. When experiencing Spatial Presence, objects and other entities are perceived in direct relationship to the own egocentric position [44]. Thus, users can directly assess distances, sizes, and ‘action possibilities’ in the virtual space. In contrast, without feeling physically present, users have no ‘direct’ reference point in the spatial scenery. Thus, if they want to construe the environment, they need to ‘reconstruct’ the spatial arrangement in their mind by implementing a reference point on a virtual basis. Thus, cognitive resources are bound if users do not feel spatially present, but set free, if they feel located in the environment.

Building on this assumption, it can be argued that usability might benefit from states of Spatial Presence, as users can allocate more cognitive resources for an effective and efficient task performance.

7. Conclusions

The theoretical analysis of the concepts usability and Spatial Presence showed that there are several joint prerequisites. In addition, it has been argued that a high usability should positively affect the experience of Spatial Presence and, in turn, that the experience of Spatial Presence should enhance the usability of VR technology. These assumptions, however, have been derived out of a theoretical conceptualization and have to be tested empirically in the future.

From a practical point of view, taking the joint prerequisites into account can be useful for researchers, designers and engineers working on VR-technology-interfaces that need to provide Spatial Presence experiences and enable high degrees of usability at the same time like vehicle simulations (e.g. aircraft and ship simulators), physical simulations (e.g. the simulation of magnetic fields for students of physics) or medical applications (e.g. tele-operated surgeries and anatomy instruction) [cf. 45].

From a scientific point of view, this means that research on Spatial Presence should include usability variables and VR-usability research should include the assessment of Spatial Presence – the interrelation of both constructs calls for interdisciplinary research.

Acknowledgements

This work has been supported by the European Community, Information, Society and Technology Program, within the research project “Presence: Measurement, Effects, Conditions (MEC)”, IST-2001-37661 (http://www.ijk.hmt-hannover.de/presence). We thankfully acknowledge the European Community´s support.

References

[1] Biocca, F. & Delaney, B. (1995). Immersive Virtual Reality Technology. In F. Biocca & M. R. Levy, *Communication in the Age of Virtual Reality*. (pp. 57-124). New Jersey: Lawrence Earlbaum Associates.

[2] Nielsen, J. (1993). *Usability Engineering*. Boston: Academic Press.

[3] Bevan, N. (2001). International Standards for HCI and usability. *Int. J. Human-Computer-Studies*, 55, 533-552.

[4] Lombard, M. & Ditton, T. (1997). At the Heart of It All: The Concept of Presence. *Journal of Computer-Mediated Communication 3*(2). Available Online: http://www.ascusc.org/jcmc/vol3/issue2/lombard.html [04.01.2004].

[5] Draper, J. V., Kaber, D. B. & Usher, J. M. (1998). Telepresence. *Human Factors, 40*(3), 354-375.

[6] Barfield, W., Hendrix, C., Bjorneseth, O., Kaczmarek, K. A., & Lotens, W. (1995). Comparison of Human Sensory Capabilities with Technical Specifications of Virtual Environment Equipment. *Presence: Teleoperators and Virtual Environments, 4*(4), 329-356.

[7] Riva, G., Davide, F., & Ijsselsteijn, W.A. (2003). (Eds.), *Being there: concepts, effects and measurement of user presence in synthetic environments*. Amsterdam: Ios Press.

[8] Lee, K. M. (2004). Presence, explicated. *Communication Theory, 14*(1), 27-50.

[9] Reeves, B.R. (1991). *“Being There:” Television as symbolic versus natural experience.* Unpublished manuscript. Stanford University, Institute for Communication Research, Stanford, CA.

[10] Minski, M. (1980). Telepresence. *Omni, 2,* 45-51.

[11] ISO 9241-11. (1998). *Ergonomic requirements for office work with visual display terminals (VDT)s – Part 11 guidance on usability.*

[12] Stanney, K.M., Mollaghasemi, M., Reeves, L. Breaux, R. & Graeber, D. A. (2003). Usability engineering of virtual environments (VEs): identifying multiple criteria that drive effective VE system design. *Int. J. Human-Computer-Studies*, 58, 447-481.

[13] Hix, D. & Gabbard, J. L. (2002). Usability Engineering of Virtual Environments. In K. M. Stanney (Ed.), *Handbook of Virtual Environments: Design, Implementation and Applications.* Mahwah, NJ: Lawrence Erlbaum Associates. pp. 681-699.: vehand.engr.ucf.edu/handbook/Chapters/chapter39.PDF [04.01.2004]

[14] Gamberini, L. & Spagnolli, A. (2003). On the relationship between Presence and Usability: A Situated, Action-Based Approach to Virtual Environments. In G. Riva, F. Davide &

W. A. Ijsselsteijn (Eds.*), Being There: Concepts, effects and measurement of user presence in synthetic environments.* Amsterdam: Ios Press. pp. 97-107.

[15] Kirste, T., Diener, H., Encarnação, L. M., Hein, O., Stork, A. & Bimber-Fraunhofer, O. (2002). Die Zukunft interaktiver Systeme: Herausforderungen an die Usability-Forschung. [The future of interactive systems: Challenges for usability research] *it + ti - Informationstechnik und Technische Informatik 44*(1), 40-48.

[16] Anderson, J., Nahella, A., Douther, C., Mervyn, A. J. (2001). Presence and Usability in Shared Space Virtual Conferencing: A Participatory Design Study. *CyberPsychology & Behavior, 4*(2), 287-305

[17] Tromp, J. G., Steed, A. & Wilson, J. R. (2003). Systematic Usability Evaluation and Design Issues for Collaborative Virtual Environments. *Presence: Teleoperators and Virtual Environments, 12*(3), 241-267.

[18] IJsselsteijn, W. A., de Ridder, H., Freeman, J., & Avons, S.E. (2000). Presence: Concept, determinants and measurement. *Proceedings of the SPIE, Human Vision and Electronic Imaging V, 3959-76.* Presented at Photonics West - Human Vision and Electronic Imaging V 23-28 January 2000, San Jose, CA.

[19] ISPR (2001). *What is Presence?* [Online]. Available: http://www.temple.edu/mmc/explicat.htm [9.10.2002].

[20] Endsley, M.R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, 37, 32-64.

[21] Endsley, M.R. (1988). Design and evaluation for situation awareness enhancement. *Proceedings of the Human Factors Society 32nd Annual Meeting* (pp. 97-101). Santa Monica, CA: Human Factors Society.

[22] Wirth, W., Hartmann, T., Böcking, S., Vorderer, P., Klimmt, C., Schramm, H., Saari, T., Laarni, J., Ravaja, N., Gouveia, F. R., Biocca, F., Sacau, A., Jäncke, L., Baumgartner, T., & Jäncke, P. (submitted for publication). *A Process Model of the Formation of Spatial Presence Experiences.*

[23] Hartmann, T., Böcking, S., Schramm, H., Wirth, W., Klimmt, C., & Vorderer, P. (2005). Räumliche Präsenz als Rezeptionsmodalität: Ein theoretisches Modell zur Entstehung von Präsenzerleben [Spatial presence as modus of reception: A theoretical model]. In V. Gehrau, H. Bilandzic & J. Woelke (Eds.), *Rezeptionsstrategien und Rezeptionsmodalitäten* [Strategies and modalities of reception] (pp. 21-37). München: R. Fischer.

[24] Cutting, J. E. (1997). How the eye measures reality and virtual reality. *Behavior Research Methods, Instruments & Computers, 29*(1), 27-36.

[25] Faubert, H. (2001). Motion parallax, stereoscopy, and the perception of depth: Practical and theoretical issues. In J. Bahram, M. Storrs, F. Okano (Eds.*), SPIE Proceedings Vol. CR76.*

[26] Steuer, J. (1992). Defining Virtual Reality: Dimension determining Telepresence. *Journal of Communication, 42*(4), 73-93.

[27] Anderson, D. R., & Burns, J. (1991). Paying attention to television. In D. Zillmann & J. Bryant (Eds.), *Responding to the screen: Perception and reaction processes* (pp. 3-26). Hillsdale, NJ: Erlbaum.

[28] Surdick, T., Davis, E. T., King, R. A., & Hodges, L. F. (1997). The perception of distance in simulated visual displays: A comparison of the effectiveness and accuracy of multiple depth cues across viewing distances. *Presence: Teleoperators and Virtual Environments, 6*(5), 513-531.

[29] Murch, G. M. & Woodworth, G. L. (1977). *Wahrnehmung* [Perception]. Stuttgart: Kohlhammer.

[30] Hendrix, C., & Barfield, W. (1996). The sense of Presence within auditory virtual environments. *Presence: Teleoperators and Virtual Environments, 5*(3), 290-301.

[31] Sallnäs, E.-L. (1999). Presence in multimodal interfaces. Presentation at the 2nd International Workshop on Presence, 6-7 April 1999, Essex, UK.

[32] Riecke, B.E., von der Heyde, M., & Bülthoff, H.H. (2001). How do we know where we are? Contribution and interaction of visual and vestibular cues for spatial updating in real and virtual environments. In H. H. Bülthoff, K. R. Gegenfurtner, H. A. Mallot & R. Ulrich (Eds.) TWK 2001. *Beiträge zur 4. Tübinger Wahrnehmungskonferenz* [Proceedings of the 4th Tubingen Conference on Perception] (p. 146). Kirchentellinsfurt: Knirsch-Verlag.

[33] Witmer, B. G., & Singer, M. J. (1998). Measuring Presence in virtual environments: A Presence questionnaire. *Presence: Teleoperators and Virtual Environments 7*(3), 225-240.

[34] McCauley, M. E., & Sharkey, T. J. (1992). Cybersickness: Perception of Self-Motion in Virtual Environments. *Presence: Teleoperators and Virtual Environments, 1*(3), 311-318.

[35] Kelso, M. T., Weyhrauch, P., & Bates, J. (1993). Dramatic Presence. *Presence: Teleoperators and Virtual Environments, 2*(1), 1-15.

[36] Klimmt, C. & Vorderer, P. (2003). Media Psychology „is not yet there“: Introducing theories on media entertainment to the Presence debate. *Presence: Teleoperators and Virtual Environments, 12*(4), 346-359.

[37] Sacau, A., Laarni, J. & Hartmann, T. (2005). *Theory report on user characteristics related to Spatial Presence.* Unpublished report to the European Community IST Programme "Presence Research Activities".

[38] Vorderer, P., Ritterfeld, U. & Klimmt, C. (2001). Spaß am Hören – Hörspielkassetten als sprachförderliche Unterhaltungsangebote für Vorschulkinder [The fun of listening]. *Medien- und Kommunikationswissenschaft, 49*(4), 462-479.

[39] Bandura, A. (1997). *Self-Efficacy: The exercise of control.* New York: Freeman.

[40] Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York: Harper Row.

[41] Takalato, J. (2002). Presence and flow in virtual environments: An explorative study. [Online available]. http://eve.hut.fi/Publications/ttalo_thesiswoa.pdf (March 10 2005).

[42] Sas, C., O'Hare, G.M.P., & Reilly, R. (2004). Presence and task performance: an approach in the light of cognitive style. *Cognition, Technology & Work, 6*, 53-56.

[43] Youngblut, C., & Huie, O. (2003). The relationship between presence and performance in virtual environments: results of a VERTS study. *Proceedings of the IEEE Virtual Reality* (pp.277-278).

[44] Hofman, J., Jäger, T.J., Deffke, T. & Bubb, H. (2001). Effects of Presence on Spatial Perception in Virtual Environments. Electronic Proceedings of the 4th International Workshop on Presence, Temple University, Philadelphia, 21-23 May, 2001.

[45] Ellis, S. R., Begault, D. R. & Wenzel, E. M. (1997). Virtual Environments as Human-Computer-Interfaces. In Helander, M., Landauer, T. K. & Prabhu, P., *Handbook of Human-Computer-Interaction. 2nd edition*. Amsterdam: Elsevier.

Session 9

23nd September, 2005

Multimodal Presence

12.00-12.30 *Multi-Modal Stimulation, Response Time and Presence*
David Hecht[1], Miriam Reiner[1] and Gad Halevy[2]
[1] Technion - Israel Institute of Technology Department of Education in Technology and Science Haifa, 32000, Israel
[2] The University of Haifa Haifa, 31905, Israel

12.30-13.00 *A comparison of the effect that the visual and haptic problems associated with touching a Projection Augmented model have on object-presence*
Emily Bennett and Brett Stevens
Department of Information Systems and Computer Applications, University of Portsmouth, UK

13.00-13.15 *Difficulties Using Passive Haptic Augmentation in the Interaction within a Virtual Environment.*
R. Viciana-Abad, A. Reyes-Lecuona and F.J. Cañadas-Quesada
Department of Electronic Technology, University of Málaga, Spain

13.15-13.30 *The Value of Reaction-Time Measures in Presence Research: Empirical Findings and Future Perspectives*
Christoph Klimmt[1], Tilo Hartmann[1], Andre Gysbers[1] and Peter Vorderer[2]
[1] Department of Journalism & Communication Research, Hanover University of Music and Drama, Germany
[2] Annenberg School for Communication, University of Southern California, Los Angeles, USA

Multi-Modal Stimulation, Response Time, and Presence

David Hecht[1,2], Miriam Reiner[1] & Gad Halevy[1]
[1] Technion – Israel Institute of Technology
Department of Education in Technology and Science
Haifa, 32000, Israel
[2] The University of Haifa
Haifa, 31905, Israel

davidh@techunix.technion.ac.il, miriamr@techunix.technion.ac.il, gadh@techunix.technion.ac.il

Abstract

Multi-modal virtual environments succeed better, than single-channel technologies, in creating the 'sense of presence'. We hypothesize that the underlying cognitive mechanism is related to a faster mental processing of multi-modal events. Comparing reaction time in uni-modal, bi-modal and tri-modal events, we show that the processing speed is: uni-modal > bi-modal > tri-modal. Given this advantage, users of multi-modal VEs start their cognitive process faster, thus, in a similar exposure time they can pay attention to much more informative cues. This results in a more rich, complete and coherent experience and a greater sense of presence.

***Keywords**: Multi-Modal, Visual, Audio, Haptic, Processing-Speed, Presence.*

1. Introduction

Multi-modal virtual environment systems able to combine efficiently sensory information from two or three channels (vision, audio, haptic) have an advantage in generating the 'sense of presence'. This "multi-sensory" experience differentiates them from older technologies, communicating only via a single sensory channel, which can generate only a limited degree of immersion and presence. Therefore, it is assumed that the more *multi*-modal a virtual environment is designed, the greater the sense of presence it generates [1] [4] [6] [7] [8] [11]. However, the *underlying cognitive mechanisms* in which multi-modal environments succeed to create an enhanced sense of presence are still elusive and unknown. In the following sections we present the ideas introduced by researchers and then suggest another possible mechanism.

1) Environmental richness results in a complete and coherent experience

A rather intuitive idea suggests that a *single channel* media is relatively *sensory-poor* and conveys *limited and insufficient* information to the senses, thus it engenders only a lower sense of presence. Conversely, *multi*-modal environments provide a greater extent of sensory information to the observer. This *sensual richness* translates into a *more complete and coherent experience*. And therefore, the sense of "being there", in the virtual realm, is felt stronger [4] [8] [11].

2) Multi-modal VEs mimic "reality" better

Another way in which *multi*-modal environments succeed in creating a stronger sense of presence is by better mimicking "reality" [6]. An elaboration of this idea argues as follows: many of our *natural daily experiences* in the real world are fundamentally *multi*-modal by their nature, for instance, reaching to grasp an object or even simple posture and movement control are a co-production of visual, haptic and vestibular systems [5]. Communicating with another person through speech is a fine combination of producing and receiving audio and visual cues – sound, lip movements and gestures [10]. Our gastronomic pleasures result from a fine integration of taste, smell and vision [2] [3].

Therefore, *multi*-modal VEs have a clear advantage, in mimicking a multi-modal phenomenon, since they stimulate not only the user's *auditory* and *visual* sensory systems, but they do it with a *realistic 3D depth* perception. In addition, as a result of capturing the entire perceptual field (via head mounted display or 360^0 presentation) they stimulate also the *proprioceptive* and *vestibular* systems, as evidenced by the simulators sickness phenomenon and user's 'natural body movements' in virtual environments. The experience is especially felt as "real" if it includes also haptic sensations.

3) "Filling in" of missing information

Biocca et. al [1] proposed another mechanism which may help *multi*-modal virtual environments gain an edge in creating the sense of presence. They argue that the sensation of presence in virtual environments is related to the mind’s attempt to integration. Since synthetic virtual environments provide fewer sensory cues than most physical environments in which we act, the user needs to interpolate sensory stimuli to create a functional mental model and use these cues to walk towards, reach out, and manipulate objects in the environment. During the process of integrating and augmenting impoverished sensory cues, information from one sensory channel may be used to augment and help ambiguous information from *another* sensory channel.

Thus, the process of *inter-modal integration* enables an inter-sensory *"filling in" of missing information*. This is a rather active and creative process, depending on the user abilities, and this *active* cognitive process results in an enhanced immersion into the virtual scene and a greater sense of presence.

4) Faster processing enables deeper and richer experience

While the above explanations - coherent experience, mimicking reality and filling in missing information - focus mainly on *higher cognitive functions*, occurring at the *end* of the *cognitive processing stream*, we suggest another possible mechanism which occurs *earlier* in *beginning* of the processing stream, at the *initial perception level*, which gives an advantage to multi-modal environments over single channel systems in creating the sense of presence.

Using a *simple reaction time* paradigm we compared the brain processing speed of *uni*-modal events (audio, visual or haptic) with the processing speed of *bi*-modal combinations of these signals and a *tri*-modal combination of these signals. Our hypothesis suggested an advantage, in processing speed, for bi-modal signals over uni-modal signals. Furthermore, we hypothesized that *tri*-modal signals will be processed *even faster than all bi-modal combinations*.

The rational for this study is that a processing speed advantage in multi-modal events may indicate a *greater focus of attention*, which may affect the entire event to be experienced as richer, more complete and coherent. In addition, a faster processing speed in the initial perceptual stage (at the first 300-400 msc.) allow users *more time in the consequent cognitive stages* enabling them to 'fill in' missing information and thus create a richer experience.

2. Experimental design

Materials

We used a touch-enabled computer interface which can generate for the users visual, auditory and haptic sensation. The haptic device (shown in figure 1) is based on a force-feedback mechanism which can generate haptic sensations felt by the user as a resisting force. Full technical descriptions of this system are available at: http://www.reachin.se and http://www.sensable.com.

Figure 1: While users held the pen-like stylus (on the right) performing writing-like movements, the attached force-feedback mechanism generated a resisting force – haptic stimulation. Users responded by pressing a button on the stylus.

Participants

Sixteen students, 11 males and 5 females, (mean age - 25.5 years) participated in this study. They were recruited at the Technion, thus having had a minimum of 12 years education. All had normal hearing and normal or corrected to normal vision. They were paid for their participation but were *not unaware* of the purpose of the experiment, except that it has to do with eye-hand coordination. The experiment was carried out under the guidelines of the ethical committee and with its approval.

Stimuli

Seating in front of the computer system, participants were presented visually with 2 parallel green lines. Their task was to hold the stylus in their hand and move it by crossing these lines as if they are writing (see figure 2). On each trial the computer generated a sensory stimulation, either *uni-modal* (visual (V), auditory (A) or haptic (H)), *bi-modal* - a combination of the visual and auditory (VA), the haptic and visual (HV) or the haptic and auditory (HA) stimulations, or *tri-modal* – a combination of the haptic, visual and auditory (HVA) stimulations. The visual stimulus consists of the 2 lines changing their color from *green* to *red*. The auditory stimulus was a compound sound pattern (8 KHz, 560 msc.) emitted from 2 loudspeakers located at both sides of the subject. The haptic stimulus was a *resisting force* (4 Newton) delivered through the stylus.

Procedure

Participants were instructed to react, by pressing a button on the stylus, as soon as they detect *either* one of the three stimuli or *any* of their combinations. Reaction time was measured, from the beginning of the stimulation until the subject's reaction, and recorded by the computer. Participants used *the same hand* to move the stylus and to react by pressing the button (with the index finger). The other hand rested freely on the table.

In order to prevent participants from knowing and/or expecting the exact timing of the stimulation, they were delivered *randomly* in the following manner. The computer counted each crossing (of both, upper and lower, lines) made by the subject and generated the stimulation, randomly, between the 5th and the 13th crossings. (For example, in the 1st trial, the stimulation was delivered immediately after the 5th crossing, in the 2nd trial, the stimulation was delivered only after the 12th crossing. and in the 3rd trial, the stimulation was delivered after the 10th crossing etc.). In this way, although the participants' movements triggered the stimulations, they were not aware of this arrangement so they could not predict the timing of the next stimulation, thus, they continued to cross the lines until they were actually stimulated.

Before the beginning of the experiment, each subject was trained briefly how to perform his task. The experiment consisted of 6 blocks of trials, 3 performed with the dominant hand and 3 with the other hand. Each of these 6 blocks consisted of 105 single trials, in which each of the 7 conditions (V, A, H, VA, HV, HA, HVA) appeared 15 times. All 7 conditions were randomly intermixed in order to prevent participants from expecting a stimulus in a *specific modality* [9] so in each block, every consecutive 7 trials contained one trial of every condition, but their *internal* arrangement - within the 7 - differed randomly (For instance, the initial seven were: A, HV, H, VA, HVA, V, HA, the next seven were: H, V, VA, HA, A, HVA, HV etc.). Total number of trials for each subject was 630 (105 (trials) x 3 (blocks) x 2 (both hands).

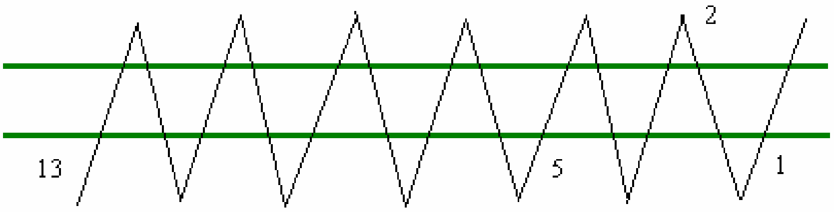


Figure 2: Participants performed writing-like movements with the stylus crossing the parallel green lines. Between the 5th and the 13th crossings, the computer generated, randomly, a sensory stimulation, either *uni*-modal, *bi*-modal or *tri*-modal.

3. Results

A repeated measures ANOVA (GLM) indicated a significant main effect for condition, in both, the dominant

[F(6,10) = 32.71 , p < 0.000] and non-dominant hands [F(6,10) = 29.54 , p < 0.000].

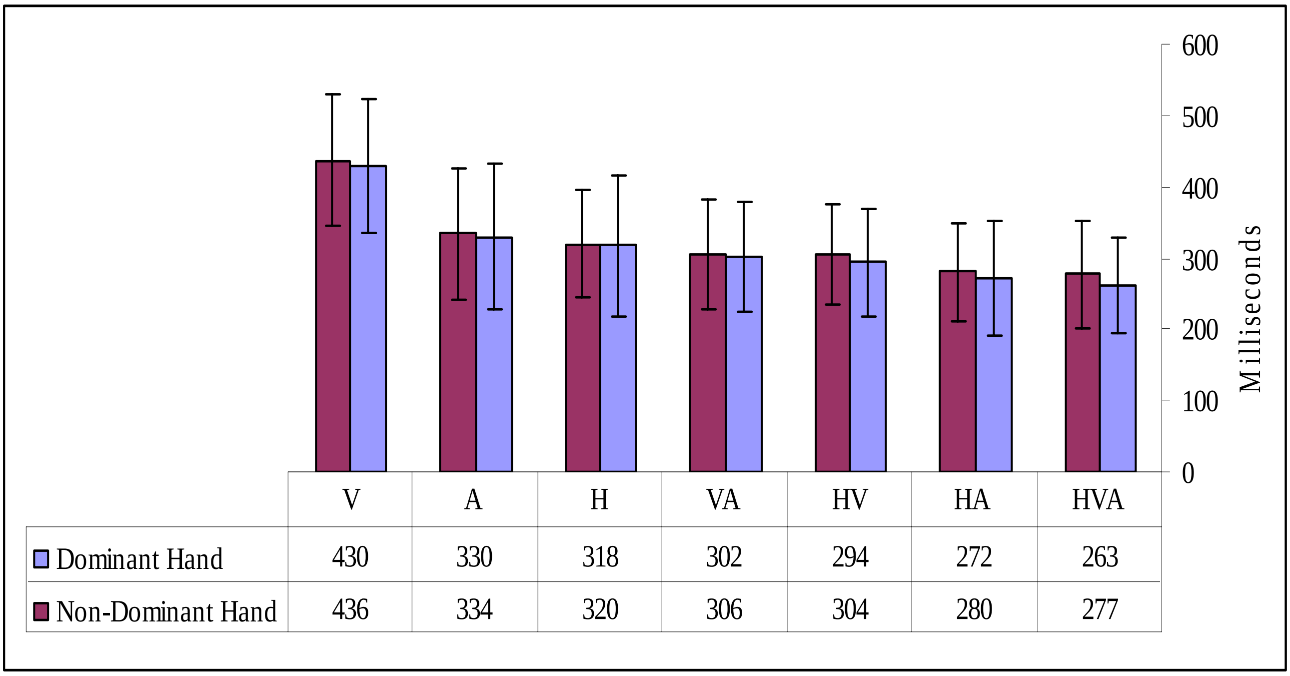
Dominant hand

Mean RT in the *uni*-modal conditions were the longest. 430 ms. for the visual stimulus, 330 ms. for the auditory stimulus and 318 ms. for the haptic stimulus. All three *bi*-modal conditions were *shorter than any uni-modal condition*, 302 ms. for the audio-visual combination, 294 ms. for the haptic-visual combination and 272 ms. for the haptic-audio combination. RT in the *tri*-modal combination was the *shortest* – 263 ms. See figure 3 for a summary of the results.

Paired comparisons analysis revealed that: a) When participants received a bi-modal combination of auditory and visual cues simultaneously, their RT [mean = 302, SD = 78] was faster than the *shortest* of their *uni*-modal component – auditory - [mean = 330, SD = 103]. The difference between these two conditions was highly significant [paired-t(15) = 3.60, p = 0.001]. b) When participants received a bi-modal combination of haptic and visual cues simultaneously, their RT [mean = 294, SD = 75] was faster than the *shortest* of their *uni*-modal component – haptic - [mean = 318, SD = 99]. The difference between these two conditions was also highly significant [paired-t(15) = 3.05, p = 0.004]. c) When participants received a bi-modal combination of haptic and auditory cues simultaneously, their RT [mean = 272, SD = 81] was faster than the *shortest* of their *uni*-modal component – haptic - [mean = 318, SD = 99]. The difference between these two conditions was also highly significant [paired-t(15) = 5.64, p < 0.000]. d) When participants received a tri-modal combination of haptic, visual and auditory cues simultaneously, their RT [mean = 263, SD = 69] was faster than the *shortest* of their *bi*-modal component – haptic and auditory - [mean = 272, SD = 81]. The difference between these two conditions was also significant [paired-t(15) = 2.2, p = 0.02].

Non-dominant hand

In the non-dominant hand, mean RT in the *uni*-modal conditions were also the longest. 436 ms. for the visual stimulus, 334 ms. for the haptic stimulus and 320 ms. for the auditory stimulus. All three *bi*-modal conditions were *shorter than any uni-modal condition*, 306 ms. for the haptic-visual combination, 304 ms. for the visual-auditory combination and 280 ms. for the haptic-auditory combination. RT in the *tri*-modal combination was the *shortest* – 277 ms. See figure 3.



	V	A	H	VA	HV	HA	HVA
□ Dominant Hand	430	330	318	302	294	272	263
■ Non-Dominant Hand	436	334	320	306	304	280	277

Figure 3: Reaction times in the uni- bi- and tri-modal conditions.

Paired comparisons analysis revealed that: a) When participants received a bi-modal combination of haptic and visual cues simultaneously, their RT [mean = 306, SD = 77] was faster than the *shortest* of their *uni*-modal component – haptic - [mean = 334, SD = 91]. The difference between these two conditions was highly significant [paired-t(15) = 3.4, p = 0.001]. b) When participants received a bi-modal combination of visual and auditory cues simultaneously, their RT [mean = 304, SD = 70] was faster than the *shortest* of their *uni*-modal component – auditory - [mean = 320, SD = 76]. The difference between these two conditions was also highly significant [paired-t(15) = 3.72, p = 0.001]. c) When participants received a bi-modal combination of haptic and auditory cues simultaneously, their RT [mean = 280, SD = 69] was faster than the *shortest* of their *uni*-modal component – auditory - [mean = 320, SD = 76]. The difference between these two conditions was also highly significant [paired-t(15) = 5.27, p < 0.000]. d) When participants received a tri-modal combination of haptic, visual and auditory cues simultaneously, their RT [mean = 277, SD = 76] was faster than the *shortest* of their *bi*-modal component – haptic and auditory - [mean = 280, SD = 69]. However, the difference between these two conditions was *not* significant [paired-t(15) = 0.51, p = 0.30].

Comparison of RT *between hands* in each condition, revealed *insignificant* differences (P values well above 0.05) between the dominant and the non-dominant hand in all *uni*- and *bi*-modal conditions, except for the *tri*-modal condition, in which there was a clear difference between the hands [paired-t(15) = 2.49, p < 0.01]. These *preliminary* results are still under analysis, especially the apparent difference between the dominant/non-dominant hands. Nevertheless, the results indicate a clear enhancement in all three *bi*-modal conditions, as compared to the *uni*-modal conditions, in *both* hands.

4. Discussion

These results provide evidence for a clear processing-speed advantage in all three *bi*-modal stimulations (VA, HV, HA) over *any* uni-modal stimulation (V, A, H). This advantage appeared in *both* hands. Furthermore, the results suggest a *special tri*-modal (HVA) processing-speed advantage over all three *bi*-modal conditions, at least in the dominant hand.

From a neuro-cognitive perspective, a possible explanation of these phenomena may be that our brain allocates *greater attention* to events activating *several neural systems simultaneously*, in comparison to events activating fewer neural systems. This *enhanced attention* may be the factor beyond the faster processing of these multi-modal events.

Although, reaction-time measurements do not *directly* indicate presence, we suggest the possibility that *they both share a common factor - enhanced attention*. That is to say, *multi-modal virtual environments may achieve a greater sense of presence, since they employ their users' attention and receptiveness to its maximum.* This greater attentional focus enables them to absorb more details and subtle cues from the display and integrate them creatively. An enhanced attention leads at the end of the cognitive process to a richer, more complete and coherent experience, and possibly, a greater sense of presence.

In addition, the advantage of multi-modal events at the initial perceptual stage (at the first 260-300 msc.) allow users *more time in the consequent cognitive stages* to creatively 'fill in' missing information and form a richer experience. For instance, in processing an event which lasts a *similar* time period, person A, stimulated by a single channel environment, is processing the incoming information *slower* than person B, stimulated by a multi-modal environment. Thus, *in a similar exposure time person B finishes the initial perception stage faster and can advance much further in the cognitive stream by paying attention to much more details and subtle cues in the graphic/auditory/haptic display*.

Therefore, *multi*-modal virtual environments *provide their users with a 'cutting edge' already early in the perceptual stage*, in the beginning of the cognitive stream, since multi-modal informative cues are perceived *faster*. Their clear advantage over users of single channel technologies, in the *starting point*, allow them *more time at the consequent stages to:* a) *Acquire a wider range of details and subtle cues* from the display. b) 'Fill in' missing information from one sensory channel with cues from another sensory channel. c) Integrate these informative cues from different sensory systems in an active and creating manner. As a result, the end product of this *longer, detailed and active cognitive effort* is a robustly richer, more 'colorful' and coherent experience, and possibly, a greater sense of being present in the virtual scene.

Implications for VE simulators

Designers of virtual driving and flying simulators may find special interest in this study as these simulators can be upgraded by using multiple signals (visual, auditory, haptic and proprioception) simultaneously. Since, in these simulators, one of the most important parameters for assessing driving and flying skills is the time it takes users to detect a car, a traffic sign, an object or a topographic view, creating multi-modal environments in which information is presented via multiple channels may significantly shorten reaction time.

These multi-modal simulations may be especially important to teach and assess driving and flying *during limited-vision conditions* such as twilight time, night, sharp curves on the road etc. as users can amplify the weak visual data and *"fill it in"* with appropriate auditory, proprioceptive and haptic cues.

References

[1] Biocca, F., Kim, J., & Choi, Y. (2001). *Visual Touch In Virtual Environments: An Exploratory Study of Presence, Multimodal Interfaces, and Cross-Modal Sensory Illusions.* ***Presence: Teleoperators & Virtual Environments.*** 10(3); 247-266.

[2] Dalton, P., Doolittle, N., Nagata, H. & Breslin, P.A. (2000). *The merging of the senses: integration of subthreshold taste and smell.* ***Nature Neuroscience.*** 3(5): 431-2.

[3] Gottfried, J.A. & Dolan, R.J. (2003). *The nose smells what the eye sees: crossmodal visual facilitation of human olfactory perception.* ***Neuron.*** 39(2): 375-86.

[4] Held, R. & Durlach, N. (1992). *Telepresence.* ***Presence: Teleoperators and Virtual Environments.*** 1 (1); 109–112.

[5] Mergner T, Rosemeier T. (1998). *Interaction of vestibular, somatosensory and visual signals for postural control and motion perception under terrestrial and microgravity conditions--a conceptual model.* **Brain Research: Brain Research Review.** 28(1-2):118-35.

[6] Romano, D.M., & Brna, P. (2001). *Presence and Reflection in Training: Support for Learning to Improve Quality Decision-Making Skills under Time Limitations.* ***CyberPsychology & Behavior.*** 4(2); 265-278.

[7] Sanchez-Vives, M.V. & Slater, M. (2005). *From presence to consciousness through virtual reality.* ***Nature Reviews: Neuroscience***. 6: 332-9.

[8] Sheridan, T. B. (1992). *Musings on Telepresence and Virtual Presence.* ***Presence: Teleoperators and Virtual Environments***. 1(1); 120–125.

[9] Spence, C., Nicholls, M.E. & Driver, J. (2001). *The cost of expecting events in the wrong sensory modality.* ***Perception & Psychophysics.*** 63(2): 330-336

[10] Bernstein, L. E., Auer, E. T. Jr., Moore, J. K. (2004). *Audiovisual speech Binding: Convergence or Association?* In Calvert. G., Spence, C. & Stein B. E. (Eds.) ***The Handbook of Multisensory Processes.*** Pp. 203-223. (MIT Press).

[11] Witmer, B. G., & Singer, M. J. (1998). *Measuring presence in virtual environments: A presence questionnaire.* ***Presence: Teleoperators and Virtual Environments.*** 7(3); 225-240.

A comparison of the effect that the visual and haptic problems associated with touching a Projection Augmented model have on object-presence

Emily Bennett and Brett Stevens
Department of Information Systems and Computer Applications, University of Portsmouth, UK
emily.bennett@port.ac.uk, brett.stevens@port.ac.uk

Abstract

A Projection Augmented Model (PA model) is a type of haptic Augmented Reality display. It consists of a real physical model, onto which a computer image is projected to create a realistic looking object. Users can physically touch the surface of a PA model with their bare hands, which has experiential value for the types of applications for which they are being developed. However, the majority of PA models do not provide haptic feedback for material properties such as texture, and hence feel incorrect when they are touched. In addition, most PA models are front-projected which means the projected image appears on the back of the user's hand, and their hand casts a shadow on the display. Previous research has found that touching this type of PA model reduces a user's sense of object-presence. The empirical study reported in this paper investigated which of the problems had a greater effect on object-presence. It was found that object-presence was significantly higher when correct haptic feedback for material properties was provided; however eliminating the visual projection problems often did not effect object-presence. These results have implications for the direction in which PA model technology should be developed. They also have implications for theory on how the haptic and visual senses contribute to a person's sense of object-presence, and indeed presence.

1. Background

Figure 1. PA model with projection on and off.

A Projection Augmented model (PA model) is a type of projection based haptic Augmented Reality display. It consists of a physical three-dimensional model, onto which a computer image is projected to create a realistic looking object [1]. For example, the PA model in figure 1 consists of smooth white plaster models of various objects that are commonly found in a garden shed [2]. The image projected onto these objects provides color and visual texture, which makes them appear to be made from different materials. PA models can either be front-projected (e.g. figure 1), or if a semi-transparent physical model is used, they can be back-projected (e.g. [3]).

PA models can create very realistic looking objects, which multiple people can view and interact with in a natural and intuitive way, for example using a touch or gesture based interface. These properties have lead to PA models being developed for applications such as museum displays [4][5], and for product design applications such as cars [6] and mobile telephones [7]. The latter application highlights the importance of the fact people can touch a PA model with their bare hands, and feel its physical shape. PA models can also provide some haptic feedback for the material properties of the object that it is representing. For example, a PA model of a fossil which was created using a cast from a real fossil, provides haptic feedback for texture [4]. This accurate haptic feedback for material properties makes the PA model feel more realistic to touch.

Thus, considering the examples outlined above, there are four types of PA model: front-projected with haptic feedback for material properties; front-projected without haptic feedback for material properties; back-projected with haptic feedback for material properties; and back-projected without haptic feedback for material properties.

The majority of the technology that enables PA models to be a viable display option has been developed for front-projected PA models that do not provide haptic feedback for material properties (*simple PA models*). Technology has been developed that semi-automates the setup procedure of a simple PA model, for example software that correctly registers the projected image onto the physical model [8][9]. This technology can be combined with rapid prototyping and 3D scanning techniques to semi-automate the whole construction process [6][9]. Additionally, it is possible for a simple PA model to be created that changes its physical shape within a limited range [10][11]. Technology has also been developed which enables dynamic visual effects to be simulated using the projected image; examples include animated specula highlights, apparent motion and different lighting conditions (see [12] for a review). It is also possible for the projected image to remain registered on the surface of a moving object [5][8].

Given that PA models aim to create realistic objects, an important issue is the user's sense of object-presence. Object-presence is the subjective feeling a particular object exists in a person's environment, when that object does not [1]. PA models use real physical objects; therefore it could be argued that PA models do actually exist in a person's environment. However, object-presence is the sense that the specific object the PA model is representing exists, as opposed to a white physical model and a projected computer image. Thus, a PA model is an essentially computer generated display because it is the projected image that gives the dummy physical objects meaning. Previous work has found that when a simple PA model is touched, a user's sense of object-presence is reduced [13].

Touching a simple PA model may reduce object-presence for two reasons. Firstly, when a person touches the surface of a simple PA model, the projected image appears on the back of their hand, and their hand casts a shadow onto the display. This draws attention to the fact the PA model is not a coherent object, and thus may reduce object-presence. Indeed, it has been found that the use of shadows on virtual objects viewed using an optical see-through augmented reality display can affect object-presence [14]. Additionally, a study that directly compared front and back projected flat-screen displays, reported that people found the shadows on the front-projected display very distracting [15]. The second possible reason why object-presence is reduced, is the mismatch between the visual and haptic feedback for material properties, such as texture, becomes apparent when a simple PA model is touched. This suggestion is supported by previous work which has shown that touch is very sensitive for perceiving material properties, such as texture [16].

Although using a back-projected display and providing haptic feedback for material properties could overcome these problems, technology is currently not sufficiently developed for these to be viable options (e.g. no automated setup procedures exist for back-projected PA models). These two areas are separate and distinct from each other, which suggests that research is needed to determine which is the most effective way of increasing object-presence, and hence the most effective area in which technological efforts need to be focused.

2. Research questions

This paper focuses on a user's sense of object-presence when touching a PA model. It investigates whether object-presence is increased by eliminating the visual projection problems, or by providing haptic feedback for material properties and hence making the PA model feel correct to touch. The results will indicate the most effective direction in which technology should be developed.

It is possible that the effect of eliminating the projection problems and providing haptic feedback for material properties will differ depending on the object that the PA model is representing. Different objects have different 'key' properties, which are most suited to being perceived by either the visual or the haptic sense [17]. The properties that are most suited to being perceived through the haptic sense are material properties (e.g. temperature and texture), whilst geometric properties (e.g. size and shape) are most suited to being perceived through the visual sense [17]. Clearly all objects have both 'haptic' and 'visual' properties; however their relative importance may be biased towards one or the other depending on the task being completed with the object. When considering where an object is on a 'visual' to 'haptic' scale, it is useful to think in terms of general activities, as opposed to specific tasks. By considering general activities in combination with an object's most prominent features, a rough position on the scale can be identified. For example, when choosing a box of breakfast cereal, the salient property is the visual design on the package, and hence the object can be classed as a 'visual' object. Whereas, the salient property when choosing high quality printing paper is how the paper feels to touch, and hence the object is classed as a 'haptic' object.

It is possible that if a PA model is touched that represents a 'haptic' object, object-presence will be reduced more by the lack of haptic feedback for material properties than the visual problems associated with the projected image. Conversely, if a PA model is touched that represents a 'visual' object, object-presence may be reduced more by the visual problems associated with the projected image.

However, it has been argued that touch is a human's reality sense [18], which suggests that the lack of correct haptic feedback for material properties will always reduce object-presence more than the problems associated with the projection. Conversely, the theory of Visual Capture, which argues that vision is the primary sense that dominates over the others [17], suggests that the visual projection problems will always reduce object-presence more than the lack of haptic feedback for material properties. In fact, it has been shown that visual feedback can be used to 'fool' the haptic sense, and make a person believe that they have felt something when they have not [19]. For example, a person can be made to believe that they have felt a specific texture by manipulating the visual feedback [20].

Thus, it is possible that the type of object a PA model represents does not matter; either eliminating the projection problems, or providing accurate haptic feedback for material properties, will always be the most effective solution. However, whilst one solution may always be more effective, the amount by which it is more effective may vary depending on the type of object that the PA model represents. Given that previous research does not provide a clear indication as to which of the two solutions is the most effective, this study investigates the following questions:-

1) *When touching a PA model that represents a **'visual'** object, to what extent is a person's sense of object-presence affected by the visual problems associated with the projection, and to what extent is it affected by the incorrect haptic feedback for material properties?*
2) *When touching a PA model that represents a **'haptic'** object, to what extent is a person's sense of object-presence affected by the visual problems associated with the projection, and to what extent is it affected by the incorrect haptic feedback for material properties?*
3) *Does the extent to which object-presence is affected by the projection problems and by the haptic feedback for material properties differ depending on the type of object (visual/haptic) that a PA model is representing?*

3. Experiment 1: Investigating question 1

3.1 Design and procedure

This experiment investigated the first research question (section 2). A 2x2 factorial design was used; the independent variables (IV) were *ProjectionDirection* (*Front-Projected* and *Back-Projected*), and *FeelsToTouch* (*FeelsCorrect* and *FeelsIncorrect*). Therefore 4 PA models were constructed, and were arranged so that the participant could use them singularly or simultaneously (figure 2).

The experiment was split into two parts; each participant completed part 1, and then part 2 directly afterwards. Part 1 had a between-participants design; each participant did a set of tasks which required them to touch just one of the 4 PA models (section 3.1.1). An additional condition was included in this part, in which participants used the Front-Projected+FeedsIncorrect (i.e. simple) PA model without touching it. This was to verify the reliability of a previous study, which concluded that touching a simple PA model decreases object-presence [13]. Therefore in part 1 there were actually 5 conditions with different participants in each; four 'touch' conditions and one *NoTouch* condition.

Part 2 of the experiment had a within-participants design; all participants completed the same tasks which required them to touch all four PA models simultaneously (section 3.1.2). A total of 50 participants were used, who were all students on computing related degree courses.

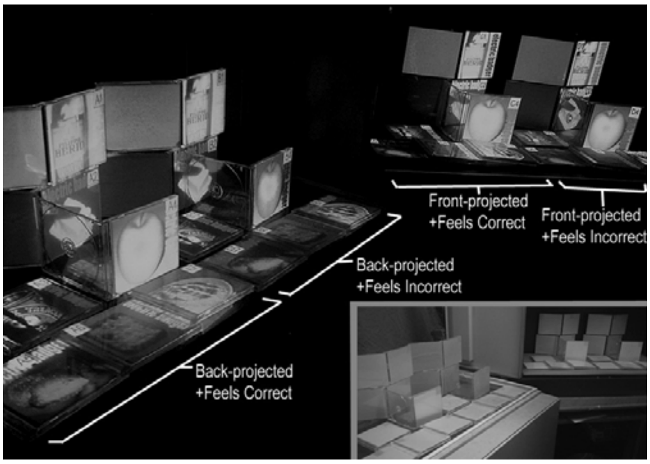


Figure 2. First experiment - 'visual' PA models.

The PA models represented CD cases and each of the four PA models consisted of a set of 7 cases (figure 2). CD cases were chosen because their most prominent feature is the visual design on the sleeve, and hence they can be classed as 'visual' objects. The CD cases were all firmly attached to the base and could not be moved. The image on the front-projected PA models was projected down from a 45 degree angle, which meant the projection problems only occurred when the participants touched the display. The projection equipment was equally visible for both the front and back-projected PA models.

The precise design of the PA models (CD cases) was directly related to the experimental tasks (sections 3.1.1 and 3.1.2). The sets of CD cases in the two FeelsCorrect conditions felt like normal plastic cases. Whereas, in the two FeelsIncorrect conditions, 4 of the CD cases felt like paper to touch, and 3 cases felt like paper with a sticky area in the center. The images on the CD cases were of foods that could either be described as sticky (e.g. treacle) or smooth (e.g. butter). The same set of images was used on each of the 4 sets of CD cases. On each CD case the words 'electric badger' were written (as if it was the title of the album). The combination of the text colour and the image was different for every CD case. Importantly, all four sets of CD cases were the same in terms of image content (apart for the text colour) and image quality.

The experiment aimed to be ecologically valid, and hence a fair refection of a real life situation. This was achieved by designing tasks that appeared to be 'natural' and 'sensible' activities to do with a PA model. The participants were told that the PA models were a new type of design system which was being developed to assist in the design of products. They were lead to believe that they were doing the experiment to investigate how people evaluate products (i.e. CD cases) using the system. Throughout the experiment the participants had to touch the CD cases to 'select' them (section 3.1.1 and 3.1.2). This gave participants the opportunity to perceive both the haptic properties of the CD cases and the projection problems. To make the participants feel that this was a 'sensible' activity, they were told they had to do this because their hand was being tracked. This 'story' was made believable by placing a camera above the display.

To operationalise the definition of object-presence, high object-presence was defined as '*a strong sense that the paper sleeves are inside the CD cases, and the images are physically printed onto them'*. This was based on the notion that if participants found the projection problems very noticeable, they may have troubling imagining that the images are physically printed on the paper sleeves of the front-projected CD cases. And, if they found the haptic properties of the CD cases noticeable, they may have trouble imagining that the paper sleeve is on the *inside* of the cases that felt incorrect to touch.

Ten 'measures' were designed based on this definition (sections 3.1.1 and 3.1.2). Not all of the measures investigated object-presence directly. The measures initially explored the issues related to object-presence, and gradually asked increasingly more direct questions. Measures 5, 6 and 7 aim to measure object-presence directly. The principal behind these measures was to ask direct, but essentially subjective, questions relating to the visual and the haptic problems. If the participants indicate that they can suspend their disbelief when they notice a problem, it suggests the 'problem' does not effect object-presence. For example, participants are asked '*The design system aims to give you the sense that the paper sleeves are inside the CD cases, and the images are physically printed onto them. Put the 4 sets in order based on how strongly you get this sense. You may give 2 or more sets same ranking.*' (measure 6) (section 3.1.2). If the participants ranked the two PA models that felt correct to touch as joint 1st, and the two that felt incorrect as joint 2nd, then the projection problems have not affect the ranking. This would suggest that people can suspend their sense of disbelief when they encounter the projection problems, but they cannot when they encounter the haptic problem. Thus, the haptic problems effect object-presence, but the projection problems do not.

A number of pilot studies (16 participants in total) were conducted to ensure that the participants understood the wording of the questions, and understood the questions were interested in their subjective opinion. After testing several versions of the questions and instructions, the following procedure was decided upon.

3.1.1 Part 1 (between-participants) procedure: In part 1 the participants completed a set of tasks which required them to touch just one of the 4 sets of CD cases (i.e. PA models). An additional condition was included in which participants used the Front-Projected+FeedsIncorrect (i.e. simple) PA model without touching it.

Each of the five conditions contained 10 participants. Firstly, the participants in the four 'touch' conditions touched each of the CD cases in the set that they were using. This was to ensure they had the opportunity to notice how the CD cases felt, however to make this appear to be a 'natural' activity they were told it was to calibrate the hand tracking device. Participants in the NoTouch condition had to look at each of the CD cases in the Front-Projected+FeelsIncorrect set; they were told this was to familiarize them with the system.

The participants in the four 'touch' conditions then did the following tasks by touching the CD cases to indicate their answer. The participants in the NoTouch condition did the tasks by saying the number on the CD cases.

The participants were asked to categorized the 7 CD cases into two groups based on the ones they thought *'go together or are similar in someway'* (**Measure 1**). This was to investigate the attention they paid attention to how the CD cases felt to touch; the participants in the two FeelsIncorrect conditions could answer this question based on the images on the CD cases, or based on how they felt to touch (section 3.1). This approach to assessing object-presence is derived from the concept of Cognitive Presence, which argues the way in which people interpret a question signifies which reality they are in [21]. They then categorized the 7 CD cases into two groups based on the ones that they *'would describe as sticky and those that they would describe as smooth'*. (**Measure 2**). The aim of this measure was the same as the last, except the participants were prompted towards 'sticky' and 'smoothness'.

They then completed a 'realistic' design task, in which they were asked to 'select' the CD cases in the order that they liked the images. It was this task and the 'realistic' design task in part 2 (section 3.1.2), that the participants believed was the focus of the experiment. However, this was simply a way of getting them to do a 'realistic' seeming task with the display, and nothing was recorded.

After completing these tasks, the participants faced away from the display and were asked to describe their *'main memory of the design system'* (**Measure 3**) and their *'main memory of what they saw'* (**Measure 4**). This was to investigate how noticeable the participants found the image being projected on their hand, and the shadows. They also completed the following object-presence questionnaire by giving an answer on a five point Likert Scale, where 1 = strongly disagree and 5 = strongly agree (**Measure 5**). Reverse scoring was used for questions 3 and 7. They were told to answer the questions based on *'their subjective sense or feeling, and not what they know to be true'*.

1) *I had a strong sense or feeling that the CD cases had a smooth clear plastic front.*
2) *The design system was a very natural way of presenting information.*
3) *I constantly paid attention to the design systems deficiencies / problems.*
4) *I had a strong sense that the paper sleeve was inside the CD cases.*
5) *I had a strong sense that the images were physically printed in coloured ink.*
6) *I can easily believe that the front of all of the CD cases felt like plastic to touch.*
7) *I had a strong sense that parts of the design system were computer generated.*

3.1.2 Part 2 (within-participants) procedure: In part 2, all 50 participants did the same tasks using the four sets of CD cases (i.e. all 4 PA models) simultaneously.

The participants first completed another 'realistic' design task (related to their preference for the text colour), which ensured they touched all of the CD cases and nothing was recorded. The participants were then asked to put the four sets of CD cases in order based on the sense of object-presence that they felt. Specifically, they were asked: *'The design system aims to give you the sense that the paper sleeves are inside the CD cases, and the images are physically printed onto them. Put the 4 sets in order based on how strongly you get this sense.'* (**Measure 6**). They were allowed to give two, three or all the sets the same ranking, and they could touch the CD cases to make their decision. The participants were then asked to decide for each set of CD cases, whether touching the cases increased or decreased their sense of object-presence. Specifically, they were asked: '*Consider each set in turn. Can you decide whether touching the CD cases increases or decreases your sense that the paper sleeve is inside the case and the image is physically printed onto it, or does touching make no difference?'* (**Measure 7**). Again, they could touch the CD cases to make their decision.

After this, they were asked a direct question about their sense of object-presence with regards to how the CD cases felt to touch: '*Which set or sets gives you the strongest sense that the paper sleeve is inside the CD cases*' (**Measure 8**). And, they were asked a direct question about their sense of object-presence with regards to the projection problems: '*Which set or set gives you the strongest sense that the images are physically printed on the paper sleeve?'* (**Measure 9**). For both questions they could touch the CD cases to make their decision.

Finally, the participants faced away from the display and were asked whether or not they had noticed the projected image on their hand, the shadows, and how the CD cases felt to touch. (**Measure 10**).

3.2 Results of the first experiment

3.2.1 Measure 1 (allocating the CD cases into two groups based on the ones which *'you think go together or are similar in some way'*). 48/50 participants did this task based on the images on the CD cases. 2 participants (who were unsurprisingly in the two FeelsIncorrect conditions) did this task based on touch.

3.2.2 Measure 2 (categorization of the CD cases based on *'those that you would describe as sticky and those that you would describe as smooth'*). As expected, all participants (except for one) in the two FeelsCorrect conditions and the NoTouch condition did this task based on the images on the cases. 7/10 participants in the Back-Projected+FeelsIncorrect condition, and 8/10 participants in the Front-Projected+FeelsIncorrect condition did this task based on how the cases felt to touch; therefore a significant number of participants in the FeelsIncorrect conditions (combined) categorized the CD cases based on how they felt to touch ($X^2(1)=5.0$, $p<0.05$).

3.2.3 Measure 3 (*'what did you find most noticeable?')* and **measure 4** *('what was your main memory of what you saw?').* The participants responses to both measures were divided into the following categories; 'feels to touch', 'projection on hand', 'shadows hand cast', 'task', 'design on the CD case' and 'equipment setup'. The results were very similar for both measures and there was virtually no difference between conditions. For both measures, the majority of participants' responses (38/50 for measure 3 and 44/50 for measure 4) fell in the 'design on the CD case' category (e.g. they recalled the treacle); this figure was made up of an approximately equal number of responses from each condition. For both measures, only one participant's response fell into the 'image on hand' category (both in the Front-Projected+FeelsIncorrect condition), and no participants mentioned the shadows.

3.2.4 Measure 5 (questionnaire) (figure 3) .

	Condition / PA model				
Question	NoTouch	BP+FC	BP+FIC	FP+FC	FP+FIC
1 Mean	3.70	4.00	2.00	3.90	2.40
s.d	(0.82)	(0.94)	(0.82)	(0.88)	(0.97)
2 Mean	3.80	3.40	3.10	3.20	3.00
s.d	(0.92)	(0.84)	(0.88)	(1.03)	(1.05)
3 Mean	3.30	3.10	3.10	3.10	2.90
s.d	(0.95)	(1.10)	(0.88)	(0.87)	(0.74)
4 Mean	3.70	3.50	2.30	4.00	2.20
s.d	(1.49)	(1.18)	(0.82)	(1.05)	(0.79)
5 Mean	4.00	3.70	3.80	3.50	3.30
s.d	(1.15)	(1.49)	(1.03)	(1.35)	(1.25)
6 Mean	4.20	4.20	2.70	4.40	3.30
s.d	(0.92)	(1.03)	(1.16)	(0.69)	(1.16)
7 Mean	2.30	2.20	2.30	3.00	2.00
s.d	(1.16)	(1.14)	(0.95)	(1.15)	(1.15)

Figure 3. First experiment - measure 5 results. (BP=Back-projected, FP=Front-projected, FC=Feels correct, FIC=Feels incorrect).

A Cronbach Alpha test for internal consistency was found to not be significant (0.43). A principal component factor analysis was conducted on the results to determine whether groups of questions were measuring separate constructs. Three subscales were found; questions 1, 4, and 6 (which relate to how the CD cases felt), 2 and 3 (which relate to the 'naturalness' of the display) and questions 5 and 7 (which relate to the projection problems). However, when the analysis was conducted on each individual condition, this pattern was not found. Although this is unsurprising because each condition contained only 10 participants, it means that the questions must be analyzed separately.

Firstly, comparing between the four 'touch' conditions (i.e. not including the NoTouch condition), and hence investigating question 1. A suitable nonparametric test was not available (i.e. a nonparametric two-way ANOVA type test). Work is currently being done to develop this type of test [22], however a usable version is not yet available. It has been argued that ordinal data, such as Likert Scales, can be treated as interval data for the purpose of statistical analysis [23], therefore this approach was taken. Levene's test for equality of variance was conducted on each question and found no significant difference between conditions. Therefore a two-way between-participants MANOVA was conducted. A significant effect was found for the FeelsToTouch IV for questions 1 ($F(1,36)=37.63$, $p<0.001$), 4 ($F(1,36)=23.68$, $p<0.001$) and 6 ($F(1,36)=15.93$, $p<0.001$), but not for the other questions. No significant effects were found for the ProjectionDirection IV for any of the questions, and no significant interactions were found for any questions.

Now comparing the NoTouch condition to each of the four 'touch' conditions (i.e. which could confirm the conclusion from previous work). A Man Whitney test found a significant difference between the NoTouch condition and the Back-Projected+FeelsIncorrect condition for questions 1 ($U(18)=8$, $p<0.01$), 4 ($U(18)=23$, $p<0.05$) and 6 ($U(18)=16$, $p<0.01$), but not for the other questions. A significant difference was found between the NoTouch condition and the Front-Projected+FeelsIncorrect condition for questions 1 ($U(18)=15$, $p<0.01$) and 4 ($U(18)=22$, $p<0.05$), but not for the other questions. No significant difference was found between the NoTouch condition and the Back-Projected+FeelsCorrect condition, and the NoTouch condition and the Front-Projected+FeelsCorrect condition, for any of the questions.

3.2.5 Measure 6 (Putting the four sets of CD cases in the order of the sense of object-presence that they create). Although the data that this measure generates is ipsative (i.e. the ranking that a participant gives one set of CD cases affects the ranking they give another), previous research has concluded that it is valid to perform an ANOVA on this type of data [24]. Moreover, the participants were given the option of giving two (or more) sets of CD cases the same ranking (e.g. joint second place). This meant that the participants were not forced to create an 'artificial' rank order. Therefore the data was analyzed using an ANOVA by converting each ranking into a score of 1 (lowest) – 4 (highest). If a participant said that two sets gave them the same sense of object-presence, the mean between the two ranks was given for both. (Figure 4).

Condition	Mean (standard deviation)
Back-Projected+FeelsCorrect	3.13 (0.92)
Back-Projected+FeelsIncorrect	2.02 (1.00)
Front-Projected+FeelsCorrect	3.02 (1.01)
Front-Projected+FeelsIncorrect	1.83 (0.66)

Figure 4. First experiment - measure 6 results.

A within-participants ANOVA was conducted. A significant effect was found for the FeelToTouch IV (F(1,49)=52.16, p<0.001). No significant effect was found for ProjectionDirection IV (F(1,49)=1.06, p=0.31) and no significant interaction was found (F(1,49)=0.08, p=0.78).

3.2.6 Measure 7 (For each set of CD cases, what effect does touching have on object-presence.) (Figure 5).

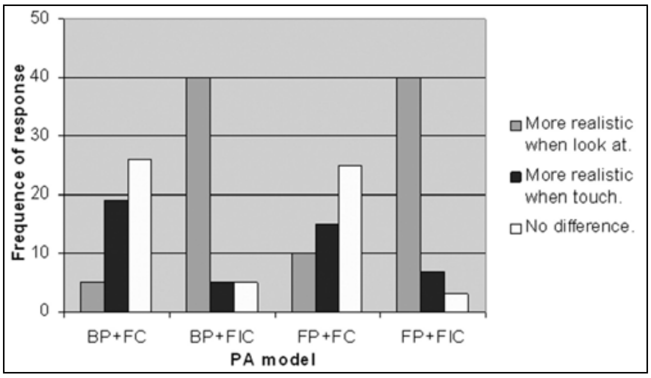


Figure 5. First experiment - measure 7 results. (BP=Back-projected, FP=Front-projected, FC=Feels correct, FIC=Feels incorrect).

3.2.7 Measure 8 (*'which set or sets gives you the strongest sense that the paper sleeve is inside the CD cases?'* i.e. direct question relating to how the CD cases felt to touch). 44/50 participants chose one or both of the sets of CD cases that felt correct to touch.

3.2.8 Measure 9 (*'which set or sets gives you the strongest sense that the images are physically printed on the paper sleeve?'* i.e. direct question relating to the projected image). 17/50 participants said the four sets of CD cases were the same. 4/50 participants chose both the Back-Projected sets, 4/50 chose the Back-projected+FeelsCorect set and 16/50 participants chose the Back-Projected+FeelsIncorrect set. 6/50 chose either of the Front-Projected sets, and 3/50 chose both the FeelsIncorrect sets.

3.2.9 Measure 10 (did participants actually notice the projection and haptic problems). 19/50 participants noticed the image on the back of their hand, and 26/50 participants noticed the shadows that their hands cast. A 2x2 Chi-Squared test found a significant relationship between noticing the image on the hand and noticing the shadows ($X^2(1)=12.74$, p<0.001). 49/50 participants noticed the CD cases did not all feel the same to touch, and 49/50 noticed that some CD cases felt sticky.

3.2.10 Comparing between measures. The data can be split based on whether or not the participants noticed the projected image on their hand (measure 10). Re-running the within-participants ANOVA showed no difference in the pattern of results for any of the measures, except for measure 9. Measure 9 asked '*which set or sets gives you the strongest sense that the image is physically printed on the paper sleeve?'*. 74% of the participants who noticed the projection on their hand gave a back-projected set of CD cases as their answer, whereas only 32% of participants who did not notice gave a back-projected set as their answer. Splitting the data based on whether participants noticed the shadows that their hands cast on the CD cases (measure 10) showed no significant difference in the patterns of results for any of the measures.

3.3 Discussion of the first experiment

Participants did not question the true goals of the experiment, and they appeared to put much thought into the 'realistic' design tasks. Moreover, no participants questioned whether or not their hand was really being tracked. This suggests that they believed the scenario they were told, and hence the results can be considered to be a fair reflection of a real task. Although there was a slight possibility that using the visual design scenario could focus participants' attention towards the visual projection problems, this did not occur.

In fact, the experiment found that a surprisingly high number of participants did not actually notice the projection problems, whereas (virtually) all participants noticed the haptic problems (measure 10, section 3.2.9). This *Innattentional Blindness* is the failure to notice information in the visual angle of the fovea [25]. Object-presence is concerned with whether participants can suspend their sense of disbelief when they encounter a problem, which they obviously cannot do if they have not actually noticed the problem. Therefore only the results of the participants who noticed the problems should be taken into consideration when addressing the research questions. However, when the results were split based on whether the problems were noticed, the same pattern of results were found for the measures that directly addressed object-presence (i.e. 5, 6 and 7) (section 3.2.10).

The results confirm the finding of previous work [13]; the majority of participants reported that object-presence was lower when the Front-Projected+FeelsIncorrect (i.e. simple) PA model was touched, compared to when it was just looked at (figure 5).

The main research question asked '*When touching a PA model that represents a 'visual' object, to what extent is a person's sense of object-presence affected by the visual problems associated with the projection, and to what extent is it affected by the incorrect haptic feedback for material properties?'*. The two measures directly addressed this question were the questionnaire (measure 5) and the ranking of the sets of CD cases based on the sense of object-presence they induce (measure 6). However, the questions in the questionnaire cannot be considered together because it was found they did not measure a unified construct (section 3.2.4), therefore the questionnaire cannot be used as a direct measure of object-presence.

Considering measure 6 (section 3.2.5); it was found that the two sets of CD cases that provided correct haptic feedback for material properties and hence felt correct to touch, were ranked significantly higher than the sets that did not. However, the direction of the projected image did not effect how the sets were ranked. This suggests that it is how a ‘visual’ PA model feels to touch that effects object-presence, as opposed to the projection problems. The other measures provide more subtle indicators, which generally support this conclusion.

Firstly, when classifying the CD cases as ‘sticky’ and ‘smooth’ (measure 2, section 3.2.2), a significant number of participants in the two FeelsIncorrect conditions did this based on how the CD cases felt to touch. This shows that participants found how the cases felt very noticeable. The fact the results for this measure follow the same pattern as the other measures suggests that the Cognitive approach to investigating presence [21] is a reliable method.

The results from the questionnaire (measure 5, section 3.2.4) show a similar pattern. Although the questions cannot be considered together, the individual questions can be examined. It was found that the haptic feedback the CD cases provided for material properties significantly affected responses to the questions relating to how they felt to touch. Moreover, the results support the argument that the problems associated with the projection are less important because it was found the direction of the projected image had no significant effect on any of the questions.

Further support for this pattern of results can be found by examining the effect that touching the CD cases had on object-presence, in comparison to when they were not touched (measure 7, figure 5). It was found that object-presence was lower when the two sets of CD cases which felt incorrect were touched. Whereas the projection problems did not appear to reduce object-presence when the front-projected CD cases were touched.

When questioned separately about the haptic issues, a similar pattern was also found. The majority of participants (44/50) reported that the two sets of CD cases that felt correct to touch gave them a stronger sense that the paper sleeve was inside the CD case (measure 8, section 3.2.7); hence supporting the argument that the haptic feedback for material properties is important.

However, when questioned separately about the projection problems, the results are not as consistent (measure 9). This is the only measure that showed a different pattern between the participants who noticed the projected image on their hand and those who did not (section 3.2.10). 74% of the participants who did notice it, chose only the back-projected sets of CD cases when asked to say which set/s gave them a stronger sense that the image was physically printed, compared to only 32% of the participants who did not notice it. This suggests that if people notice the projection on their hand, it can have an effect when asked a more direct question.

The only measures that do not support the general pattern of results are measures 3 and 4. When asked questions about what they found most noticeable and what was their main memory of what they saw, the vast majority of the participants’ responses were related to the content of the projected image (e.g. they remembered the apple design). This suggests that how the CD cases felt to touch, or indeed the projection problems, were not considered important enough to mention. However, the participants were told that the experiment was investigating how people evaluate the designs of CD cases, so it is possible that they were responding to demand characteristics.

To summarize, when a PA model is touched that represents a ‘visual’ object, object-presence is strongly affected by the haptic feedback provided for material properties, i.e. how it feels to touch. Although people tend not to find the visual projection problems noticeable, when they are noticed, they can affect responses to very direct questions relating to object-presence.

4. Experiment two – investigating question 2

4.1 Design and procedure

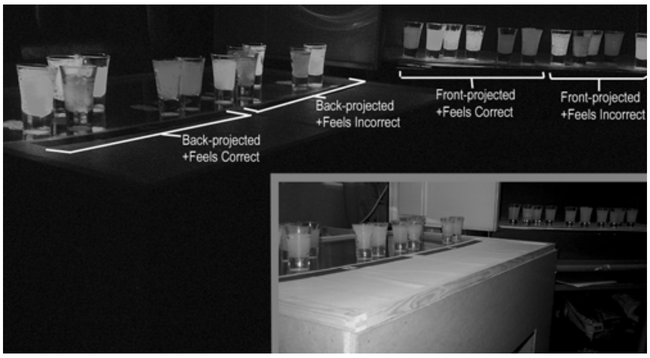


Figure 6. experiment 2 - ‘haptic’ PA models.

This experiment investigated the second research question (section 2). The same experimental design and procedure as the first experiment was used (section 3.1), however in this experiment the PA models represented ‘haptic’ objects. The specific objects that the PA models represented were vodka jellies (figure 6). (A vodka jelly is a fruit jelly, which is made with vodka and set into a shot glass; they are sold in many of the bars on the university campus). Vodka jellies are ‘haptic’ objects because their prominent feature is that they are made from, and hence feel like, jelly. Each of the four PA models used in this experiment consisted of a set of 6 vodka jellies. The models were actually made from gel candle wax, which was set into shot glasses. The vodka jellies in the two FeelsCorrect conditions felt like jelly to touch, and the vodka jellies in the two FeelsIncorrect conditions felt hard (this was achieved by setting a thin layer of clear resin on top of the wax). Importantly, all four sets of vodka jellies were visually identical, and they were all firmly attached to the base.

The same scenario as the first experiment was used; the participants were told that they were doing the experiment to investigate how people evaluate products using the system. The participants did equivalent ‘realistic’ design tasks as the participants in the first experiment, which involved them giving their preference for the colours. They had to touch the jelly itself (as opposed to the glass) to indicate their answer, which gave participants the opportunity to perceive both the haptic properties of the jellies and the projection problems. Again, they were told they were doing this because their hand was being tracked.

The same procedure and measures as experiment 1 were used, except for measures 1 and 2 which were not used because it was not practical to create an equivalent measure. However, the numbering of the measures starts at 3 to maintain consistency. The questions for each measure were reworded to assess the following operationalised definition of object-presence: high object-presence is '*a strong sense that the glasses contained jelly that is physically coloured*'. This was based on the notion that if participants found the projection problems noticeable, they may have troubling imagining the jelly is actually coloured when using the front-projected jellies. And, if they found the haptic properties of the jellies noticeable, they may have trouble imagining that the jellies which felt incorrect are actually made from jelly. Again, a pilot study was conducted before running the experiment. Detailed explanations of each stage of the procedure can be found in sections 3.1.1 and 3.1.2, however the following sections provide a summery. Again a total of 50 participants took part (not the same people as in experiment 1), who were students on computing degree courses.

4.1.1 Part 1 (between-participants) procedure: In part 1, the participants completed a set of tasks which required them to touch just one of the 4 sets of vodka jellies (i.e. PA models). An additional condition was included in which participants used the Front-Projected+FeedsIncorrect (i.e. simple) PA model without touching it. Each of the five conditions contained 10 participants. They first completed equivalent 'calibration' and 'design' tasks as the participants in experiment 1. The participants in the four 'touch' conditions did these tasks by touching the vodka jellies. The participants in the NoTouch condition did the tasks by saying the colour of the jellies. After this, they faced away from the display and were asked '*what did you find most noticeable?*' **(Measure 3)** and '*what was your main memory of what you saw?*' **(Measure 4).** They then completed the following object-presence questionnaire by giving answers on a five point Likert scale, where 1 = strongly disagree and 5 = strongly agree **(measure 5)** (questions 1 and 3 are reversed scored):-

1) *I had a strong sense that parts of the design system were computer generated.*
2) *The design system was a very natural way of presenting information.*
3) *I constantly paid attention to the design systems deficiencies / problems.*
4) *I had a strong sense that there were Vodka Jellies present in front of me.*
5) *I had a strong sense that each glass in the set that I focused on had a different colored material inside of it.*
6) *I can easily believe that there was jelly inside the shot glasses.*

4.1.2 Part 2 (between-participants) procedure: In the second part, all 50 participants completed the same tasks using the four sets of vodka jellies (i.e. PA models) simultaneously. They first completed an equivalent 'design' task as the participants in the first experiment, which required them to touch all four sets of vodka jellies. They were then asked to put the four sets of vodka jellies in order based on the sense of object-presence they felt **(Measure 6)**. Specifically, they were asked '*The design system aims to give you the sense that the shot glasses contain coloured jelly. Can you put the sets in order based on how strongly they give you this sense*'. They were allowed to give two or more sets the same ranking. After this, they had to decide for each set of vodka jellies, whether touching increased or decreased object-presence **(Measure 7).** Specifically, they were asked '*Consider each set in turn. Can you decide whether touching the vodka jellies increases or decreases your sense that the shot glasses contain coloured jelly, or does touching make no difference*'. They were then asked '*which set or sets gives you the strongest sense that the objects are made from the correct material?*' i.e. a direct question about object-presence relating to how the PA models felt to touch **(Measure 8).** And, they were asked '*which set or sets gives you the strongest sense that the jelly is actually coloured?'* i.e. a direct question about object-presence relating to the projected image **(Measure 9).** Whilst completing measures 6, 7, 8 and 9 the participants were allowed to touch the vodka jellies.

Finally, they faced away from the display and were asked whether they actually noticed the projection problems and the haptic feedback for material properties **(measure 10)**.

4.2 Results of the second experiment

4.2.1 Measures 3 and 4. The same categories as in the first experiment were used to group the responses to measures 3 and 4 (section 3.2.3), except for the 'design on the CD cases' category which was replaced with 'colours'. For both measures, the pattern of results was very similar and there was virtually no difference between conditions. For both measures, the majority of responses (31/50 for measure 3, and 40/50 for measure 4) fell into the 'colours' category (e.g. the participant listed all the colours of the vodka jellies); this figure was made up of roughly an equal number of responses from each condition. For measure 3, 3 of the participants' responses fell into the 'projection on hand' category, and for measure 4, 4 of the participants' responses fell into this category; all of these responses came from participants in the Front-Projected+FeelsIncorrect condition. For both measures 3 and 4 no participants mentioned the shadows their hands cast on the display.

4.2.2 Measure 5 (Figure 7).

	Condition / PA model				
Question	NoTouch	BP+FC	BP+FIC	FP+FC	FP+FIC
1 mean	3.40	3.60	3.40	3.70	3.20
s.d	(0.97)	(1.26)	(0.70)	(0.82)	(1.03)
2 mean	4.10	3.50	2.70	3.50	2.50
s.d	(0.99)	(1.35)	(1.25)	(0.97)	(1.08)
3 mean	4.00	3.80	2.80	3.50	2.60
s.d	(0.94)	(0.79)	(1.23)	(0.97)	(1.26)
4 mean	3.30	4.10	2.50	3.80	2.90
s.d	(0.95)	(1.10)	(1.43)	(1.13)	(1.52)
5 mean	3.80	3.80	3.80	4.10	3.30
s.d	(1.23)	(0.92)	(1.23)	(0.74)	(1.25)
6 mean	3.90	4.20	2.20	4.40	2.10
s.d	(1.19)	(1.03)	(1.40)	(1.26)	(1.19)

Figure 7. Second experiment - measure 5 results, (BP=Back-Projected, FP=Front-projected, FC=Feels correct, FIC=Feels incorrect).

A Cronbach Alpha test for internal consistency was found to not be significant (0.52). A principal component factor analysis was then conducted on the results. Similar subscales to the first experiment were found; questions 2, 3 and 4 ("naturalness of the display"), questions 1 and 5 (projection problems) and question 6 (how the display felt to touch). However, when the principal component factor analysis was conducted on each condition, this pattern was not found. This means that the questions will have to be analyzed separately.

Firstly, comparing between the four 'touch' conditions (i.e. not including the NoTouch condition), and hence investigating question 2. Levene's test for equality of variance was conducted on each question and found no significant difference between conditions. Therefore a two-way between-participants MANOVA was conducted (see section 3.2.4 for justification of test). A significant effect was found for the FeelsToTouch IV for questions 2 (F(1,36)=4.41, p<0.05), 3 (F(1,36)=5.48, p<0.05), 4 (F(1,36)=5.53, p<0.05) and 6 (F(1,36)=21.62, p<0.001), but no significant effects were found for questions 1 and 5. No significant effects were found for the ProjectionDirection IV for any of the questions, and no significant interactions were found for any questions.

Now comparing the NoTouch condition to each of the four 'touch' conditions. A significant difference was found between the NoTouch condition and the Back-projected+FeelsIncorrect condition for questions 2 (U(18)=20.00, p<0.05), 3 (U(18)=23.00, p<0.05), and 6 (U(18)=19.50, p<0.05), but not for the other questions. A significant difference was found between the NoTouch condition and the Front-Projected+FeelsIncorrect condition for question 2 (U(18)=14.00, p<0.05), 3 (U(18)=19.00, p<0.05), and 6 (U(18)=16.50, p<0.05), but not for the other questions. No significant differences were found between any questions when comparing the NoTouch condition to the Back-Projected+FeelsCorrect condition, and to the Front-Projected+FeelsCorrect condition.

Condition	Mean (standard deviation)
Back-Projected+FeelsCorrect	3.42 (0.52)
Back-Projected+FeelsIncorrect	1.85 (0.67)
Front-Projected+FeelsCorrect	3.39 (0.64)
Front-Projected+FeelsIncorrect	1.69 (0.70)

Figure 8. Second experiment - measure 6 results.

4.2.3 Measure 6. (Figure 8). A within-participants ANOVA was conducted (see section 3.2.5 for justification of test). A significant effect was found for the FeelsToTouch IV (F(1,49)=317.14, p<0.001). No significant effect was found for the ProjectionDirection IV (F(1,49)= 0.91, p=0.35) and no significant interaction was found (F(1,49)= 0.39, p=0.54).

4.2.4 Measure 7. See figure 9.

4.2.5 Measure 8. 48/50 participants chose one or both of the sets of the vodka jellies that felt correct to touch.

4.2.6 Measure 9. 23/50 participants chose either one or both of the back-projected sets of vodka jellies, 10/50 participants chose either one or both of the front-projected sets, 17/50 participants said the four sets were the same.

4.2.7 Measure 10. 29/50 participants noticed the image on the back of their hand, and 34/50 noticed the shadows their hands cast on the PA model. A 2x2 Chi-Squared test found a significant relationship between noticing the image on the hand and noticing the shadows ($X^2(1)=20.00$, p<0.001). All of the participants noticed the vodka jellies did not all feel the same to touch.

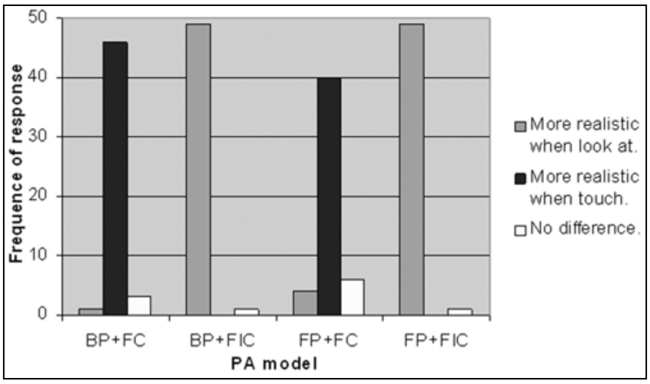


Figure 9. Second experiment - measure 7 results. (BP=Back-projected, FP=Front-projected, FC=Feels correct, FIC=Feels incorrect).

4.2.8 Comparing between measures. Splitting the data based on whether participants noticed the projected image on their hand showed no difference in the pattern of results, and neither did splitting the data based on whether they noticed the shadows that their hand cast on the display.

4.3 Discussion of the second experiment

The results of the second experiment are very similar to those of the first experiment. The participants accepted the scenario that they were told, which indicates that the results can be considered to be a fair reflection of a real task. Again, there was a slight possibility that using the design scenario could focus participants' attention towards the visual projection problems, however, this did not occur.

This experiment also found that Inattentional Blindness occurred for a high number of participants and they did not notice the projection problems (measure 10, section 4.2.7). Again, this raised the issue that object-presence is concerned with whether participants can suspend their sense of disbelief when they encounter a problem, which they obviously cannot do if they have not actually noticed the problem. Therefore only the results of the participants who noticed the problems should be taken into consideration when addressing the research questions. However, similar to the first experiment, when the results were split based on whether the problems were noticed, the same pattern of results were found (section 4.2.8).

The results from measure 7 (figure 9) confirm the finding of previous work [13]; touching a simple PA model reduces object-presence. Considering how the results relate to the main research question: *Question 2. 'When touching a PA model that represents a 'haptic' object, to what extent is a person's sense of object-presence affected by the visual problems associated with the projection, and to what extent is it affected by the incorrect haptic feedback for material* *properties?'*. The two measures that aimed to directly address this question were the questionnaire (measure 5) and the ranking of the sets of vodka jellies based on the sense of object-presence that they induce (measure 6). However, the questions in the questionnaire cannot be considered together because it was found that they did not measure a unified construct (section 4.2.2). This means that the questionnaire cannot be used as a direct measure of object-presence.

Measure 6 found (section 4.2.3) the two sets of vodka jellies that provided correct haptic feedback for material properties, and hence felt correct to touch, were ranked significantly higher than the sets that did not. However, the direction of the projected image did not effect how the sets were ranked. The results indicate that how a 'haptic' PA model feels to touch is important, whereas the problems associated with the projection are relatively unimportant. The other measures generally support this conclusion.

The results from the questionnaire (measure 5) show a similar pattern. Although the questions cannot be considered together, the individual questions can be examined. The haptic feedback provided for material properties significantly affected the responses to the questions relating to how the vodka jellies felt to touch. Additionally, the results suggest the problems associated with the projection are relatively unimportant because no significant effect was found for the ProjectionDirection IV on any of the questions.

Further support for this pattern of results can be found by examining the effect that touching the vodka jellies had on object-presence, in comparison to when they were not touched (measure 7, figure 10). It was found that object-presence was lower when the two sets of vodka jellies that felt incorrect were touched. Whereas the problems associated with the projected image did not appear to reduce object-presence when the Front-Projected vodka jellies were touched.

Finally, when questioned separately about the haptic issues and projection problems, a similar pattern was found again. Virtually all participants reported that the two sets of vodka jellies that felt correct to touch, gave them a stronger sense that the objects were made from the correct material (measure 8, section 4.2.5); hence supporting the argument that the haptic feedback for material properties is important. Moreover, participants did not tend to select back-projected PA models when asked which set/s gives the strongest sense that the jellies are actually colored (measure 9, section 4.2.6), which supports the argument that the projection problems are relatively unimportant.

Similar to the first experiment, the only measures that do not support the general patterns of results are measures 3 and 4 (section 4.2.1). When asked questions about what they found more noticeable and what was their main memory of what they saw, the vast majority of the participants' responses were related to the colors (e.g. they remembered the green jelly). This suggests that how the vodka jellies felt to touch, or indeed the projection problems, were not considered important enough to mention. However, the participants were told that the experiment was investigating how people evaluate the designs of vodka jellies, so it is again possible that they were responding to demand characteristics.

To summarize, when a PA model is touched which represents a 'haptic' object, object-presence is strongly affected by the haptic feedback provided for material properties, i.e. how it feels to touch. However, there is no evidence to suggest that object-presence is affected by the visual problems associated with the projection.

5. General discussion

The two experiments can be examined together to investigate the 3rd question: *'Does the extent to which object-presence is affected by the projection problems and by the haptic feedback for material properties differ depending on the type of object (visual/haptic) that the PA model is representing?'*. The overall pattern of results suggest that how a PA model feels to touch is the most important factor, and the projection problems are relatively unimportant (sections 3.3 and 4.3). In fact, Inattentional Blindness often occurred, and many participants did not notice the projection problems (sections 3.2.9 and 4.2.7).

However, the results do suggest that the projection problems are more of an issue for PA models that represent 'visual' objects, than for those that represent 'haptic' objects. It was found that if participants did notice the projection problems when using the 'visual' PA model (CD cases), they tended not to suspend their sense of disbelief when questioned directly (section 3.2.10). However, this did not occur when the PA model represented 'haptic' objects (vodka jellies) (section 4.2.8). Further support comes from measure 7. Measure 7 investigated whether the participants' sense of object-presence increased, decreased or stayed the same when touching each PA model, compared to when it was not touched. Touching the two 'haptic' PA models (vodka jellies) that felt correct always increased the participants' sense of object-presence (figure 9). Whereas when the participants touched the two 'visual' PA models (CD cases) that felt correct, similar numbers reported their sense of object-presence remained the same as those who reported it was increased (figure 5). This suggests that whilst providing incorrect haptic feedback for material properties will always decrease object-presence regardless of the object that the PA model represents, providing correct haptic feedback is more important for PA models that represent 'haptic' objects.

To conclude the answering of the third research question; the results suggest that how a PA model feels to touch is the most important factor regardless of the object that it is representing. However, the results support the original argument (section 2) that the projection problems are more important when a PA model represents a 'visual' object, and providing correct haptic feedback is more important when the PA model represents a 'haptic' object.

Focusing on the implications that the results have for PA models; the results suggest that technology needs to be developed to overcome the problems associated with a PA model feeling incorrect to touch. One possibility is to create the physical model which naturally provides haptic feedback for material properties, for example giving it a physical texture. This is suitable for displays that have a fixed physical shape, and only their colour and visual information are altered. For example, a PA model that represents a fossil could have a fixed shape and only the text annotations on the fossil may be altered depending on whether an expert or novice is viewing it, e.g. [4]. However, this approach reduces the flexibility of a PA model to a level that may be unacceptable for some applications. Additionally, for many objects it is not possible to create a PA model that provides accurate haptic feedback for material properties because it would not provide a suitable surface on which to project an image.

An alternative solution is to provide some form of haptic feedback through a separate device. For example, a hand-held tracked tool could be used to provide vibration feedback to simulate the feeling of physical texture, e.g. [26]. Indeed, it has been found that humans are good at perceiving material properties, such as texture, through vibrations simulated using a probe [27]. Another solution is to use a visual cue to indicate texture. For example, a PA model could be interacted with using a normal mouse, whose cursor could be animated so that it deforms to suggest that it is moving over a textured object. This type of ‘pseudo haptic’ feedback has been shown to be effective for flat screen displays [28]. It should be noted that a user should not touch a PA model with their bare hand if haptic feedback is provided through a separate device because the illusion would be broken. However, the visual effect of the physicality that a PA model gives to computer graphics is still compelling.

The results also suggest that developing technology to overcome the projection problems may be useful when a PA model represents a ‘visual’ object. This could be achieved by using a back-projected PA model, however there are some practical issues that need to be considered. Firstly, back-projected PA models can only be constructed for a limited range of shapes because the projection needs to be directed from behind. Secondly, the projection has to travel through the PA model, which means the material the PA model is made from is important. These factors mean that the construction is more complex and probably best suited to ‘one-off’ installations, such as a museum display. An alternative to using a back-projected PA model is to track the users’ hands and ‘turn off’ the pixels that would be projected onto them. This technique has already been developed for eliminating the shadows cast by people using flat projection screens [29]. Although this does not overcome the shadow problem, the results suggest that it is the projected image on a user’s hand that reduces object-presence, as opposed to the shadows (section 3.2.10).

The results also have implications for the design of other types of computer generated displays. The results suggest that the object a display is representing should be taken into consideration when predicting the effect adding feedback to different modalities will have on object-presence/presence. However, the finding that participants always noticed how a PA model felt to touch suggests that when designing a haptic device, one cannot rely on people not noticing any inconsistencies in haptic feedback.

Finally, considering the results reported in this paper together with the results from previous research, predictions can be made about the effect of adding feedback to different sensory modalities to different types of computer-generated displays. Computer-generated displays range from *‘non-realistic’* to *‘realistic’*, where realism is determined by the naturalness and unintrusiveness of the equipment, in addition to the fidelity of the graphics. A PA model is a type of realistic computer-generated display, whereas displays such as a head-mounted-display may be considered to be non-realistic because the user is required to wear the equipment. Sensory feedback can range from being *‘basic’* to *‘advanced’*. ‘Basic’ feedback is when feedback is only provided for one aspect of the environment, for example the PA models that felt incorrect to touch provided ‘basic’ haptic feedback for shape. Whereas ‘advanced’ haptic feedback is when feedback is provided for several aspects of the environment, for example shape and texture. Previous research has shown that ‘basic’ haptic feedback increases presence when added to a non-realistic computer-generated display e.g. [30]. However, the experiments reported in this paper found ‘basic’ haptic feedback reduced object-presence. This suggests the addition of ‘basic’ haptic feedback to a ‘realistic’ computer-generated display will reduce object-presence.

This argument supports Mori’s ‘uncanny valley’ hypothesis, which predicts that the believability of a simulation increases as its fidelity increases, until it reaches a point where only the differences with the real world are noticed, and hence believability decreases [31]. Whilst this hypothesis originally comes from the field of robotics, it has recently been applied to virtual environments, e.g. [32]. With regards to haptic feedback, Mori theorized that a person can accept a realistic looking android as being human when they look at it, however when they touch the androids ‘skin’ and find it to be cold, it becomes very unrealistic and ‘horrific’ [31]. Thus it seems likely that adding ‘basic’ haptic feedback to a ‘realistic’ display will decrease a user’s sense of presence/object-presence because they will only notice how it differs from the real world. For example, a person viewing an extremely realistic looking cushion ‘placed’ on a real chair through a light-weight unintrusive Augmented Reality display, may feel that they are viewing a real cushion. However, if only ‘basic’ haptic feedback is provided, for example it feels hard and solid as opposed to feeling soft, when the user ‘touches’ it the sense that they are perceiving a real cushion may disappear. This suggests the assumption that the addition of feedback to extra sensory modalities always increases presence is flawed, and caution needs to be taken when considering the value of adding such feedback to ‘realistic’ computer-generated displays. This will become more important as displays become more ‘realistic’.

Conclusion

Currently Projection Augmented models are nearly all front-projected, and do not provide haptic feedback for material properties and hence feel incorrect to touch. This research compared the effect the projection problems and incorrect haptic feedback for material properties have on a user's sense of object-presence. It was found that overall for both PA models that represent 'visual' objects and those that represent 'haptic' objects, the incorrect haptic feedback for material properties is always the most important factor. However, the results also indicate that the projection problems are more important when a PA model represents a 'visual' object, and the providing correct haptic feedback is more important when it represents a 'haptic' object. Suggestions as to how technology could be developed to overcome these problems, and the implications the results have for other displays were discussed.

References

[1] B. Stevens, J. Jerrams-Smith, D. Heathcote, D. Callear. Putting the Virtual into Reality: Assessing Object-Presence with Projection-Augmented Models. *Presence: Teleoperators and Virtual Environments*. 11, 79-92. 2002.

[2] E. Bennett, B. Stevens, B. PARIS: Interacting with a Projection Augmented model using a mouse and a Tangible User Interface. In *Proceedings of British HCI group annual conference, volume 2*. 2004.

[3] D. Clark, R. McKeon, R. Marciano, M. Bailey. Rear-Projecting Virtual Data onto Physical Terrain: An Exercise in Two Senses Being Better Than One. In *Proceedings of IEEE Visualization*. IEEE Press, 451-454. 1998.

[4] O. Bimber, A. Emmerling, T. Kelmmer, Embedded Entertainment with Smart Projectors. *IEEE Computer (cover feature)*, 38, 1, 48-55. 2005.

[5] S. Hirooka, H. Saito. Virtual Display System Using Video Projector onto Real Object Surface. In *Proceedings of the 14th International Conference on Artificial Reality and Telexistence* (ICAT). 305-310. 2004.

[6] J. Verlinden, A. de Smit, A. Peeters, M. van Gelderen. Development of a flexible augmented prototyping system, *Journal of WSCG*, Vol. 11, No. 3, 1-1. 2003.

[7] T. Nam, W. Lee. Integrating hardware and software: augmented reality based prototyping method for digital products. In *Proceedings of CHI*. 957-957. 2003.

[8] J. Lee, P. Dietz, D. Maynes-Aminzade, R. Raskar, S. Hudson. Automatic Projector Calibration with Embedded Light Sensors. In *Proceedings of UIST*. ACM press. 2004.

[9] R. Raskar, G. Welch, K. Low, D. Bandyopadhyay. Shader Lamps: Animating Real Objects with Image-Based Illuminations. In *Proceedings of the 12th Eurographics Workshop on Rendering*. 2001.

[10] C. Ratti, Y. Wang, A. Biderman, B. Piper, H. Ishii. Phoxel-Space: an Interface for Exploring Volumetric Data with Physical Voxels. In *Proceeding of Designing Interactive Systems*. 2004.

[11] H. Zhu, W. Book. Speed control and position estimation of small hydraulic cylinders for Digital Clay. In *Proceedings of Symposium of Flexible Automation*. 2004.

[12] D. Dietz, R. Raskar, S. Mihelic-Booth, J. van Baar, J., K. Wittenburg, B. Knep. Multi-Projectors and Implicit Interaction in Persuasive Public Displays. In *Proceedings of Advanced Visual Interfaces*. ACM Press, 209-217. 2004.

[13] E. Bennett, B. Stevens, B. The effect that touching a Projection Augmented model has on object-presence. In *Proceedings of IEEE Conference of Visualization*. 2005.

[14] N. Sugano, H. Kato, K. Tachibana. The Effects of Shadow Representation of Virtual Objects in Augmented. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality*. 2003.

[15] J. Summet, G. Abowd, G. Corso, J. Rehg. Rear Projection: Do Shadows Matter? In *Proceedings of CHI*. 2005.

[16] R. Klatzky, S. Lederman. Touch. In: Healy A, Proctor R, eds. *Handbook of Psychology, volume 4: Experimental Psychology*. 147-176. 2003.

[17] M. Ernst, M. H. Bülthoff, H. Merging the senses into a robust percept. *Trends in Cognitive Sciences*. Elsevier, 8(4):162-169. 2004.

[18] M. Taylor, S. Lederman, R. Gibson. Tactual perception of texture. In E. Carterette & M. Friedman (Eds.) Handbook of perception (Vol. 3): *Biology of perceptual systems*, Academic Press, New York. 1973.

[19] F. Biocca, Y. Inoue, H. Polinsky, A.. Lee, A.. Tang. Visual cues and virtual touch: Role of visual stimuli and intersensory integration in cross-modal haptic illusions and the sense of presence. In *Proceedings of Presence 2002*.

[20] C. Cinel, G. Humphreys, R. Poli. Cross-modal illusory conjunctions between vision and touch. *Journal of Experimental Psychology – Human Perception and Performance*. 28(5), 1243-66. 2002.

[21] D. Nunez, and E. Blake. Cognitive presence as a unified concept of virtual reality effectiveness. In P*roceedings of ACM Afrigraph 2001: 1st International Conference on Computer Graphics, Virtual Reality and Visualization in Africa*. 115-118. 2001.

[22] E. Brunner, and M. Puri. Nonparametric Methods in Factorial Designs. *Statistical Papers*. 42, 1-52. 2001.

[23] W. Conover, and R. Iman. Rank transformations as a bridge between parametric and nonparametric statistics. *American Statistician*. 35:124-129. 1981.

[24] T. Greer, and W. Dunlap. Analysis of variance with ipsative measures. *Psychological Methods*, 2, 200-207. 1997.

[25] A. Mack, I. Rock. Inattentional blindness. MIT Press. 2000.

[26] M. Fiorentino, A. Uva, G. Monno. The SenStylus: A Novel Rumble-Feedback Pen Device for CAD Application in Virtual Reality. In Proceedings of *International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision*. 2005.

[27] M. Okamura, M. Hage, M. Cutkosky, J. Dennerlein, Improving Reality-Based Models for Vibration Feedback. *American Society of Mechanical Engineers, Dynamic Systems and Control Division*. Vol 69(2), 1117-1124. 2000.

[28] A. Lécuyer, J. Burkhardt, and L. Etienne. Feeling Bumps and Holes without a Haptic Interface: the Perception of Pseudo-Haptic Textures. In *Proceedings of SIGCHI,* 2004.

[29] C. Jaynes, S. Webb, R. Steel, M. Brown, W, Seales. Dynamic shadow removal from front projected displays. In *Proceedings of Visualization*, IEEE, 175-182. 2001.

[30] H. Hoffman, A. Hollander, K. Schroder, S. Rousseau, T. Furness. Physically touching, and tasting virtual objects enhances the realism of virtual experiences. In *Proceedings of IEEE VRAIS '98 Virtual Reality annual International Symposium*. 1998.

[31] J, Reichardt. *Robots: Fact, fiction and prediction*. Thames and Hudson Ltd. 1978.

[32] V. Vinayagamoorthy, A.Brogni, M. Gillies, M. Slater, A. Steed. An Investigation of presence response across variations in visual realism. In *Proceedings of 7th Annual International Workshop on Presence*. 2004.

Difficulties Using Passive Haptic Augmentation in the Interaction within a Virtual Environment

R. Viciana-Abad, A. Reyes-Lecuona, F.J. Cañadas-Quesada
Department of Electronic Technology
University of Málaga
ETSI Telecomunicación, Campus de Teatinos, CP. 29071
Málaga, Spain.
{ viciana@uma.es, areyes@uma.es, fjcq@uma.es }

Abstract

In this paper, the evaluation of an interaction technique based on the metaphor of the natural hand in a virtual environment (VE) is presented. The aim of this study is the analysis of how the inclusion of passive haptic feedback affects the interaction within a VE.

With this purpose an experiment with 18 subjects was conducted. The pilot design of this experiment and the implemented system used as a testbed are described.

The evaluation of this interaction has been developed taking into account both, objective and subjective factors. A pilot experiment was conducted to study the relationship among haptic feedback, presence and task performance, and the obtained results are discussed.

A new objective estimation of presence is also presented.

1. Introduction

Applications based on immersive VE are complex because of the interaction within the environment. Over recent years, there has been some research into 3D interaction aimed at the development of new interaction techniques and at the study of their evaluation. Moreover, the sense of presence has proved to be highly influenced by interaction mechanisms [1].

Lombard [2] interprets presence as “a perceptual illusion of non-mediation”; presence is what happens when the participant forgets that his perceptions are mediated by technology. In this sense the effect the implemented passive touch mechanism has on the illusion of non-mediation is evaluated within this testbed.

Many research studies [3] suggest that multimodal interaction compensates some constraints of interaction mechanisms. In this sense the haptic modality is being included in a wide variety of forms in Virtual Reality systems and via different devices. On the other hand, multimodality has to be carefully used because mismatches between the different sensorial sources can lead to negative effects for the user [4] [5]. Furthermore, the economical cost of devices that provide force feedback is sometimes a drawback.

In this paper we propose the use of passive force feedback as an alternative to complex, active devices for some specific applications. Moreover, the constraints of using this passive force feedback instead of active devices are discussed.

The performance assessment of the interaction techniques is difficult, mainly because its definition is unclear. A possibility is the measurement of the task completion time, the accuracy or the error rate. Nevertheless, certain applications based on VE usually treat a broad definition of performance, in which cybersickness or presence can be considered [6]. So, in this paper we propose a testbed to evaluate a passive force feedback mechanism by measuring the task performance and its relationship with presence.

In section 2, a discussion of the prior work is presented. The testbed description and the experiment design are shown in sections 3 and 4. In section 5, the results obtained from the experiment are presented. We conclude, in sections 6 and 7, with a discussion and some conclusions about the results. Finally, in section 8 we describe some ideas about further research.

2. Prior work

In recent years, a number of researchers ([7] [8]) have explored the use of new interaction techniques to enhance human performance, using objective metrics.

The use of haptic in a VE is implemented in several systems. McLean [9] discusses the use of haptic feedback as a design element for human computer interaction. Moreover different investigations measure the effectiveness of passive haptic feedback by objective and subjective metrics. Meehan [10] uses the concept of passive haptics to elicit presence. Rossember [11] shows, in a pilot study, that both active and passive force feedback can be effective in decreasing the task completion time. Hoffman [12] provides the subjective analysis of a technique based on tactile augmentation.

Furthermore, regarding the relationship between interactivity and presence and its consequences, some studies have determined that interactivity of VEs is an important cause of presence [13].

3. The testbed system

The system used as a testbed reproduces a virtual version of the “Simon says” game (Figure 1). This game is

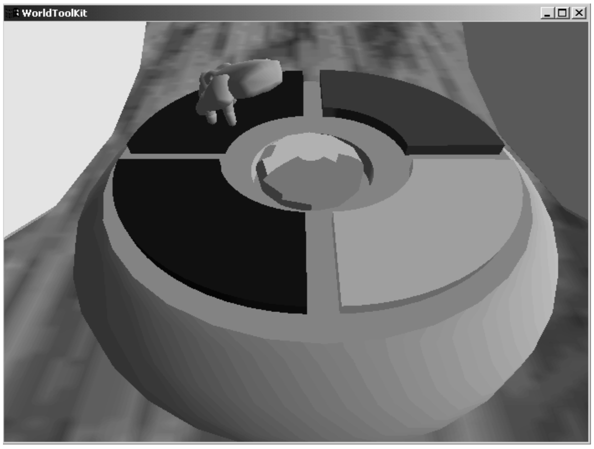


Figure 1 “Simon says” game shown to the subjects

a simple device that consists of four differently coloured buttons.

The system shows a sequence (by lighting the buttons and emitting a sound) and the user must then try to reproduce the sequence correctly. When the reproduced sequence is correct, the system increases the sequence length by adding one new step. Hence, the task grows in complexity as the sequence length increases. When the sequence is not well reproduced an error sound is emitted and two lateral plates (see Figure 1) are suddenly closed, grabbing the user’s virtual hand.

The interaction with the virtual game consists in pushing buttons. Therefore, it is simply a selection task of objects within a close range. This selection is implemented as a natural mechanism by merely touching the buttons with the fingers of the virtual hand.

To do this, we need to track the position of the user’s hand with a tracker (Flock of Bird by Ascension) and a VR glove (cyberglove by VTi). Furthermore the user sees the VE through a HMD (Head Mounted Display (VR8 by IIS)) with another tracker attached in order to sense the orientation of the head.

The system detects when a button is pressed by testing when an intersection between one of the user’s fingers and a button occurs. This intersection is checked in two ways: by testing the fingertips bounding-box, and with a ray originating from the fingertip and normal to the button surface. The button goes off when all the fingertips are removed and separated by a minimum distance threshold.

The passive force feedback is implemented in the system by means of a real surface placed under the participant’s hand. Furthermore, the tracking control system is calibrated in such a way that the real fingers touch the real surface when the virtual hand presses a button. This calibration procedure includes a slight rotation of the reference system. This allows us to reduce certain distortion of the magnetic tracker attached to the user’s hand and match the real horizontal plane with the virtual one.

4. Pilot experiment

We have conducted a pilot experiment to explore the influence of using passive force feedback in the task performance and presence achieved during the interaction with the testbed system.

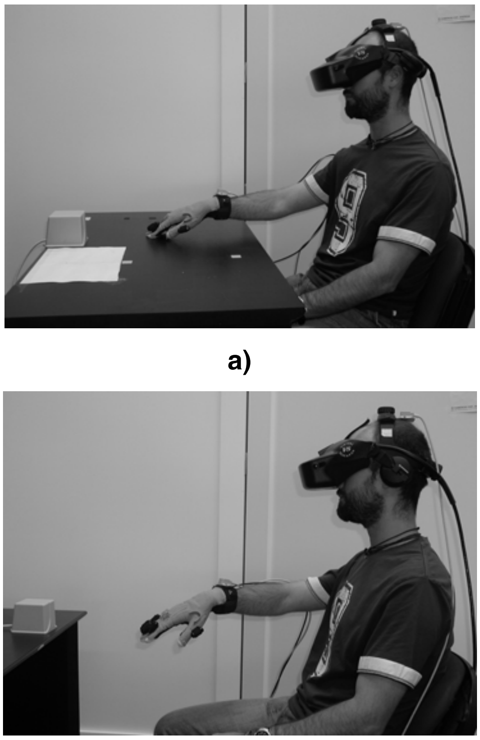
4.1. Participants

Eighteen subjects (10 males, 8 females) participated in the experiment. All of them were undergraduate first year telecommunication engineering students at the University of Málaga, aged between 17 and 19. No reward was given to them for their collaboration.

4.2. Experimental conditions and procedures

In this experiment, the independent variable was the existence or absence of passive haptic feedback when the participant presses a button. The experiment has a within-subject design. This means that every subject interacts with the system under two conditions: “haptic feedback” (FB) and “no haptic feedback” (NFB) (See Figure 2). Moreover, to eliminate the possible effect the order of the two conditions has, a counterbalanced design was made. Thus, the participants were randomly assigned into two groups. In one group, FB condition was used before NFB condition (FB-NFB), and in the other group the opposite order was used (NFB-FB).

The dependent variables considered in this experiment are the subjects’ sense of presence within the VE and the task performance when interacting with the system.



b)

Figure 2 a) FB condition b) NFB condition

The experiment took place in a research laboratory. Upon arrival, participants completed consent forms and then they received all the task instructions. Every trial lasted 6 minutes; the first two minutes being devoted to training. The difference between these two periods is that in the training phase, there is no *virtual grabbing* of the hand when an error is committed.

4.3. Measurement mechanisms

The sense of presence was operationalized by means of a subjective measurement based on questionnaires, and an objective one, based on user behaviour. The task performance was operationalized via some different measurements.

4.3.1. Presence measurements. The subjective measurement of presence was calculated by using two presence questionnaires; Presence Questionnaire (PQ), proposed by Witmer et al. [14], and the questionnaire proposed by Slater et al. [15].

In order to evaluate the subject behaviour when an error is made we record his/her hand position, for two seconds from the moment this error is made. With this data, a two-dimensional graph can be plotted representing the averaged trajectory of the hand. In order to compute this average, we consider that the hand is in the origin of coordinates (0,0,0) when an error is committed. Then, the average position for each time t after an error is computed as follows:

$$\left(\bar{X}(t),\bar{Y}(t),\bar{Z}(t)\right)=\left(\frac{\sum_{i=1}^{N}x_i(t)}{N},\frac{\sum_{i=1}^{N}y_i(t)}{N},\frac{\sum_{i=1}^{N}z_i(t)}{N}\right)$$

where $x_i(t)$, $y_i(t)$ and $z(t)$ are x, y, z coordinates respectively of the hand at the time t after the i-th error is committed. N is the total amount of errors. A linear interpolation is used to compute the average trajectory in certain temporal positions where no data is recorded.

So, every time an error occurs, two seconds of hand position are recorded, and we can follow the evolution over time of the user's hand movement. Note that following an error the game triggers the closing of the lateral plates and a sound is emitted. Therefore, differences in the trajectories made by the hand in response to this virtual event are expected to be related to different levels of presence.

4.3.2. Task performance measurements. During the trials, the score, the number of errors committed during the game, the spurious actions and the elapsed time between button pressings are recorded. The score is considered as the maximum sequence length reached by the user during the game. Spurious actions are evaluated registering the number of times the central button in the game is pressed when the user is trying to reproduce a sequence.

Furthermore, the subjects were asked how long they thought the trials lasted. The subjective estimation of time is considered to be an indication of the difficulty related to the provided interaction mechanism [16].

5. Results

5.1. Presence measurements

Regarding the sense of presence, the presence factor proposed by Slater et al. (**SF** in Figure 3) showed a slight difference between the two conditions within both groups. Moreover no differences were noted when analyzing the participants' answers without considering the order.

The PQ questionnaire showed certain significant differences between the two experimental conditions. The factors evaluated were: Presence **P**, Control/Involvement **C**/I, Natural **N**, Interface Quality **IQ**, Auditory **A**, Haptic **H** and Resolution **R**.

No dependence to the order of the two conditions (FB/NFB) was found for any of these factors. So, the difference in the factors between the two conditions in the 18 subjects is analysed, without taking into account the order.

In Figure 3, differences in the mean factors for each condition and their significance are shown. These factor values are always higher in the NFB condition in the two groups except in R (with no significance).

In Figure 4, the averaged trajectories of the users' hands in three dimensions (a, b) and their projections in the horizontal plane (c, d) under the two conditions are shown.

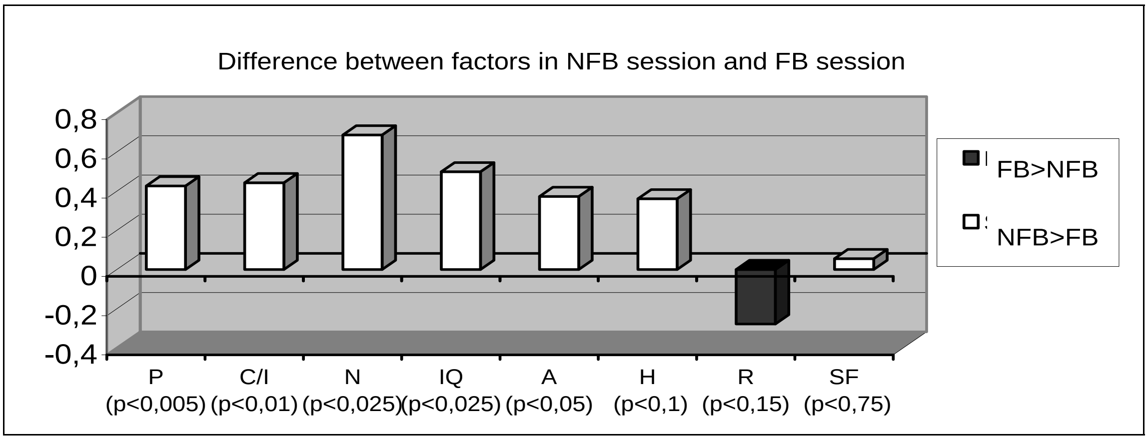
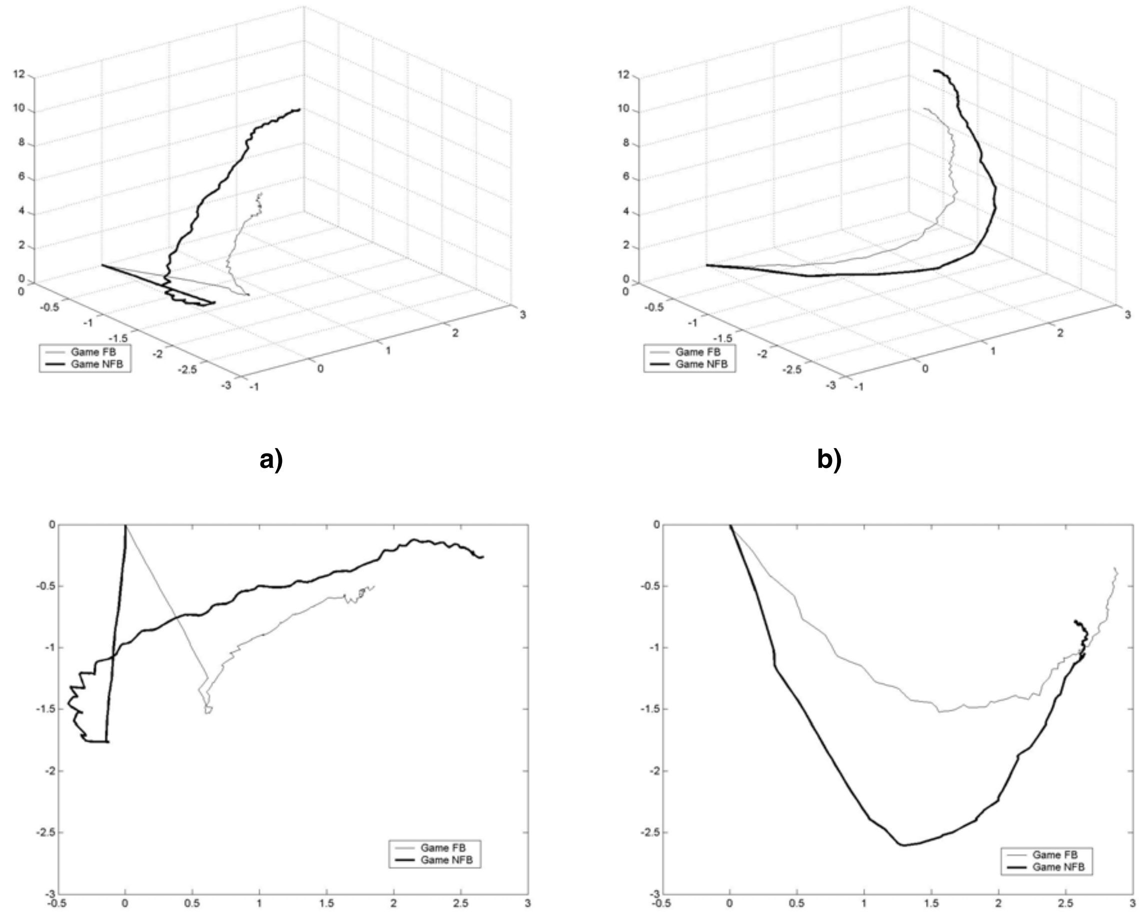


Figure 3 Differences found for each factor



a) b)

c) d)

Figure 4 a) c) Averaged trajectory of the users' hand during two seconds following an error. b) d) Averaged trajectory of the users' hand during two seconds from when the highest scores are reached just before an error is made.

Subfigures a) and c) show the average trajectories during two seconds following an error and its consequences. Subfigures b) and d) show the average trajectories during two seconds after users reach their highest score just before an error is made. When no error is made the graphics show the normal movement needed in order to clearly view the next sequence. So, they move their hand back closer to their bodies. When an error occurs this movement is the response to the clashing plates. As can be seen in this figure, a sharper change in the users' hand trajectory occurs when an error is made. In both cases the users' hand is moved over a wider range in NFB condition.

5.2. Task performance measurements

The highest score is achieved within the NFB condition. The difference in the length of the sequence between the two conditions on average is 1.72 ($p<0.025$).

Regarding spurious actions, a dependence with the order of administration of the two conditions was found. Better values are taken in the first condition, whether being FB or NFB, (difference of 2.83 $p<0.025$).

The average time elapsed between button pressings, is shorter in the NFB condition. However, it is a small non-significant difference of 166 ms.

The subjective estimation of the average time spent in each condition, was higher in FB condition. These averaged values are 6.27 min. (NFB) and 7.35 min (FB), but no significant differences were found ($p <0.25$).

6. Discussion

In this paper, a study on how passive haptic feedback affects the sense of presence and task performance within a VE is presented. In accordance with previous works, we expected haptic feedback to enhance the sense of presence and task performance [10] [11] [17].

However, our results show that, surprisingly, this passive haptic feedback diminishes the sense of presence and task performance. This could be explained by the fact that slight mismatches were detected between the position of the virtual object and the prop arranged to provide the passive haptic feedback. These mismatches can hinder the interaction because the rigid surface which can become an obstacle. In these situations, passive haptic feedback produces a sense of mediation that decreases the sense of presence. This idea is in accordance with the subjective estimation of time reported by the subjects. Subjects in the FB session reported a longer time, although the experiment duration was the same for both conditions. According to some studies [16] that relate higher time estimation with interaction difficulties, it seems that the interaction task in FB condition was more complex.

Although passive haptic feedback presents some advantages over active ([18], [19]), special care must be taken with the spatial synchronism. This kind of non-intelligent feedback might become an obstacle to the interaction when slight mismatches in that synchronism are present.

7. Conclusions

The presented testbed and pilot study have shown the importance of the spatial synchronism between real and virtual worlds for interaction. These findings indicate that further research should include techniques that improve the spatial synchronism.

We have also proposed a new objective technique to estimate presence, based on the users' reaction when an event takes place (in this case the clashing plates) via detecting changes in the hand trajectory. Our results indicate a relationship between the objective and subjective measures of presence, based on questionnaires. Furthermore, in this interaction experiment, task performance is also related to presence. In both cases, it seems that the sense of presence of the subject is higher in the NFB condition. Moreover better performance is found in the NFB condition considering the score, precision and elapsed time between button pressings.

The experiment developed for this research shows how the improvement that is expected by providing a new source of sensorial information might become a new interaction difficulty. Nevertheless, we still think that haptic feedback should improve the interaction, but we suggest that passive haptic feedback presents difficulties of implementation regarding spatial matching. Moreover, adding a new source of information requires a major effort in order for this new source to coherently join with the other sensorial sources present. In this sense, passive haptic mechanisms, whilst easier to supply than active ones, require more effort in order to overcome the lack of accuracy derived from tracking systems and the virtual reality glove.

One of the major problem sources is that there are individual differences among users. It is difficult to provide an interaction mechanism appropriate for each individual user, with different hands and different interaction and cognitive styles.

8. Further works

Further research should include techniques that improve the spatial synchronism between real objects and virtual ones. Due to the lack of accuracy of the tracking system used for the fingers (virtual reality glove) and the hand (tracker), the passive haptic feedback should be made via using a soft surface. This would provide a certain error margin that could facilitate the interaction between user and virtual environment

In order to overcome individual differences, some facilities should be provided allowing the configuration of the virtual hand in such a way that it is adapted to some physical features of the real hand. In addition, the calibration procedures should be improved.

Acknowledgements

This work has been partially supported by Spanish Ministry of Education & Science and ERDF funds through the project TIC2002-4348-C20-01.

References

[1] W.I. Jsselsteijn. Elements of a multi-level theory of presence: Phenomenology, mental processing and neural correlates. In *Proceedings of PRESENCE 2002*, 245-259. October 2002.

[2] M. Lombard, T. Ditton. At the heart of it all: the concept of presence. *Journal of Computer-Mediated Communication*, 3(2). September, 1997.

[3] F. Biocca, J. Kin, Y. Choi. Visual touch in virtual environments: an exploratory study of presence, multimodal interfaces, and cross-modal sensory illusions. Pre*sence: Teleoperators and Virtual Environments*, 10 (3), 247-265. 2001.

[4] F. Biocca. Will simulator sickness slow down the diffusion of virtual environment technology?. *Presence: Teleoperators and Virtual Environments*, 1 (3), 334-343. 1992.

[5] K. Stanney. Aftereffects and sense of presence in virtual environments: Formulation of a research and development agenda. *International Journal of Human-Computer Interaction*, 10 (2), 135–187. 1998.

[6] R.S., Kalawsky, S.T. Bee, S.P. Nee. Human factors evaluation techniques to aid understanding of virtual interfaces. *BT Technology Journal*, 17 (1), 128-141, 1999.

[7] S. Poupyrev, M. Billinghurst, T. Ichikawa. Egocentric object manipulation in virtual environments: empirical evaluation of interaction techniques. In *Proceedings EUROGRAPHICS'98*, 17 (3) , 41-52. 1998.

[8] D.A. Bowman, D.B. Johnson, L.F. Hodges. Testbed evaluation of virtual environment interaction techniques. *Presence: Teleoperators and Virtual Environments*, 10, 75-95. 2001.

[9] K.E. MacLean. Designing with haptic feedback. In *Proceedings of IEEE Robotics and Automation*, 783-788. 2000.

[10] M.Meehan, B. Insko, M.Whitton, F.P. Brooks. Physiological measures of presence in virtual environments. In *Proceedings of 4th International Workshop on Presence*. May 2001.

[11] L. Rosenberg, S. Brave. Using force feedback to enhance human performance in graphical user interfaces. In *Proceedings of CHI 96*, 291-292. 1996.

[12] H.G. Hoffman, A. Hollandera, K. Schrodera, S. Rousseaua, T. Furnessn. Physically touching and tasting virtual objects

enhances the realism of virtual experiences. In *Proceedings Virtual Reality* 3, 226-234. 1998.
[13] M.J. Schuemie, C.A.P.G. van der Mast. Presence: Interacting in VR?. In *Proceedings of Twente Workshop on Language Technology* 15, 213-217. 1999.
[14] B.G. Witmer, M.J. Singer. Measuring presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoperators and Virtual Environments*, 7(3), 225–240. 1998.
[15] D. Nunez. A connectionist explanation of presence in virtual environments. M. Phil Thesis. Dep. of Computer Science, University of Cape Town. 2003.
[16] A. Haston, A. Kingstone. Time estimation: The effect of cortically mediated attention. *Brain and Cognition*, 55 (2), 286-289. 2004.
[17] E.L. Sallnäs, K. Rassmus-Gröhn, C. Sjöström. Supporting Presence in Collaborative Environments by Haptic Force Feedback. In *Proceedings of CHI 2000*, 7 (4), 461-476. December 2000.
[18] B.E. Insko. Passive haptics significantly enhances virtual environments. Doctoral Thesis. Department of Computing Science, UNC Chape0. 2001.
[19] M.R. McGee. Investigating a multimodal solution for improving force feedback generated textures. Doctoral Thesis. Department of Computing Science, University of Glasgow. 2003.

The Value of Reaction-Time Measures in Presence Research: Empirical Findings and Future Perspectives

Christoph Klimmt[1], Tilo Hartmann[1], Andre Gysbers[1] and Peter Vorderer[2]
[1]Department of Journalism & Communication Research, Hanover University of Music and Drama
[2]Annenberg School for Communication, University of Southern California, Los Angeles

Abstract

A series of experiments (N = 168) was conducted to test the capacity of Secondary Task Reaction Times (STRT) for Presence measurement. Based on recent theories, possible connections between reaction times and Presence were examined in users of a hypertext, a film, and a virtual environment that used the same visual materials. A Presence questionnaire was employed as comparative measure. Findings indicate rather unsystematic and weak convergence between STRT and subjective measures of Presence or underlying processes. A modified STRT paradigm for Presence research is suggested.

Keywords--- **Spatial Presence, secondary task reaction times, measurement, methodology, objective measure, experiment.**

1. Introduction

The progress of Presence research depends on both theoretical and methodological advances. Measuring Presence experiences through reliable and valid indicators produced by practical, robust, and efficient methods is a precondition to resolve many research questions and to improve the user-centered design of new Presence applications based on empirical data.

Today, a large variety of methods and instruments to measure Presence is available [1]. However, many of them have not been evaluated systematically. We simply do not know enough about most of the available questionnaires, scales, and apparative procedures in respect to their reliability, validity, and practicability. Without such systematic inquiry in the value of specific methods, the research community is in danger to rely on problematic or ineffective methodological grounds.

This paper introduces one piece of such methodological evaluation. It presents results from three experiments that tested the value and usefulness of one specific approach to measure (Spatial) Presence, which is entitled Secondary Task Reaction Times (STRT).

2. Secondary Task Reaction Times and Presence Research

STRT has been employed as a measure of *attention* by psychologists for a long time. It is based on theoretical assumptions about the limitation of an individual's cognitive capacities [2]. From this perspective, people constantly distribute their perceptual and cognitive resources across different modalities. The more resources are allocated to one channel of input, the less resources remain available for other channels. Secondary task measures are constructed upon the idea that the more attention an individual devotes to a certain activity or task (the so-called primary task), the less attention 'is left' for alternative activities (i. e., secondary tasks), and the more time the organism will need to accomplish such alternative activities. Consequently, the empirical information produced by this methodology is the response time of a message receiver to a secondary input that does not belong to the actual message. The longer the response to this input takes, the more attention is devoted to the primary task, that is, the message being received by the subject [3]. Specifically, STRT are considered to indicate the amount of resources a subject is allocating to encode (as opposed to memorize) a received message [4]. However, recent findings challenge this and other existing theories on attentional resource allocation and the type of resources measured by STRT reaction times [5]. In spite of those unresolved questions, STRT is in general capable to assess "attention, arousal, and involvement" ([3] p. 93) when applied in communication studies.

Several theoretical models of Presence highlight the importance of attentional processes (e.g., [6]). The MEC model of Spatial Presence [7] [8] defines attention allocation as a key step within the formation of Presence experiences. If STRT are capable to deliver process-based information on users' attention towards a virtual environment, these data would be of great interest for Presence researchers (see [9] for a similar dual-task approach to Presence measurement).

From the perspective of the MEC model [9], STRT might even offer greater opportunities to assess (components or facilitators of) Spatial Presence: The model expects Spatial Presence to occur only through additional cognitive processes that exceed mere attention allocation. These processes include the mental representation of the media space ('spatial situation model, SSM'). STRT may produce information on the complexity of such SSMs if it is applied to users of space-related media stimuli. Moreover, cognitive involvement is considered as an important facilitator of Presence by the MEC model. Higher cognitive activities (e.g., thinking, evaluating, counter-arguing) build on attentional processes and consume much cognitive capacity [9]. Variance in STRT could therefore also indicate variations in cognitive involvement, which would be an even more relevant measure for Presence researchers [7]. There is even a possibility that STRT could measure Presence experiences directly: For instance, if the arrival-departure metaphor of Presence [10] is applied, response times to a secondary signal from the spatial environment that a person has already ‘departed from’ should be much longer than response times to signals from the environment the user ‘has arrived in’.

As a conclusion, the perspective of Presence theory suggests STRT to hold interesting capabilities to assess at least important foundational processes of Spatial Presence (i.e., attention and/or cognitive involvement) or even to address the intensity pf Spatial Presence itself.

3. Method

Three experimental investigations were conducted to assess the methodological implications of STRT in the domain of Spatial Presence; we used a hypertext, a film, and a VR environment of similar narrative and visual quality.

3.1. Stimulus materials

A set of media stimuli that was based on the same narrative and visual content was produced. The intention of using several media was to paint a more complete picture of the value of STRT and to avoid dependence of findings on one specific medium. For this purpose, a hypertext environment (with mixed text and visual elements), a film, and an interactive virtual environment were produced that all displayed the same spatial environment, which was labelled “Mozart’s house of learning” [10]. Each media stimulus was experimentally varied in order to create a broad range of Presence intensities (variance), which was required to test the reagibility of STRT to variations in Presence and/or facilitator processes. The specific settings of the three experiments are portrayed in the following sections.

3.1.1. Hypertext environment. The hypertext (HT) stimulus was similar to an old-fashioned role playing game where the location of the user is described by text and/or pictures. The museum was represented by single snapshots accompanied by explanatory and descriptive text.

This HT was experimentally varied in two ways. The first manipulated feature of the HT was the ratio between text and images. One half of the experiment’s participants used a HT version that included large images and small text sections (expectably the “high Presence” condition because of more salient visuo-spatial information), whereas the other half interacted with a HT version that displayed large text areas and comparatively small images (low Presence condition). The other experimental factor was the type of navigation. One half of participants could move through the museum by selecting desired locations (floors, rooms) from drop-down menus (non-space-related navigation, low Presence condition); in contrast, the other half of participants used navigation points posited within the HT images (space-related intuitive navigation, high Presence condition). For instance, users could click a sign on a door to proceed to the next room or click on stairs to move to another level of the museum.

3.1.2. Film stimulus. The film stimulus was a non-interactive walkthrough of the museum. It was generated from the virtual reality stimulus (see 3.1.3.). Participants were placed in front of a screen and watched the virtual walk, which included all rooms of the virtual building.

To manipulate the capacity of this film to induce Spatial Presence, the field of view (FOV) covered by the screen was varied. Participants watched the film in one of two display configurations. One half of the participants viewed the film on a 21-inch computer monitor, which covered about 20 degrees of their FOV (horizontal). The other half was posited in front of a projection screen with a diagonal of about three meters, resulting in a covered FOV of approximately 61 degrees.

3.1.3. Virtual Environment. The virtual environment (VE) used in the third experiment (built with *WorldUp*) was the actual source of all visual and auditory information used in the HT and film stimuli. A large amount of exhibits such as paintings, instruments, historical musical notes and documents, as well as other details (information desks and tables, loudspeakers, benches etc.) were placed as virtual objects in the VE.

The VE was navigated through a computer mouse. Participants could use stairs to change between levels, enter any room of the museum, and perform simple interactions with different exhibition objects.

To create variance in Presence experiences, the same manipulation of the FOV as in the film study (20 versus 61 degrees horizontal, cf. 3.1.2.) was applied in the VE experiment.

3.2. STRT Procedure

In all three experiments, the same STRT procedure was applied to maximize comparability. In order to determine the specific quality of cognitive-perceptual resources that using the media environment would (not) consume, three types of probes were developed.

One type of probes addressed only the *visual* modality. A red square (about 10 x 10 cm) appeared on an additional monitor at the right side of the screen that displayed the actual media environment (HT, Film, or VE, respectively). In those studies that manipulated field of view (see 3.1.), the size of visual probes was adjusted to the size of the primary medium in order to keep the same ratio across experimental conditions (however, eccentricity of the probes was necessarily higher in the large FOV condition). The second type was an *auditory* signal (an alarm sound produced by a typical siren). It was played at a volume that pretests had found to enforce perception in spite of the auditory background of the primary medium (approximately 70 dB). The third type, finally, combined the red square and the alarm sound to form *audiovisual* probes.

From these types of probes, a unified sequence was composed (with DirectRT software by Empirisoft, 2004). The duration of the sequence was – as the exposure times to the media stimuli – seven minutes. Within this time, 13 probes (5 visual, 4 auditory, and 4 audiovisual) were 'fired'. Each participant received the same STRT sequence. A source of unsystematic variation was, however, the program's logic to *wait* for a reaction of the participant before it continued with the probe sequence.

Participants who were exposed to the STRT measure were instructed to respond as quickly as possible to any of the mentioned types of probes by pressing the 'Spacebar' button of a computer keyboard which was unrelated to using and navigating the actual media environment. The interfaces of the interactive stimuli (HT and VE, see 3.1.) only required participants' right hand, so they could keep their left hand on the response key. Variance in motor behavior that could have biased response times were thus avoided.

Boxplot inspection was used to identify extraordinarily long response times (mostly above 1.5 sec). Corrected response times were computed to three mean index variables, one for each type of probe. These variables were used throughout the analysis.

As STRT is highly obtrusive, it was reasoned that the method could affect the actual Presence experience it was intended to measure. Thus, obtrusiveness of the method was assessed by applying the STRT procedure only to half of the participants. This strategy allowed for more rigorous testing of the method's potential for Presence measurement (see 3.3.).

3.3. Comparative Measures and Analytical Strategy

To generate baseline data for comparative analyses, an ex-post Presence questionnaire was applied in addition to the STRT procedure. The scales of the MEC-SPQ [10] were employed to measure the precursor / correlate processes of Spatial Presence – attention, strength of spatial situation model (SSM), cognitive involvement, suspension of disbelief, and the two dimensions of Spatial Presence (self-location within media space and ascription of possible action to media environment: SPSL and SPPA) elaborated in the MEC model of Spatial Presence [9].

The first step of the analytic strategy was to compute ANOVAs for each experimental setup (analysis per medium) to test the effect of Presence manipulation on STRT. For comparisons, MEC-SPQ scales from all participants were also included in this analysis.

Second, only those participants who performed both STRT and the ex-post questionnaire were examined. Correlations between response time indices and MEC-SPQ scales were computed to uncover substantial covariance between objective and subjective data. This analysis was repeated for all three media.

Third, a media comparison (independently of experimental condition within medium) was performed to detect potentially similar patterns of objective and subjective data across media. For this purpose, STRT values of those participants who had performed this measure were compared to MEC-SPQ scales only from those participants who had *not* performed the STRT procedure. This way, the obtrusive effect of STRT on (Presence) experiences was expected to be uncovered.

3.4. Procedure

Participants of all three experiments were recruited from several universities of a mid-size German city. They were offered 10 € as financial compensation. In each study, participants were randomly assigned to one experimental condition (between-subject design); gender was balanced between conditions.

On arrival in the laboratory, participants were briefly informed about the procedure of the experiment and then exposed to the stimulus (hypertext, film, or VR, respectively) for seven minutes. Prior to exposure, participants of the hypertext and VR experiments received a brief instruction on how to use the mouse to navigate through the museum. Similarly, the STRT procedure was introduced to those participants who were asked to perform this additional measure. After the seven minutes of exposure, the experimenter asked the participants to fill in the MEC-SPQ. Subsequently, s/he was thanked, received the financial compensation and additional information. Overall, experimental sessions lasted between about 25 and 40 minutes (due to participants' varying reading speeds).

For the hypertext study, 79 participants were recruited. 36 of them used the HT version with drop down navigation (low Presence condition), 43 the intuitive navigation (high Presence condition); 40 people had large text sections and small images (low Presence), and 39 small text sections and large images (high Presence) on the screen. Within each excondition, at least 8 individuals performed the additional STRT measure (35 individuals overall). Another 42 individuals participated in the film experiment (21 in the small FOV and 21 in the large FOV condition), 19 of them performed the STRT procedure (10 participants in the large FOV condition). Finally, 47 students accepted to participate in the VR experiment. 25 used the small FOV version, the remaining 22 were confronted with the large FOV version. 21 of these participants performed the STRT procedure (11 in the large FOV condition). All in all, 168 students participated in the experiments, and 75 of them produced STRT data sets.

4. Results

4.1. Experimental analysis

In this section, results of two MANOVAs for each experiment are presented. The first analysis tests the effects of experimental manipulation on Presence and its precursor processes as measured by MEC-SPQ scales, based on data from all subjects; the second tests the effects of the independent variable(s) on average response times to STRT probes and is thus necessarily based only on data from the

75 people who had performed STRT procedures. Subsequently to these two analyses, findings are briefly compared. The actual discussion is left for section 5.

4.1.1. Findings from hypertext experiment. A two-factor MANOVA (type of navigation x text/image ratio on screen) was computed to analyse experimental data for the hypertext study. Dependent variables were the scales on attention, SSM, involvement, SOD, Spatial Presence / Self-Location (SPSL), and Spatial Presence / Ascription of possible actions to media space (SPPA), which were all included in the MEC-SPQ.

Findings indicate a multivariate effect of navigation type ($F(6, 70) = 2.39$, $p < .05$), but no effect of text/image ratio and no interaction between factors. Type of navigation affected both dimensions of Presence ($F = 10.25$, $p < .01$ for SPSL; $F = 7.00$, $p = .01$ for SPPA) in the hypothesized direction: SPSL was larger for intuitive space-related navigation ($M = 2.73$, $SD = .87$) than for drop-down menu navigation ($M = 2.15$; $SD = .75$); similarly, SPPA was higher for space-related navigation ($M = 2.36$. $SD = .83$) than in the drop-down menu condition ($M = 1.90$, $SD = .65$). Values of the other MEC-SPQ scales (attention, etc.) were not affected by navigation types.

The ANOVA was repeated for those participants who had performed the STRT measurement ($n = 35$). Dependent variables were the average response times to visual, auditory, and audiovisual probes. Average response time values ranged from 471 *msec* to 610 *msec* across experimental conditions and type of probes, with standard deviations between 87 and 226 *msec*.

No multivariate or univariate effects of any of the manipulated factors on any of the response time variables were observed (all *F*s < 1). In addition, the main effect of navigation type on Presence that had been observed for the complete sample did not occur in the MEC-SPQ data of the STRT subsample, which indicates the effect of the STRT procedure on questionnaire results (obstrusiveness). As a conclusion, the hypertext experiment did not reveal a pattern that would allow to link STRT values to Presence experiences.

4.1.2. Findings from the film experiment. For the film study, a one-factor MANOVA was computed that used size of FOV (small versus large) as the only independent variable and all MEC-SPQ scales (see 4.1.1.) as dependent measures. No effect of FOV on Presence experiences or any of the related variables was detected in the questionnaire data.

In the second MANOVA (that examined only those subjects who had performed the STRT procedure, $n = 19$) FOV did again not display a significant multivariate effect and did not affect any of response time values systematically.

4.1.3. Findings from VR experiment. The MANOVAs computed for the VR study were identical with the analysis of film data. Questionnaire data again did not indicate any multivarate or univariate effect of FOV on Presence or its precursors. The MANOVA that used STRT data in addition to scale values from the STRT subsample ($n = 21$) did not find a multivariate effect and only a marginally significant influence of FOV on reaction times to auditory response times ($F = 3.02$, $p < .10$). For the other two types of probes, response times remained unaffected (both *F*s < 1).

4.2. Correlational analysis

4.2.1. Findings from hypertext experiment. Relationships between STRT values and subjective data as measured by the MEC-SPQ scales were generally weak in the HT study. Most *r* remained below +/- .20 and were not significant ($n = 35$). The strongest observed correlation occurred between the attention scale and response time index for auditory probes ($r = -.42$, $p < .01$). The negative direction opposed the hypothesized positive relationship between attention to the primary medium and response time to secondary input (see 2.1.).

4.2.2. Findings from film experiment. In general, correlations ($n = 19$) between questinnaire data and STRT values were higher in the film experiment than in the hypertext study. However, most of them were again negative, which contradicted the hypothesized relationship between STRT and Presence (precursors). A stable pattern of negative correlations was observed for SPSL ($r = -.42$, -.30 and -.33 for visual, auditory, and audiovisual probes, all *ns*); similarly, all correlations between response times and SPPA and between response times and attention were negative, with slightly lower coefficients. In contrast, all correlations between STRT and involvement scale values were rather weak, but positive ($r = .09$, .17, and .20, respectively). This finding was in line with expectations, however, the negative correlations between STRT and attention as well as Presence were unexpected and puzzling.

4.2.3. Findings from VR experiment. In the VR study, correlations ($n = 21$) were generally weaker than in the film study and more similar to the results on hypertext (4.2.1.). Whereas relationships between STRT and attention were again (weak, but) negative, rather strong positive correlations were observed between STRT and involvement ($r = .25$, .36, and .50 for visual, auditory, and audiovisual probes, with only the last coffefficient reaching statistical significance, $p < .05$). These results suggest that there maybe a stable relationship between involvement and STRT.

4.3. Media Comparison

For the media comparison, experimental conditions within each medium were ignored, which is partly justified by the failure to create effective manipulations (see 4.1.). The scope of the media comparison was to find out if average subjective values and average probe response times display similar patterns across media. If, for instance, the Presence scales would reveal higher scores for VR than for hypertext, and STRT data would display the same pattern, this would indicate a general convergence between measures. To avoid obtrusiveness effects, this analysis included questionnaire data only from those subjects who had *not* performed the STRT procedures (see 3.3.). Figure 2 displays average scale values for the most important MEC-SPQ constructs (attention, involvement, SPSL, and SPPA). Media differences occurred only in SPSL, which was substantially lower in hypertext than in the other media. SPPA was higher for VR than for hypertext and film.

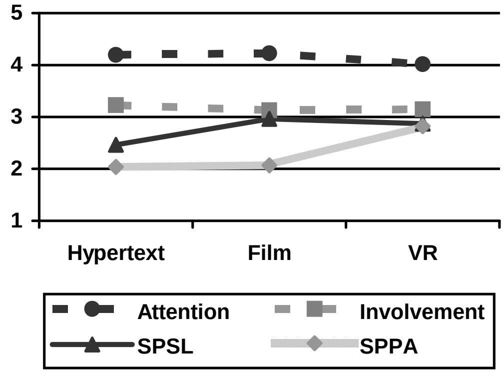


Figure 2 Average MEC-SPQ values across media (participants without STRT measurement)

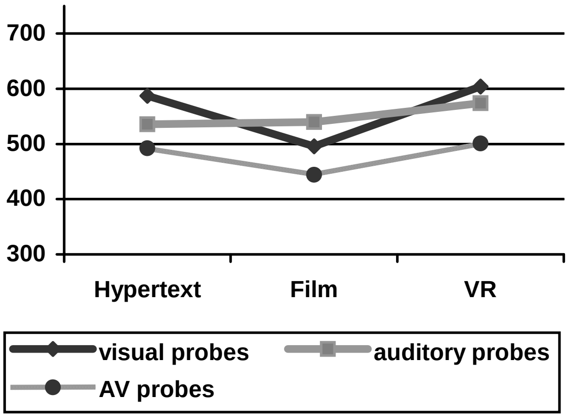


Figure 3 Average STRT values (in *msec*) across media for three probe types (see 3.4. for *n* of each medium/study)

The only interesting pattern that emerges is that response times to visual (and audiovisual) probes are faster in the film experiment than in the two other media/studies (figure 3). As film was the only non-interactive medium that was investigated, this result suggests that *visual* attention is more effectively bound by interactive media in which users have to make decisions and solve tasks (e.g., navigation) by themselves instead of merely witnessing a ready-made media product. Subjective measures do not reflect this pattern, however. In fact, the cross-media curves created by STRT data display not much congruity with the according subjective measures.

5. Discussion

Our studies produced in part unexpected covariance between STRT and questionnaire data, in part plausible connections, and mostly weak to no relationships that do not allow for a unified interpretation. Results indicate that STRT may assess involvement and, to some extent, visual attention. These conclusions would be in line with conventional STRT methodology as it is applied in TV research. Contradictory to past STRT research is the negative correlation of STRT with the attention scale (especially film and VR studies), which might be explained by the assumption that participants took the subjective attention measure as general scale of vigilance that referred to both the medium *and* to the secondary task: Highly alert people would be attentive to the medium *and* watching out for the next probe such behaviour would lead to a negative correlation between STRT and the attention scale.

Findings suggest further problems with the subjective measure of Presence and its precursors. The faster response times to visual probes in the film experiment was not reflected in subjective measures. Moreover, the expectably ‘safe’ manipulation of Presence (FOV in film and VR experiments) did not produce systematic variations in the MEC-SPQ scales.

One possible explanation for these unexpected results is that participants might have used an implicit media-specific baseline of what they would have expected to be the ‘maximum possible value’ when using a given medium. For instance, a low Presence rating made by a participant of the VR stimulus may result from that person’s consideration that still much more intense Presence experiences would have been conceivable when using a VR environment. At the same time, the low rating may, in absolute terms, still mean a much stronger Presence experience than the person would have had when using the hypertext environment (even if the person would have made a high Presence rating when using that medium). If participants have performed such relativizations when filling in the subjective measures, this would necessarily cause difficulties to identify a stable convergence between subjective measures and objective data such as STRT values, since objective data are not sensitive to such media-specific adjustments of values. Consequently, the mixed results found in the present studies should not be solely attributed to the STRT procedure.

Some additional methodological problems of the reported studies need to be addressed. One major limitation is the low power of the research design, especially due to the small number of people who actually have produced STRT values. Although the realized samples would have had sufficient sizes to detect clear and non-ambiguous patterns in subjective and objective data, they certainly do not allow for sophisticated re-analysis, e.g., to repeat analysis for several subgroups within the STRT subsample.

Moreover, the concrete STRT procedure applied in the present studies is only one possible operationalization. Alternatively, the visual probes could appear on the screen that displays the primary stimulus. In a similar fashion, alternative options for auditory probe design could be envisioned.

The majority of results illustrates the problems of STRT. Primarily its obstrusiveness [11] is a limitation in the context of Presence research, because Presence is a highly fragile experience that may be massively altered by disturbing visual or auditory signals. From this perspective, the findings do not contain indications of advantages that could compensate for the high obtrusiveness of STRT. Therefore, the main conclusion of this series of experiments is that the methodological cost/benefit ratio of STRT seems to be not very positive. However, this recommendation is only valid for conventional STRT procedures like those executed in the reported studies.

6. Outlook: Advancing STRT to „Spatial STRT“ in order to assess space-related cognition (and Presence?)

The basic idea behind the STRT paradigm is to assess the availability of attentional and/or cognitive resources that remain when a person is engaged in a certain task (e.g., media use). Thus, the conceptual target of STRT is a rather broad-defined human capacity, which may have contributed to the mixed results reported in this paper (see also [4]).

One possible improvement of STRT that may be especially useful in the context of Presence measurement is to narrow the (conceptual) focus of what STRT can measure. Instead of addressing any kind of cognitive resources through measuring response times to any kind of probe, specific processing resources might be targeted by designing special types of reaction tasks. In the context of Presence measurement, it would be interesting to assess *space*-related cognitive capacities. From the perspective of the MEC model, for instance, space-related secondary task reaction times (sSTRT) may be capable to quantify the strength or salience of a user's mental representation of the media space (SSM) or even the intensity of Spatial Presence itself.

We are currently exploring the potential use of sSTRT. This modified methodology uses spatial and non-spatial ('flat') stimuli and requires participants to decide about spatiality or non-spatiality as quickly as possible. If users' space-related processing resources are bound by the primary medium (which would be an indicator for strong SSM or even high Presence), they should need more time to make that decision (response time) and should make more mistakes (error rate), which would create two interpretable output variables of sSTRT that could be directly linked to theoretical models of Presence. At least one major experiment will be conducted to find out if the sSTRT methodology is capable to deliver results that holds greater benefits for Presence measurement than what we have found for conventional STRT.

Acknowledgements

The research presented in this paper was funded by the European Commission (project "Presence: Measurement, Effects, Conditions", IST-2001-37661). We thankfully acknowledge the Commission's support. We also thank all partners of the MEC project for their support in respect to Presence theory and measurement, as well as in regard to the development and production of stimulus materials.

References

[1] van Baren, J. & IJsselstein, W. (2004). *Compendium of Presence measures* (Online). Available: http://www.presence-research.org.

[2] Lang, A. (2000). The information processing of mediated messages: A framework for communication research. *Journal of Communication, 50,* 46-70.

[3] Basil, M. D. (1994). Secondary reaction-time measures. In A. Lang (Ed.), *Measuring psychological responses to media* (pp. 85-98). Mahwah: Lawrence Erlbaum Associates.

[4] Lang, A. & Basil, M. (1998). Attention, resource allocation, and communication research: What do secondary task reaction times measure anyway? In M. E. Rollof (Ed.), *Communication Yearbook 21* (pp. 443-474). Thousand Oaks, CA: Sage.

[5] Lang, A., Haverhals, L. M., Bradley, S. D., Park, B.-H. & Shin, M. (2003, May). *An empirical examination of the secondary task reaction times: Testing what they really measure*. Paper presented at the 53rd Annual Conference of the International Communication Associaton (ICA), May 23 - 27, 2003, San Diego.

[6] Draper, J. V., Kaber, D. B. & Usher, J. M. (1998). Telepresence. *Human Factors*, *40*(3), 354-375.

[7] Vorderer, P, Wirth, W., Saari, T., Gouveia, F. R., Biocca, F., Jäncke, F., Böcking, S., Hartmann, T., Klimmt, C., Schramm, H., Laarni, J., Ravaja, N., Gouveia, L. B., Rebeiro, N., Sacau, A., Baumgartner, T. & Jäncke , P. (2003). *Constructing Presence: Towards a two-level model of the formation of Spatial Presence*. Report to the European Commission, project Presence: MEC (IST-2001-37661). Hannover, Munich, Helsinki, Porto, Zurich.

[8] Böcking, S., Gysbers, A., Wirth, W., Klimmt, C., Hartmann, T., Schramm, H., Laarni, J., Sacau, A., & Vorderer, P. (2004). Theoretical and empirical support for distinctions between components and conditions of Spatial Presence. In M. Alcaniz & B. Rey (Eds.), *Proceedings of the VII. International Workshop on Presence – Presence 2004* (pp. 224-231). Valencia: Universidad Politécnica de Valencia.

[9] Darken, R.P., Bernatovich, D., Lawson, J., and Peterson, B. (1999). Quantitative measures of Presence in virtual environments: The roles of attention and spatial comprehension. *Cyberpsychology and Behavior, 2,*337-347.

[10] Kim, T., & Biocca, F. (1997). Telepresence via television: two dimensions of telepresence may have different connections to memory and persuasion. *Journal of Computer-Mediated Communciation, 3*(2). http://jcmc.indiana.edu/vol3/issue2/kim.html

[11] Klimmt, C., Gysbers, A., Hartmann, T., Nosper, A., Behr, K.-M. & Vorderer, P. (2005, May). *Do secondary task reaction times measure (precursors of) Spatial Presence?* Presentation at the 55th conference of the International Communication Association (ICA), May 26 – 30 2005, New York.

Session 10

23nd September, 2005

Panel Session

15.00-16.30 *Presence: Past, Present, and Future*

OMNIPRES

The panel will take a bird's eye view of the field, overview the current Presence initiative, and look ahead towards new Presence research. With views directed towards the past, present, and future of presence research, the panel will outline key presence research accomplishments, current issues, and suggest possible trends and recurring themes. The panellists will seek to engage the audience in a discussion of the current state and direction of presence research.

Presence Past
Wijnand IJsselsteijn
Human-Technology Interaction Group, Department of Technology Management, Eindhoven University of Technology, The Netherlands

Presence Present
Jonathan Freeman
Department of Psychology (i2 media research), Goldsmiths College, University of London

Presence Future
Frank Biocca
Telecommunications, Information Studies and Media Communication Arts and Sciences, Michigan State University, USA

Session 11

23nd September, 2005

Therapy

17.00-17.30 *Play Therapy Utilizing the Sony Eye Toy*
Eva Petersson and Anthony Brooks
Aalborg University Esbjerg, Denmark

17.30-17.45 *An Augmented Reality system for the treatment of acrophobia*
M. C. Juan[1], D. Pérez[1], D. Tomás[1], B. Rey[1], M. Alcañiz[1], C. Botella[2] and C. Baños[3]
[1] MedICLab, Universidad Politécnica de Valencia, Spain
[2] Departamento de Psicología Básica y Psicobiología (UJI), Spain
[3] Universidad de Valencia, Spain

17.45-18.00 *Using virtual reality to provide nutritional support to HIV+ women*
Sarah Brown[1], David Nunez[2] and Edwin Blake[1]
[1] Department of Computer Science, University of Cape Town, South Africa
[2] Department of Psychology, University of Cape Town, South Africa

Play Therapy Utilizing the Sony EyeToy ®

Anthony Lewis Brooks[1] and Eva Petersson[2],
[1]Associate Professor, [2]Assistant Professor, Aalborg University Esbjerg, Denmark
{tonybrooks@cs.aaue.dk, eva.petersson@cs.aaue.dk}

Abstract

An international collaborative explorative pilot study is detailed between hospitals in Denmark and Sweden involving rehabilitation medical staff and children where the affordable, popular and commercially available Sony Playstation 2 EyeToy® is used to investigate our goal in enquiring to the potentials of games utilizing mirrored user embodiment in therapy. Results highlight the positive aspects of gameplay and the evaluand potential in the field. Conclusions suggest a continuum where presence state is a significant interim mode toward a higher order aesthetic resonance state that we claim inherent to our interpretation of play therapy.

Keywords--- **Flow, Therapy, Training, Play.**

1. Introduction

Our hypothesis is that game playing using embodied user interaction has evaluand potentials in therapy and thus significance in quality of life research for the special needs community. A state of presence is inherent where stimulation of fantasy and imagination involves engagement and subsequent interaction with a virtual environment (VE). Once this engagement is achieved and sustained we propose that a higher order state is achievable through empowered activity toward a zone of optimized motivation (ZOOM) [1]. This is possible by using an interface to the VE that is empowering without the need for any wearable technology that is deemed encumbering or limiting for the participant. The interface data – participant motion - is mapped to control immediate feedback content that has real world physical traits of response and is interesting, enjoyable, and fun for the participant so that experience and engagement is further enhanced.

Subjective presence has predominantly been investigated in respect of optimal user state in virtual environments and has been suggested as being increased when interaction techniques are employed that permit the user to engage in whole-body movement [2].

Our findings to date indicate at the motivational potential from an enhanced state of presence achieved from game environments where the body is used as the interactive unencumbered interface [3, 4, 5, 6, 7].

1. 1. Presence and Aesthetic Resonance: as a 'sense state' continuum

We are interested in observed behaviour aspects of presence where there is evidence of only a limited body of research.

Accordingly the case is made for a continuum beyond presence that satisfies our requirement of a play therapy scenario where, from within what is termed a state of aesthetic resonance, we enquire to the potential from game systems with mirrored user embodiment by using the EyeToy®. As a result of this initial pilot enquiry we intend to reach a point from where to launch a fuller investigation with a more optimized environment, method, and analysis design.

Aesthetic Resonance (AR) is when the response to intent is so immediate and aesthetically pleasing as to make one forget the physical movement (and often effort) involved in the conveying of the intention and is in line with [4, 8].

Within targeted aesthetic resonance our strategy is to approach the same immersive engagement that occurs between a child and a computer video game that is often subject to negativity and reverse the polarity of attitude so that it is positively used to empower activities beyond the usual limits of the special needs participant through encouraging an immersed 'play' mindset rather than a 'therapy' mindset which our prior research has shown as optimal [9].

Within this set up the same information that is used as control data to the interactive feedback content is available for simultaneously performance progress monitoring.

System tailoring as a result of observations of user performance – both physiological and psychological – is opportune with related testing that supplements traditional forms of performance measurement.

This in line with our earlier approach to interaction in virtual environments with acquired brain damage patients [4, 5, 9, 11] and is related to a study concerning brain neuroplasticity and associated locomotor recovery of stroke patients that reports on users interacting with games and perceiving the activity not as exercise or therapy, but as play [10].

1.2. Play

Most play research informs about its relationship to children's cognitive development, and focuses on solitary play [12]. However, this research does not account for the totality of what is going on between children in situations of *interactive* play therapy. Our play therapy approach is activity driven and the targeted aesthetic resonant state of the user we suggest is beyond the often used all encompassing term of presence.

Significantly, others have approached presence as an activity including video games [13] - but conducted in a laboratory which we question due to the situated effect of the environment on the participants. In previous studies [1] we state that activities always are situated, which underline a complex relationship between the individual, the activity, and the environment as mutually constitutive [14]. Thus a relationship to situated presence is implied as we base our enquiry at locales of predicted use with real users. The goal being exploratory is thus implemented in a pilot study so as to define problem areas to achieve preliminary data on potential of video games in therapy.

1.3. Under used resource for therapy

With the advancement in computer vision techniques and camera advancements we claim that systems such as the EyeToy® which focus on the body as the interface are an under resourced opportunity for therapists to include into training as unlike traditional biofeedback systems specific licensing is not required as there are no attachments to the patient. The system also achieves an essential aspect of children's engagement in virtual or real worlds as within our situated interactive therapy space they are 'placed' in the midst of the experience, as in a flow state [15].

We hypothesize that tools such as the EyeToy® have potentials to decrease the physical and cognitive load in a daily physical training regime, and this is central to our concept as the child experiences a proactive multimodal state of presence that encourages an unconscious 'pushing of their limits' that they otherwise would not approach outside of the interactive framework. This supports the statement of iterative human afferent efferent neural loop closure as a result of the motivational feedback and feed-forward interaction. This process is valuable for the child's physical demands in everyday life as the pushing intensifies the child's experience of movements in practice [18].

2. Gameplaying and mastery

The investigation presented in this paper addresses the promotion of motivational feedback within empowered gameplaying activities whilst attempting at understanding motivational mechanisms. This is by analyzing the gameplaying as an action where the child's increased skills in using the video game is viewed as a process of emerged mastery [19] of their 'doings' in a form relating to cycles of action-reaction-interaction. The material of the child's action within this study is the movement as the child masters the computer game by moving the body. In Laban's [18] terminology this is described as an 'effort' and he furthermore underlines the importance of offering the child opportunities to express him- or herself through non-human directed efforts in order to keep and increase the child's immediate spontaneity in the situation (e.g. reactive content that promotes subsequent interaction from the child).

For environments to be supportive in this sense, they must engage the child in challenging ways. Even though environments provide children a sense of challenge, they have to feel that their skills meet the challenges. If there is an imbalance between the challenges and the child's skills the child will become stressed or bored. Play and exploration encourage a sense of flow (immersion in enjoyable activities) that "provides a sense of discovery, a creative feeling of transporting the person into a new reality. It pushed the person to higher levels of performance, and led to previously undreamed-of states of consciousness" [15, p.74]. Optimal experience is also described as "a sense that one's skills are adequate to cope with the challenges at hand, in a goal-directed, rule-bound action system that provides clear clues as to how well one is performing" [15, p.71].

These activities are intrinsically rewarding and the enjoyment derives from the gameplaying activity in itself, which is related to the notion of the Zone of Proximal Development in learning situations [20]. In an explorative manner the child's cycle of movements can be shown to be fluent and intense or segmented without connection.

Laban [18] defines such changes in movements as important as they indicate whether there is a presence or absence of flow from one action and state of mind to another. As such the ZOOM [1] is important in its encouragement of the child's unintentional and/or intentional explorations, without immediate goals as in play, or curious discovery, and as a foundation of evoked interest [21]. This kind of interest indicates that the state of aesthetic resonance facilitates a foundation of creative achievements.

The motivational feedback loop described in this paper is also influenced by Leont'ev's [22] description of the formation of an internal plane. We have chosen to use the term of mastery to describe such processes where emphasis is on how the child's use of the game features leads to development of certain skills rather than on internalization [20], or more generalized abilities.

Thus, gameplaying actions do not need to be conscious, as at a certain level they can be unconscious skills, which, supported by playful aspects of the game, proactively push the child's limits towards new levels of movements.

As a preliminary investigation, we attempt to understand movements according to a semiotic interplay between the child’s inner and outer world [23] and relate the understanding to presence, through which spontaneous movement engagement and intensity is assigned [18].

We compare this to Wenger’s [24] and Vygotsky’s [20] description of emergent development processes. Bigün, Petersson and Dahiya [25] characterize such processes as non-formal, where exploration and curiosity are central conditions, rather than traditional formal training conditions.

The movement cycle of the gameplaying child includes a construal of rhythm. The movement cycle is concentrated on the game’s external achievement and by moving the body to achieve the external goal the child relates the inner world to the outer. However, it is not so that every movement unifies the inner and outer worlds, there has to be a “reciprocal stimulation of the inward and outward flow of movement, pervading and animating the whole of the body” [18, p.110] in order to enhance a sense of aesthetic resonance. In this way there is a range of flow through presence, from excitement to stillness, which increases and decreases the child’s participation in the gameplaying activity.

This range embraces an orchestration of expanding bodily action in space, or, in terms of Laban [18], includes different trace forms of movements that demands continuity of gestures and it is these gestures that we analyse.

3. Method

In consequence with our interpretation of the referenced theories and to fulfil the goals of the investigation we used a triangulation of qualitative methodologies to qualitatively analyze the combined materials from the two hospitals:

- Video observations of children playing with the Keep Up EyeToy® game;
- Interviews with children and facilitators;
- Questionnaires to the facilitators involved;
- Diaries/field notes from the facilitators involved.

The subjects in the studies were 18 children (10 females and 8 males) between the ages of 5 and 12 years, mean age 7.66 years, in 20 gameplaying sessions. The children were selected by the hospitals and were well functioning. The control group was similar children from the hospitals not in sessions [5, 9, 11]. The facilitators involved at the hospital were two play therapists and three doctors.

3.1. Description of material

In 2003 Sony Computer Entertainment Inc. released the EyeToy® as a new video game series for its market leading PlayStation®2 (PS2) platform which is based upon using the player’s body movements as the interface to the game.

This controller is unique in concept as all interactions to the game are through the video window rather than through the more common handheld gamepad or joystick device. The system is thus ideal for our enquiry.

The EyeToy® game chosen for this study was called ‘Keep Up’ due to its immediate action content, built in scoring, and cross gender qualities. A monitoring system based on multiple cameras supplemented so that post session analysis was available.

3.2. Description of procedure

EyeToy® games have ‘tasks’ for the participants to accomplish. The task within this game is to keep a virtual football - with animated real-world physical properties - ‘up’ within a virtual environment.

One game sequence is limited to three balls and three minutes.

After three balls, or alternatively three minutes, the game agent turns up and gives the player negative or positive feedback related to the scores of the game. The player can increase or decrease the scores by hitting monkeys and other animated characters with the ball as the game progresses.

At both hospitals the studied activities took place in rooms that also were used for other purposes, such as staff meetings and parent information. The children were not normally playing in this room and the system had to be set up around positional markers on floor and tables.

Parents were approached about the project, informed of the goals, and were asked to give their permission on behalf of their children beforehand.

Following the parents signing their permission the children were also asked to sign their permission to participate.

The process started with positioning the child in the calibration upper torso outline on the screen and after an introduction the game was started.

The gameplaying activity was observed and video recorded by the play therapists and doctors.

After the ending of the session the children were immediately asked follow up questions concerning their experiences of the gameplaying activity.

After the end of all sessions the play therapists and doctors were asked to fill in a questionnaire concerning their own experiences.

A final interview with the play therapists and doctors was also carried out to conclude the field materials.

3.3. Description of the set up

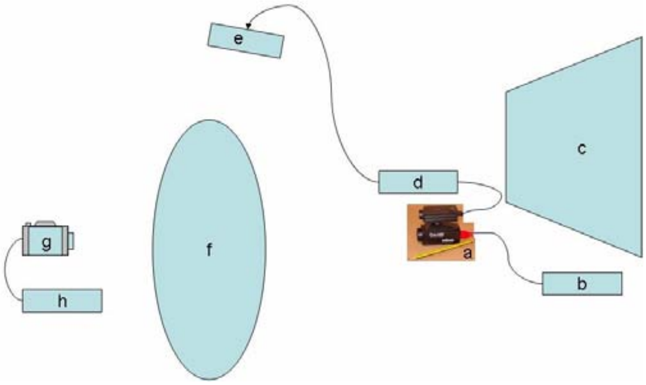


Figure 1 The set up

In previous research on camera capture as game interface [6, 10] standard TV monitors were apparently used. Our approach uses a LSD projector for large image projections approaching a 1:1 size ratio of the child (mirroring). This strategy is built upon our prior research investigations [1, 3, 4, 5, 8, 9, 11, 16] to optimize the experience. A related study is reported in the case of presence and screen size [17]. Traditional use of mirroring is used in therapy training at institutes for people with disabilities and thus our design is ‘fit appropriate’ to this context. Figure 1 (above) demonstrates the set up of the gameplay. The components included in the set up was: (a) EyeToy® camera plus front monitoring camera to capture face and body expression (b) VHS tape recorder (c) screen (d) PS2 (e) projector (f) the user space (g) rear camera to capture scene and screen (h) VHS tape recorder #2.

3.4. Description of analysis

The video recordings underwent numerous tempospatial analyses [26] where the units of analysis were the qualitatively different expressions of movement. The material attained from the sessions consisted of 36 x 1 (one) hour mini digital videos (rear and front views) – and corresponding additional backup VHS video tapes - of the 240 video games that were played by the children (n = 18) in 20 sessions at the two hospitals. Each video was digitized for the subsequent analysis; similarly, all video interviews, written notes, memos and written interviews were transcribed and transferred onto a computer workstation.

3.4.1. Manual analysis

Annotation was conducted by two coders. An initial series of four manual annotations of the video materials were conducted. These accounted for observed expressive gesture of the children (facial & body) (see Figure 2, and Appendix 4: Table 3).

Figure 2 Fully engrossed in the interaction with attention on content

In addition each video archive game and pause duration was time logged and the first, last, and best performance extracted for later analysis (example charts of three children in Appendix 1: Figure 3). Annotation of parameters of the games and pauses (*between*) before/after best and worst performance were also subject of closer analysis. An extra annotation was carried out on same child multiple sessions (n = 2) including *t* element task scores (ball 1, 2, 3).

The temporal specifics concerns rhythm as a periodic repetition and include dynamic kinetic change as well as structural patterns. Examples of temporal events are the qualities that are in play when the child affects the ball from one spot to another by swinging the body/hands or arms to and fro, which is often a challenge for those with functionality problems. The repetition of a movement develops a sense of enjoyment and engagement of the activity, which, in turn, motivates the child to continue to experience the movement. Laban [18] states that the repetition creates a memory of the experience, which is needed for new inspiration and insight to develop. More specifically the temporal data was classified into discrete units for analysis by applying the specifics of speed, intensity, and fluency of movements [18, and Efron in 26].

The spatial specifics concerns where the body moves through extended movements towards another situation in the spatial environment. Example of spatial events are the qualities that are in play when the child seeks another situation in the spatial environment, e.g. moving like jumping or leaning the body from one side of the screen to the other whereby the central area of the child's body is transported to a new position when keeping the virtual game ball up in the air. The spatial data was classified into discrete units for analysis by applying the specifics of range and intentionality of movements [18, and Efron in 26]. Alongside with these tempospatial qualities children's face expressions and utterances were analyzed.

Thus, a detailed manual multimodal analysis of the videos was realized so that:

- each video was watched twice before the detailed analysis began;
- the analysis of the first eight videos was realised twice each and the following eight videos once each;
- each minute of video was systematically analysed and transcribed into an excel flowchart in relation to the categories described above. The categories analysed represented high or low degrees of the specific movement trait. This flowchart also included analysis of a facial expression, a description of what happened on the screen (Appendix 4: Table 3);
- every category (n = 8) was analysed separately, which means that the first eight videos were watched in total 18 times each, and the remaining being watched 10 times each. Additionally the multi-sessions were annotated further four times.

3.4.2. Computer analysis

Toward a goal to amass indicators of the overall motion attributes of each child an automated low-level movement analysis was computed on the videos utilising software modules from the 'EyesWeb Gesture Processing Library' specific to the quantity and contraction aspects of the movement[1] . The data was then exported to a spread sheet for further analysis.

Our strategy for the automated computer video analysis was to supplement the manual annotations toward our overall goal in development of the methodology by (a) following a background subtraction on the source video to segment the body silhouette a Silhouette Motion Image (SMI) algorithm that is capable of detection of overall quantity, velocity and force of movement is used. Extraction of measures related to the 'temporal dynamics of movement' is computed and a threshold value slider can be adjusted according to each child's functional ability so that he or she is considered to be moving if the area of the motion image is greater than the related (to threshold) percentage of the total area [27]. The adjustment of the threshold value is achieved in real-time annotation of the videos (Appendix 2: Figure 4); (b) a contraction index (CI with range 0-1) algorithm is used with a bounding rectangle that surrounds the 2D silhouette representation of the child within the minimal possible rectangle. The CI is lower if the child has outstretched limbs compared to an image showing the limbs held close to the body where the CI approaches 1 (Appendix 2: Figure 5). Problems were apparent with the child encroaching towards the camera, and background noise. A correcting normalisation algorithm was unsuccessful in correcting the problem and thus refinement is needed [27].

4. Results

Our explorative question concerned the potential of video games in therapy and requirements toward a meaningful and optimized full investigation. Our findings present the facts that: (1) more care in the set up of the room background is needed – some videos had curtains blown with wind and people walking behind the child, (2) attire of children should contrast background – if light background and light shirt, then camera software problems occur with differentiating between child and background, (3) lighting of child/room should be optimised, (4) the system is developed for upper torso single person play but many of the children used all of their bodies, especially in kicking when the ball was lower in the screen (5) facilitators should not talk or be in line of sight. Our instructions were also interpreted differently by each hospital in so much that (1) in Sweden a time limit of 10 minutes was established for each session, (2) a long practise period was included within the Swedish ten minute period, (3) in Denmark one of the doctors also included practice periods for his children, (4) in Sweden multiple sessions were held in the same day whilst in Denmark single session per day.

4.1. Tempospatial movements

In annotating the games Start – Middle - End segmented zones were interpreted in respect of game and pause data. As expected the best performance was achieved in the end segments on an 8:15:17 ratio (even accounting for extended play boredom through no level change). The

---

[1] www.bris.ac.uk/carehere & www.eyesweb.org

shortest game ratio was 18:13:9; the longest pause ratio 16:12:12; and the shortest pause ratio 8:14.18.

These figures indicate that the virtual environment interaction with the EyeToy® met with predicted balance of performance and learning curve. Of interest within the figures was the fact that in most cases the best performance was preceded by the child's shortest pause and that following the best game it was often the case that the next two games declined in performance drastically. This matches the manual annotation where the activity (play) peaks and in most cases the emotional expression from face and body gesture before and after relates.

A general result was the faces of the children giving a defined statement of their presence (and aesthetic resonance) in the interaction with the content of the game, which was mostly pleasing and a challenge for their skills.

The detailed analysis showed a connection between tempospatial movements and aesthetic resonance through a correlation between the categories of intensity and intentionality. When there was a high, medium, or low degree of movement intensity, the same degree was always appearing in the category of intentionality of movements. Furthermore, there was a higher degrees of aesthetic resonance related to spatial movements than to temporal as the categories of range, intentionality, and shifts had high or medium degree of movements. The categories of speed and fluency, on the other hand, had low or medium degrees of movements, while the degree of intensity in temporal movements was high (Appendix 3, table 2). The computed data analysis supported the manual analysis so as to indicate higher or lower degrees of quantity of movements (QOM) and through the threshold of motion and non motion segmentation (Appendix 2: Figure 4).

Our findings in the multi-sessions were limited to two children. The standard deviation in scores between the sessions is significantly reduced with the girl [duration] 46% [between] 30% [1st ball duration] 79% [2nd ball duration] 1% [3rd ball duration] 49% - the boy, who notably in the first session had an intravenous attachment, showed insignificant change in total. Overall, consistent to our single sessions were reduced 'between' times for both the girl (12%) and the boy (9%) which we claim as a possible indicator of motivation, which we relate to the enjoyment and fun in playing the game. This involves emergent learning of navigation modes and is an attribute to aesthetic resonance through its inherent presence factor. In the multi-sessions we conducted a preliminary computer analysis for duration of last pause and motion phases (Appendix 2: Figure 4). Our findings were that both the girl and the boy had increased standard deviation and average of duration of last pause phase combined with a reduced duration of motion phase from the first to second session. This may indicate that over a number of sessions less motion is required to achieve similar tasks, thus more effective movement is learnt as the child gets acquainted with the game. Further investigation in relating such findings to presence would seem in order.

To sum up, aesthetic resonance was indicated partly through the high degree of intensity and intentionality in movements. Intensity and intentionality was shown through the children's concentration and also through their force and passion when playing the game. Aesthetic resonance was indicated by the degree of movements of range and shifts in the children's movements. The categories of speed and fluency did not have any influence on aesthetic resonance as they did not influence the intensity, intentionality, range, or shifts in movements.

4.2. Interface and activities

In interviews with children concerning their positive and negative experiences of the EyeToy® game the main part of the children expressed positive experiences. 61.1% (n = 11) of the children thought the EyeToy® game was fun, while 22.2% (n = 4) said that they liked it. One (1) child said that the EyeToy® game was difficult, but he also said that the gameplaying was fun. Concerning positive and negative specifics of the gameplay 38.8% (n = 7) of the children answered on the interface attributes and 61.1% (n = 11) on the activity attributes of the game (Table 1). The children's negative experiences of the game only concerned activity attributes regards the content of the game. Two children answered that they enjoyed the whole EyeToy® game. Six children referred to movements – using the body and to move – when they were asked about the positive attributes of the game. Four children said that the ball-play attribute was the best, while seven children stated that the ball-play attribute was the most difficult. These facts indicate that the ball-play attribute in itself was a challenging activity, as three of the children also confirmed.

Table 1 Attributes

Positive?			
Interface	Children	Activity	Children
Body used	22.2% (4)	Ball-play	22.2% (4)
To move	11.1% (2)	Monkeys	16.6% (3)
Mirroring	5.5% (1)	Challenge	16.6% (3)
		Scoring	5.5% (1)
SUM	38.8% (7)	SUM	61.1% (11)
Negative?		Difficult?	
Activity	Children	Activity	Children
Monkeys	5.5% (1)	Ball-play	38.3% (7)
Repetition	5.5% (1)		
Pauses	5.5% (1)		
SUM	16.6% (3)	SUM	38.3% (7)

The game agents were the main attributes when the children referred to negative aspects of the EyeToy® game experiences as it repeatedly gave negative feedback to the children. The monkeys were stated as difficult by one child, but were also considered as fun by three of the children.

In summary, the children's experiences of the EyeToy® game indicated that the interface supported the gameplaying activity in a challenging way and aesthetic resonance was achieved through this challenge

4.2. Resource for therapy

In interviews and the field notes from the play therapists and the doctors positive, negative, and practical aspects of the children's gameplay with the EyeToy® game was started. They also gave indications on potential with the EyeToy® game in therapy.

***Positive aspects*:**
The EyeToy® game was great fun for the children who were concentrated on the tasks in the game.

***Negative aspects*:**
The children quickly became bored as it was either too hard or too easy to play; three balls were too few; the game ended quickly limiting the challenge; the game agent mostly gave negative feedback, which many of the children commented upon.

***Practical aspects*:**
A room allocated for the test is necessary for future research; the camera set-up was too complicated to handle; the camera set-up limited some of the children's movements; both hospitals wish to continue with future EyeToy® research.

***Potentials with EyeToy® in therapy*:**
The game activity is fun and the training aspect simultaneously involved, becomes fun as well; the game activity brings in movements to the therapy, which make sense and benefits the children's rehabilitation; playing the EyeToy® game becomes physiotherapy; if there was more of challenge and action in the games, the potentials for therapy would increase as the fun and motivation for moving probably would increase.

To sum up, the results from field notes and interviews with the play therapists and doctors underlined the potential with the EyeToy® system in therapy emphasizing flow and fun aspects of the gameplaying as beneficial for the therapy training.

5. Discussion

The purpose of the study was to qualify the initial use of the system for children in rehabilitation in a hospital scenario with a consideration of the inherent logistics and practicalities. We restricted our unit of analysis to different expressions of tempospatial movements in process as indicators of a possible presence state related to behaviour and situation within play therapy. Through our exploratory investigation our findings indicate that aesthetic resonance through intensity and intentionality is related to flow and conscious reactions when a child interacts with the EyeToy® game. Furthermore, presence enhanced aesthetic resonance through range and shift related to movement increments. As far as we can ascertain, the limited computed data supports the manual annotations and our claim where observation of activity mediated within a human afferent efferent neural loop closure as a result of interaction to content of a virtual environment. The field-experiments we consider as a start toward understanding the mechanisms of motivation promoted by multimodal immersion, and the triangulations of actions becoming reactions resulting in interaction in play activities.

Conclusions

Our approach relates to the heuristic evaluation strategy of Nielsen [28] where natural engagement and interaction to a virtual environment having 'real-world' physical traits and being compatible with user's task and domain is such that expression of natural action and representation to effect responsive artefacts of interesting content feedback encourages a sense of presence. Beyond presence we seek a *sense state* continuum that stimulates intrinsic motivated activity, and from prior research we have termed this aesthetic resonance. To engage an actor in aesthetic resonance we implement a strategy toward creating enjoyment and fun as the user perceived level of interaction where emotional expression of body is the control data of the feedback. In this way an afferent efferent neural feedback loop is established. The data that is controlling the feedback content is available for therapeutic analysis where progression can be monitored and system design adapted to specifics of the task centred training. The user experience however is targeted at being solely play based.

In this document we report on our pilot study which is the first phase of an extended full scale research investigation based on our hypothesis that the positive attributes in utilizing digital interactive games that embody the actor in VE therapy will relegate the negativity tagged to video games and offer new opportunities to supplement traditional therapy training and testing. Our prior research informs that intrinsic motivation is a potential strength of game interaction where the user becomes aware only of the task and in an autotelic manner extends otherwise limiting physical attributes beyond what may otherwise be possible to achieve, and this supports our hypothesis. This study discovered that problems to overcome are the video recording system, the interpretation of instruction, and the room availability. A new single button system for optimizing the video recording system has been designed and budget planned to improve the next phase of the project. Similarly, the hospitals promise a designated space in future. The children's quantity, dynamic, and range of movements when immersed in the gameplaying activity were over and above their usual range of movements. Their facial expression and emotional outbursts further substantiated our claim that an initial state of presence was achieved.

Acknowledgements

Hospitals, Staff, & children: *Laenssjukhuset* in Halmstad, Sweden; *Sydvestjysk Sygehus* in Esbjerg, Denmark. SCEE, Egmont/Nordisk Film, Scandinavia. Sony Denmark.

This study was part financed by cooperation between Sony Computer Entertainment Europe; Egmont Interactive, Scandinavia; Sony Denmark, SoundScapes ApS, Denmark, and the authors.

"Playstation" is a registered trademark of Sony Computer Entertainment Inc. "EyeToy" is a registered trademark of Sony Computer Entertainment Europe.

Algorithms adapted from those created with partial support from IST Project CARE HERE where author¹ was researcher [ref - 27].

References

[1] A. Brooks, E. Petersson. Recursive Reflection and Learning in Raw Data Video Analysis of Interactive 'Play' Environments for Special Needs Health Care. *Healthcom 2005, 7th International Workshop on Enterprise Networking and Computing in Healthcare Industry*, 83-87. June 2005.

[2] M. Slater, A. Steed, J. McCarthy, F. Maringelli. The influence of Body Movement on Subjective Presence in Virtual Environments. *Human Factors* 40:3, 469-477, 1998.

[3] A. Brooks. Enhanced Gesture Capture in Virtual Interactive Space. *Digital Creativity*, 16, 1:43-53. 2005.

[4] A. Brooks, S. Hasselblad. CAREHERE - Creating Aesthetically Resonant Environments for the Handicapped, Elderly and Rehabilitation: Sweden. In: Sharkey, P., McCrindle, R., Brown, D. (eds.): *5th Int. Conf. on Disability, Virtual Reality, and Associated Technologies*, 191-198, September 2004.

[5] A. Brooks, E. Petersson. Humanics 2: Human Computer Interaction in Acquired Brain Injury Rehabilitation. Proceedings of *HCI International 2005*, Las Vegas, USA. July 2005. CD Rom.

[6] D. Rand, R. Kizony, P.L. Weiss. Virtual reality rehabilitation for all: Vivid GX versus Sony PlayStation II EyeToy. In: Sharkey, P., McCrindle, R., Brown, D. (eds.) *5th Intl. Conf. On Disability, Virtual Environments and Assoc. Technologies* 87-94, September 2004.

[7] R. Kizony, N. Katz, H. Weingarden, P.L. Weiss. Immersion without encumbrance: adapting a virtual reality system for the rehabilitation of individuals with stroke and spinal cord injury. In P. Sharkey, C. Sik Lányi & P. Standen (eds.) *4th intl Conf. on Disability, Virtual Reality and Assoc. Technologies*, 55-62. September 2002.

[8] A. Brooks, S. Hasselblad, A. Camurri, N. Canagarajah. Interaction with shapes and sounds as a therapy for special needs and rehabilitation. In P. Sharkey, C. Sik Lányi & P. Standen (eds.) 4th *Int. Conf. On Disability, Virtual Reality, and Associated Technologies,* 205-212. September 2002.

[9] A. Brooks. Virtual Interactive Space (V.I.S.) as a Movement Capture Interface Tool Giving Multimedia Feedback for Treatment and Analysis. *13th International Congress of The World Confederation for Physical Therapy*, 66. May 1999.

[10] S.H. You, S.H. Jang, Y.H. Kim, M. Hallett, S.H. Ahn, Y.H. Kwon, J.H. Kim, M.Y. Lee. Virtual Reality-Induced Cortical Reorganization & Associated Locomotor Recovery in Chronic Stroke, retrieved 28/7/2005 www.hamptonu.edu/News_Publications/

[11] A. Brooks, Humanics 1: A study to create a home based telehealth product to supplement acquired brain injury therapy. In Sharkey, P., McCrindle, R., Brown, D. (eds.) *5th Intl. Conf. On Disability, Virtual Environments and Assoc. Technologies* 43-50, September 2004.

[12] B. Rogoff. *Apprenticeship in Thinking. Cognitive Development in Social Context*. NY: Oxford Uni. 1990.

[13] X. Retaux. Presence in the environment: theories, methodologies and applications to video games. *Psychology Journal,* 1: 3, 283-309, 2003.

[14] J. Lave, E. Wenger, *Situated learning. Legitimate peripheral participation.* NY: Cambridge University, 1991.

[15] M. Csikszentmihalyi. *Flow: The psychology of optimal experience*. Natur och Kultur, Stockholm. 1992.

[16] A. Brooks. SoundScapes – A concept and methodology of "Being There". In *Proceedings 6th Annual Intl Workshop on Presence*. [Abstract] 67. 2003.

[17] K.M. Lee, W. Peng. Effects of Screen Size on Physical Presence, Self Presence, Mood, and Attitude toward Virtual Characters in Computer/Video Game Playing. In *Proceedings 6th Annual Intl Workshop on Presence*. [Abstract] 23. 2003.

[18] R. Laban. *Modern Educational Dance*. 3rd edn. MacDonald and Evans, Bungay Suffolk. 1963.

[19] J.V. Wertsch. *Mind As Action*. Oxford University Press, New York Oxford. 1998.

[20] L.S. Vygotsky. *Mind in Society: The development of higher psychological processes*. Harvard University Press, Cambridge, MA. 1978.

[21] L-E. Berg, *Den lekande människan. Lund:* 1992.

[22] A.N. Leont'ev. The problem of activity in psychology. In J.V. Wertsch, (ed), *The concept of activity in Soviet psychology*. Armonk, NY, 37-71. 1981.

[23] A.N. Leont'ev. *Activity, consciousness and personality*. NJ: Prentice-Hall, Englewood Cliffs. 1981.

[24] E. Wenger. *Communities of Practice. Learning, Meaning, and Identity*. Cambridge University Press, NY. 1998.

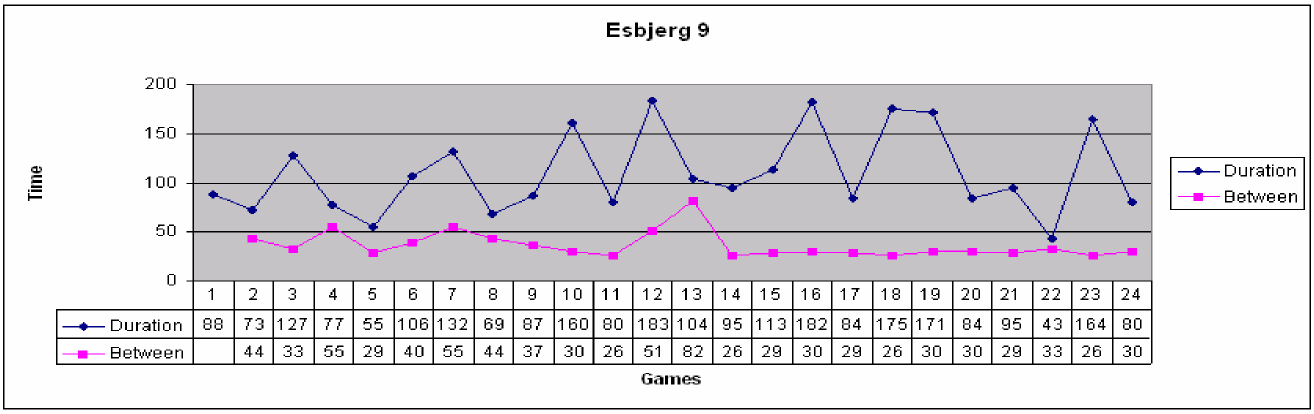
[25] J. Bigün, E. Petersson, R. Dahiya. Multimodal interfaces designing digital artifacts together with children. *7th Intl. Conf. on Learning and Educational Media*, Slovakia. 2003.

[26] J. Ruesch, W. Kees. Nonverbal communication: *Notes on the visual perception of human relations*. 76. UCLA, 1970.

[27] A. Camurri, B. Mazzarino, G. Volpe. A tool for analysis of expressive gestures: The EyesWeb Expressive Gesture Processing Library. Ret. 18/2/2005 www.megaproject.org

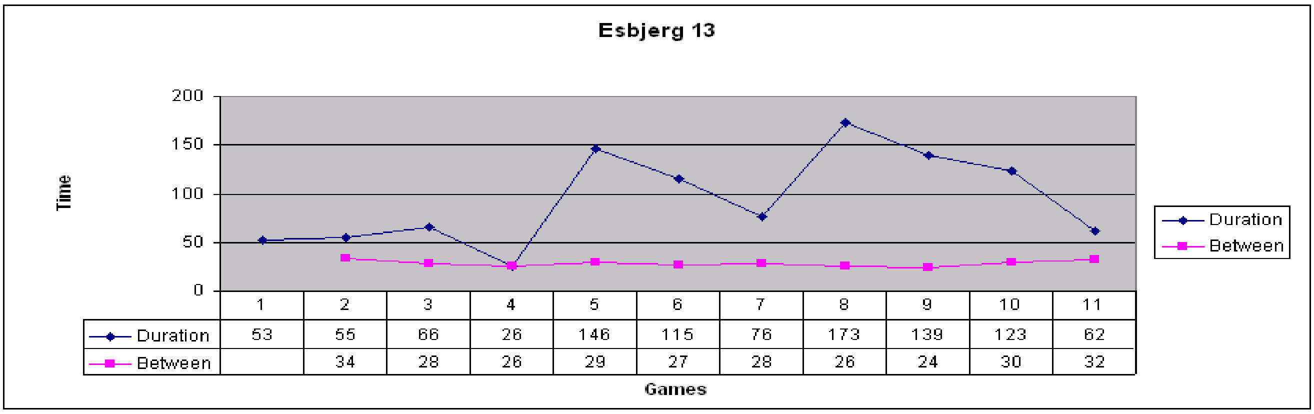
[28] J. Nielsen. *Usability engineering*. NY: Academic. 1993.

Appendix 1



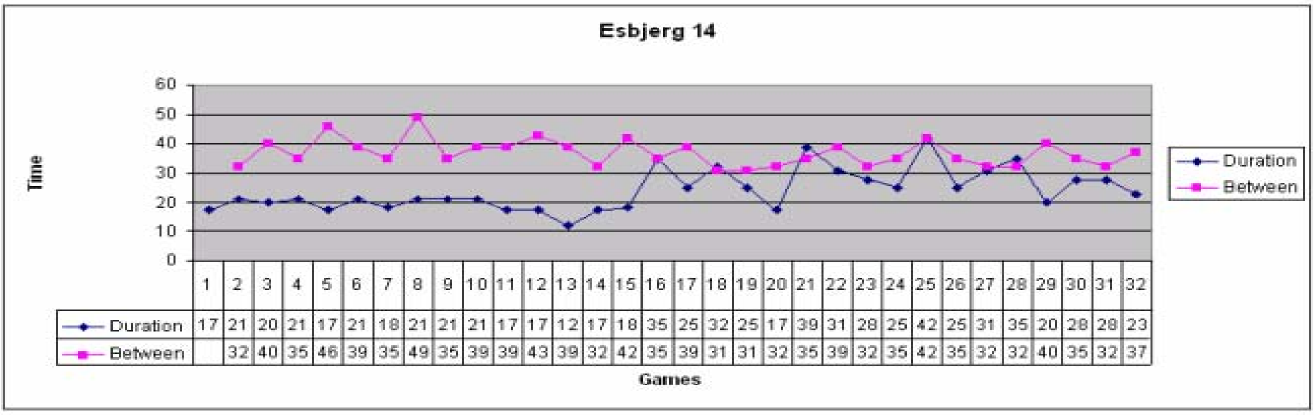
Esbjerg 9

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Duration	88	73	127	77	55	106	132	69	87	160	80	183	104	95	113	182	84	175	171	84	95	43	164	80
Between		44	33	55	29	40	55	44	37	30	26	51	82	26	29	30	29	26	30	30	29	33	26	30



Esbjerg 13

	1	2	3	4	5	6	7	8	9	10	11
Duration	53	55	66	26	146	115	76	173	139	123	62
Between		34	28	26	29	27	28	26	24	30	32

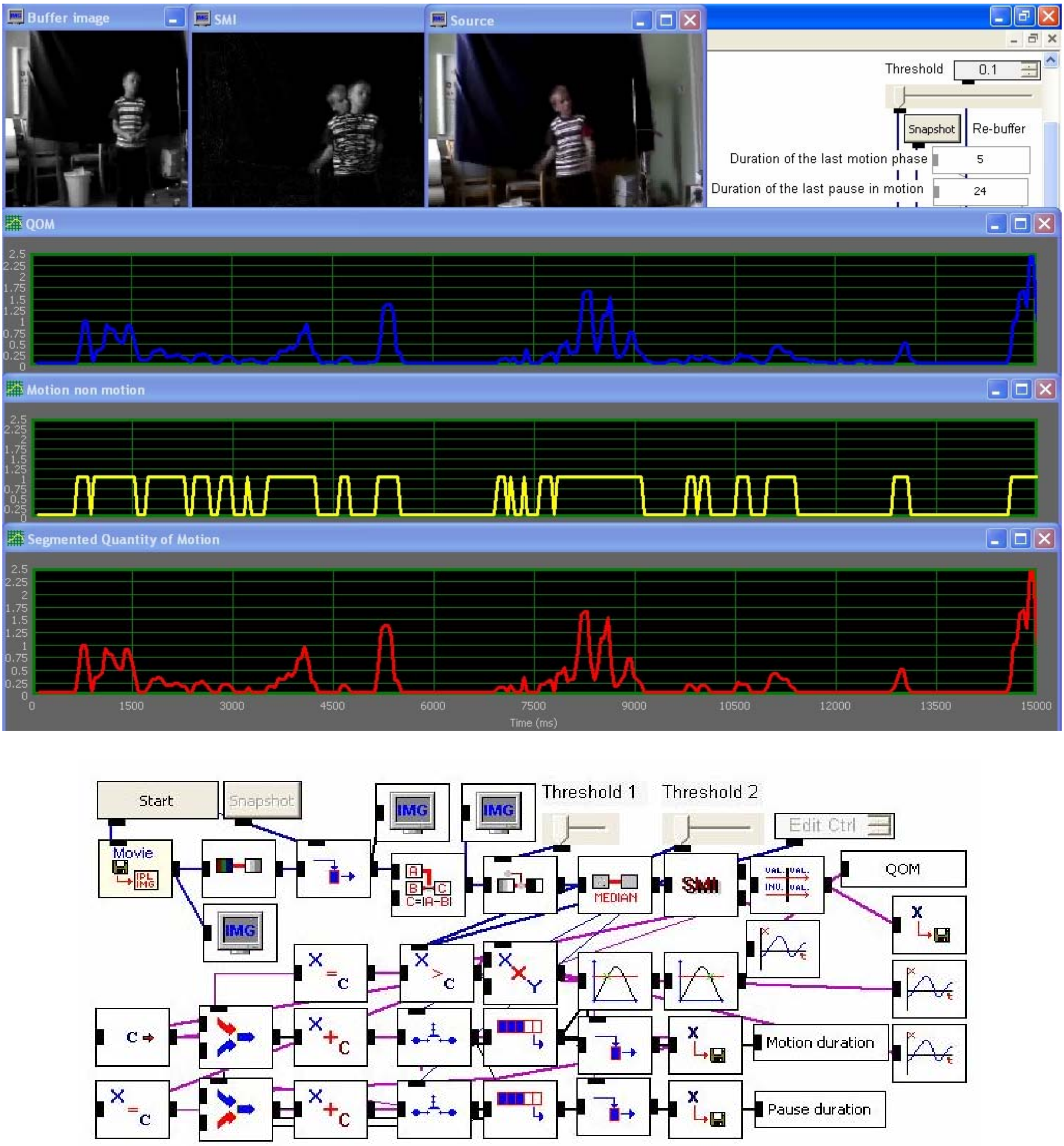


Esbjerg 14

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
Duration	17	21	20	21	17	21	18	21	21	21	17	17	12	17	18	35	25	32	25	17	39	31	28	25	42	25	31	35	20	28	28	23
Between		32	40	35	46	39	35	49	35	39	39	43	39	32	42	35	39	31	31	32	35	39	32	35	42	35	32	32	40	35	32	37

Figure 3 Three examples showing game play results: (top graph) Esbjerg 9 (male 7 years of age) where successes are inconsistent and possibly due to unstable presence. Game 13 is where a higher level was attempted shown by his 'between time' high. Esbjerg 13 (girl of 8 years of age –middle graph) achieved completion of the full game (8th game) resulting in an affirmative comment from the game agent. Esbjerg 14 (female 10 years of age – low graph) had most problems (game duration average 24/56.6) this reflective of her functional condition (brain tumor), however she achieved the most number of games (32) whilst continuously pushing her limitations and at conclusion interview described the "great fun" despite her difficulties.

Appendix 2



Threshold 0.1

Duration of the last motion phase 5

Duration of the last pause in motion 24

Figure 4 Quantity and segmentation of movement. Threshold/buffer/motion phase indicators (upper right). Buffer image, SMI & source windows (upper left), Halmstad hospital, Sweden. Algorithm for QOM, pause and motion phase duration available from authors.

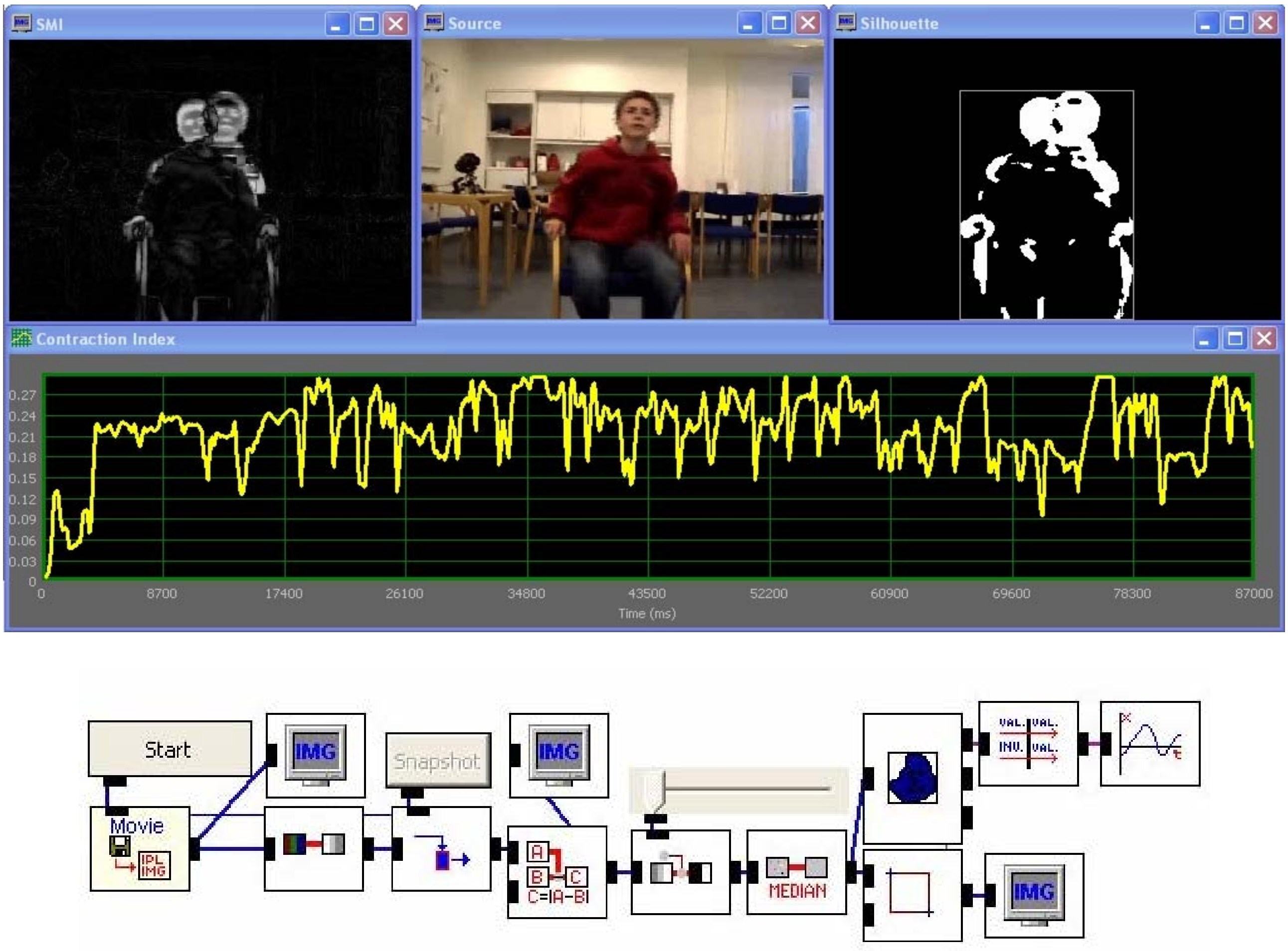


Figure 5 Contraction Index (CI) analysis. Upper right shows silhouette bounding rectangle initially set on buffer image, Esbjerg hospital, Denmark. Algorithm is made available from the authors.

Appendix 3

Table 2: Session overview: Upper = Sessions/Games (g)/Pauses (p). Lower = Movement analysis

Session	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Total games	16	15	28	10	6	7	13	5	24	14	5	5	11	32	14	8	5	6	8	8
Longest g #	7	13	25	9	4	7	5	5	12	7	3	2	8	25	11	2	1	1	3	5
Shortest g #	1	15	7	4	2	3	9	1	22	1	4	3	4	13	3	1	2	3	7	2
Longest p #	16	8	6	2	5	2	5	2	13	2	3	4	2	8	2	3	5	4	8	6
Shortest p #	4	14	9	9	3	6	8	4	14	13	4	2	9	18	8	4	4	6	5	2

Category of movement trait	High degree (%)	Medium degree (%)	Low degree (%)
Speed	33.4	22.2	44.4
Intensity	61.1	16.6	22.3
Fluency	16.6	55.5	27.9
Range	72.2	16.6	11.2
Intentionality	55.5	22.2	22.3
Shifts	66.6	16.6	16.8

Appendix 4.

Table 3: Tempospatial Analysis: An example of one annotated session video file.

Time min.	Temporal						Spatial						Screen
	Speed		Intensity		Fluency		Range		Intentionality		Shift		
	Hi	Lo	Hi	Lo	Hi	Lo	Hi	Lo	Hi	Lo	Hi	Lo	
1		1		1		1		1		1		1	Start screen/Character
2		1		1		1		1		1		1	Wave/Ball/Game Over
3		1	1		1			1	1			1	Character/Wave/Ball/Game Over
4	1		1			1		1	1			1	Ball/Monkeys/Game Over/Character/Wave
5	1		1			1	1			1		1	Wave/Ball/Monkeys
6		1	1			1		1	1			1	Monkeys/Game Over/Character/Wave/Ball
7		1	1			1	1		1			1	Ball/Monkeys/Game Over/Character/Wave
8	1		1			1	1		1			1	Wave/Ball/Monkeys/Game Over
9	1		1			1		1	1		1		Character/Wave/Ball/Monkeys
10	1		1			1	1		1		1		Monkeys/Game Over/Character/Wave
11	1		1			1	1		1		1		Ball/Monkeys
12	1		1			1	1		1		1		Ball/Monkeys/Game Over
13	1		1			1	1		1		1		Character/Wave/Ball/Game Over
14	1		1			1	1		1		1		Character/Wave/Ball/Monkeys (shortly)
15		1	1			1		1	1		1		Monkeys/Game Over/Character/Wave
16	1		1			1	1		1		1		Ball/Monkeys/Game Over/Character/Wave
17	1		1		1		1		1		1		Wave/Ball/Monkeys
18		1	1			1		1	1			1	Monkeys/Game Over/Character/Wave/Ball
19	1			1		1		1		1		1	Ball/Monkeys/Game Over/Character/Wave
20		1	1			1	1		1		1		Wave/Ball/Monkeys/Game Over
21	1			1	1			1		1		1	Character/Wave/Ball/Monkeys
22		1		1		1		1		1		1	Game Over/Character/Wave
23		1		1		1		1		1		1	Ball/Monkeys/Game Over/Character
24		1		1		1		1		1		1	Wave
25		1		1		1	1			1		1	Ball/Monkeys/Game Over
SUM	13	12	17	8	3	22	12	13	16	9	10	15	

An Augmented Reality system for the treatment of acrophobia

M. C. Juan[1], D. Pérez[1], D. Tomás[1], B. Rey[1], M. Alcañiz[1], C. Botella[2], R. Baños[3]
[1] MedICLab (Universidad Politécnica de Valencia)
[2] Departamento de Psicología Básica y Psicobiología (UJI)
[3] Universidad de Valencia

Abstract

This paper presents the technical characteristics of the first prototype that uses Augmented Reality to treat acrophobia. The immersive photographs are the virtual elements that represent the locations that the user fears. A total of 36 different immersive photographs have been included in the system (12 different locations with 3 parallel photographs in each location). At first, the system shows the central photograph. If the user rotates his/her head and stays in the same position, he/she can spin over the immersive photograph, changing his/her point of view inside the photograph. If he/she moves to the left/right (i.e. the physical position) the photograph will change and the related left/right photo will appear.

Keywords--- **Acrophobia, Augmented Reality, immersive photographs, virtual therapy**

1. Introduction

In an Augmented Reality (AR) system, users see an image made up of a real image and virtual elements that are superimposed over it. The most important aspect in AR is that the virtual elements add relevant and helpful information to the real scene. AR can be a successful tool in many fields as it can be applied to any field where the information superimposed on the real world can help the user.

This article presents an AR system for the treatment of acrophobia. This is the first system that uses AR to treat this type of phobia, but it is not the first system that treats acrophobia using Virtual Reality (VR) [1]-[5].

VR is currently a very useful tool for the treatment of several psychological problems (fear of flying, agoraphobia, claustrophobia, eating disorders, etc.). The number of studies showing the efficacy of VR environments as therapeutic tools has increased in the last few years.

AR and VR share some advantages with respect to traditional treatments. However, AR also presents advantages with respect to VR. In the specific case of acrophobia, creating different locations of high quality is extremely costly. VR applications could include avatars that simulate patients' bodies, however they cannot see their own feet, hands, etc. as can be seen in AR.

Our group has recently presented an AR system for treating phobias to small animals (cockroaches and spiders) [6]. In our work, we have demonstrated that, with a single one-hour session, patients significantly reduced their fear and avoidance. The system was tested on eleven patients. Given that this first application has proved to be effective, we believe that AR will be also effective with acrophobia.

1.1. Immersive photography

Immersive photography is a technique wherein the entirety of a space is captured from a single point and digitally processed to create a 360-degree photograph. When an immersive photograph is viewed, it appears to be a standard two dimensional photograph, but when manipulated by the user, it spins 360-degrees in any direction. This allows a user to look around a terrace, for example, in any direction that he/she chooses. He/she can look at the view out the balustrade, the hammock to the right, or the sky over him/her. He/she can even turn all the way around and look at all the details.

Immersive photography was widely developed during the 1990s [7] [8]. There are panoramic visualization systems like QuickTime VR [7] or Surround Video [9]. These systems are based on 360º cylindrical panoramic static images.

Immersive photography has been used to create VR environments, but it has not been included in an AR system. For example, VideaLab Research group (videalab.udc.es) has used immersive photography in several VR projects.

1.2. Acrophobia

According to the DSM-IV [10], specific phobias consist of the persistent fear of a circumscribed stimulus and consequent avoidance of that stimulus, where the person having this fear knows it is excessive or unreasonable. The phobia interferes significantly with daily life. Acrophobia is an intense fear of heights. A person who suffers from acrophobia tries to avoid: balconies, terraces, lifts, skyscrapers, bridges, planes, etc. People who suffer from acrophobia are fearful in any situation that implies heights; they even become anxious when other people are in those situations. The greatest fear is falling.

The incidence of acrophobia ranges from 2% to 5% of the general population; twice as many women as men suffer from this fear. Acrophobia usually has an early onset and is usually associated to having an aversive experience in a high place. It can also be indirectly acquired, by receiving information about distressing experiences related to closed spaces or by seeing someone having such a distressing experience.

The first treatments for acrophobia were graded exposures in-vivo. In these treatments, the avoidance behaviour is broken by exposing the patient to a hierarchy of stimuli. After a time, habituation occurs and the fear gradually diminishes.

The first VR system used to treat acrophobia was tested on 32 patients and had offered a 90-percent success rate [1]. Later, other VR systems have shown that VR is effective in the treatment of acrophobia, for example [2][3]. Several experiences comparing the effectiveness of VR with exposure in vivo have also been presented, two of these are [4][5]. These experiences have shown that VR exposure is as effective as in vivo exposure.

2. Material

2.1. Hardware

There are two types of hardware components: the hardware needed to run the system, and the hardware needed to obtain the immersive photographs. Immersive photographs were taken using a digital colour Coolpix 4500 Nikon Camera and the FC-E8 Fisheye converter. The system can run on a typical PC, without any special requirements. The real world is captured using a USB camera. We used Logitech QuickCam Pro 4000. The AR image is shown in a HMD and on a monitor. Thus, the therapist has the same visualization as the patient. We used 5DT HMD (5DT Inc., 800 H x 600 V, High 40º FOV). The camera was attached to the HMD so that it focuses wherever the patient looks. The system must also know the position of the patient's head in order to spin the immersive photograph in according with the patient's head movements. We have used the intertrax2 tracker to detect the patient's head rotation. We have also attached the tracker to the HMD.

2.2. Development tool

The application was developed using Brainstorm eStudio (www.brainstorm.es). Brainstorm eStudio is an Advanced, Multiplatform Real Time 3D Graphics presentation tool.

We have included ARToolKit [11] into Brainstorm as a plugin which was programmed in C++. ARToolKit is an open source Library in C that allows programmers to easily develop AR applications. It was developed at Washington University by Kato and Billinghurst. The required elements of the application are: a USB or Firewire camera, and a marker. Markers are white squares with a black border inside of which are symbols or letter/s. ARToolKit uses computer vision techniques to obtain the position and orientation of the camera with respect to a marker. Virtual elements are drawn over these markers.

By including ARToolKit possibilities into Brainstorm eStudio, we have AR options in a 3D graphic presentation tool, which offers many advantages. ARToolKit recognizes the markers and obtains the position and orientation where virtual 3D objects must be placed. Brainstorm eStudio uses this information to draw the virtual 3D objects. This plugin can work with more than one marker. The position and orientation of each marker is assigned to as many different 3D objects in Brainstorm eStudio as needed. The plugin is loaded as a dynamic library (dll) in execution time.

3. AR system for acrophobia

The system allows the therapist to perform a graded AR exposure. The patient starts with the minimum height, in case of our study a first floor. The therapist can change to the next height when the patient is prepared to do so. The maximum height is a 15th floor. Twelve different places have been included.

The patient is treated in a room with a balustrade and five markers on the floor. The markers are located in front of the balustrade. At first, the patient is placed next to the balustrade and in front of the central marker. If the patient rotates his/her head and stays in the same position, he/she can see the corresponding part of the immersive photograph. The patient has the same sensation/visualization as if he/she were rotating his/her head in the real location. If the patient moves to the left/right (the physical position) the photograph changes, and the related left/right photo appears. This process is shown in Figure 1.

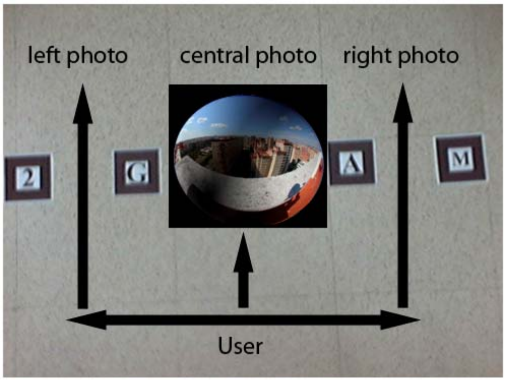


Figure 1 Possible movements of the user

The development of the system can be divided into two main steps:

- The creation of the immersive photographs
- The development of the application

3.1. The creation of the immersive photographs

The steps that were followed to create a 360-degree photograph that is suitable to be mapped as texture in Brainstorm eStudio were:

1. To take a 180-degree photograph
2. To retouch the photograph
3. To create a 360-degree photograph
4. To assign a transparency to the 180-degree white image

1. Taking a 180-degree photograph

We used the digital colour camera and the fisheye converter mentioned in section 2.1. The digital camera together with the Fisheye converter cover a field of view of over 180 degrees and is capable of capturing a full spherical panorama.

We took photographs of twelve different locations. In each location, we took three parallel photographs. The process was the following: The photographer was located next to the balustrade if there was one, or as closer as possible to the edge. The photographer took a photo and moved one meter towards his/her right, maintaining his/her parallel position with respect to the edge of the location. Then he/she took another photograph and repeated the process. Once this was done, there were three parallel immersive photographs that were one meter apart. Figure 2 shows this process.

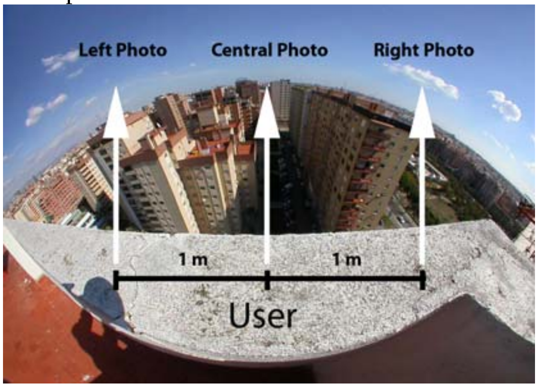


Figure 2 Process of taking 3 immersive photographs

2. Retouching the photograph

The photographs were retouched using Adobe Photoshop. In this step, undesirable information was removed from the image (for example, the feet of the photographer).

3. Creatint a 360-degree photograph

As the photograph is 180 degrees, a new 360-degree image must be created. In our system, we created a 360-degree image by sewing the 180-degree photograph and a transparent 180-degree image. We used PTStitcher to achieve this goal. This program belongs to the software Panorama Tools of Helmut Dersch (fh-furtwangen.de/~dersch). Figure 3 shows an image of this process.

4. Assigning a transparency to the 180-degree white image

Brainstorm eStudio uses the 360-degree image as texture. The 180-degree white image must be converted into a transparent image, otherwise the white 180-degree image would cover the user's position and he/she would not see his/her body. The system maps this new image as a 360-degree texture. This process was performed using Adobe Photoshop.

3.2. Characteristics of the system

As mentioned above, Brainstorm eStudio is the tool we used to develop our system. The plugin of ARToolKit included in Brainstorm eStudio is used to deal with the AR part of the system.

The system includes 12 different locations/levels. These locations were chosen by expert psychologists. We attempted to select typical locations that a therapist uses in the treatment of acrophobia. The locations are the following:

- Images taken from a window of a building located on the first floor, the second floor, the third floor, the fourth floor, the fifth floor and the fifteenth floor.
- Interior of a University. View of the stairwell from the second and third floors.
- View of a dam: both sides.
- Images taken from a terrace located on the second floor and third floors.

At first, the system shows the first level (the minimum height). Changing from one level to the next can be done using the option menu or control keys. The system uses five different markers that ARToolKit recognizes. If the camera focuses on the central marker, the system shows the central photograph of the selected level on this central marker. If the camera is focuses to the left of this central marker (left markers), the system will show the left photograph of the selected level. The same occurs for the right photograph. The immersive photograph is mapped as a spherical texture on a sphere. The appropriate image of this sphere is determined by the orientation of the user (information given by intertrax2 tracker) and is shown over the marker. Therefore, the marker focussed that is focussed by the camera determines the immersive photograph that must be shown at the selected level. The part of this photograph to be shown is determined by the intertrax2 tracker. Figure 4 shows the initial position and orientation of the user with respect to the sphere. If the user rotates his/her head, 90 degrees (up) or -90 degrees (down), the user will visualize part of the immersive photograph. If the user rotates his/her head, more than 90 degrees (up) or less than -90 degrees (down), the user will visualize part of the immersive photograph and part of the image taken by the USB camera (real image).

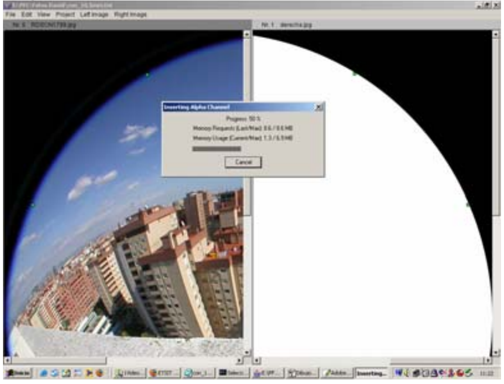


Figure 3 Creation of a 360-degree image using PTSitcher

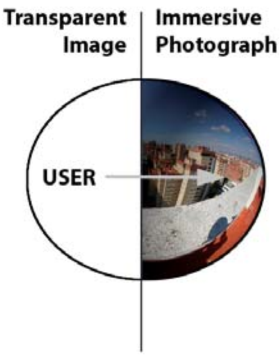


Figure 4 View of the user inside the 360-degree sphere

4. Visual examples

In this section, we include two examples captured during the execution of the system. We are currently testing the system with patients suffering from acrophobia.

Figure 5 shows an example taken during the execution of the application with an immersive photograph of a dam. In this figure, the virtual elements are the mapped images (immersive photographs), and the real images are the floor of the room and the feet of the person that is using the system. Figure 5.a) shows a view of the immersive photograph where the user is totally inside the 180 degrees of the immersive photograph. Figure 5.b) shows a view of the immersive photograph where the user is partly inside the 180 degrees of the immersive photograph and partly inside the 180 degrees of the transparent image. Figure 5.a) alone does not indicate that the image has been captured from an AR system; however, Figure 5.b) clearly indicates that it is an AR application.

5. Conclusions

AR and VR share advantages with respect to exposure in vivo. One of these advantages is that both realities offer control over the feared situations. Our AR system has one great advantage over VR. It offers more versatility than VR. With immersive photographs, the system can create on demand as many environments as the therapist desires with a high level of realism (the photograph is real) at a very low cost.

We have developed an AR system to treat acrophobia and we have added immersive photographs as virtual elements. This is the first prototype that uses AR to treat acrophobia and it is also the first time immersive photographs are included in a system like this. We are currently testing the system with patients suffering from acrophobia.

References

[1] Schneider, J., Lens-assisted in vivo desensitisation to heights. Journal of Behavior Therapy and Experimental Psychiatry, (1982) 13: 333-336.

[2] North, M.M., North, S.M. Coble, J.R., Effectiveness of virtual environment desensitization in the treatment of agoraphobia. Presence: Teleoperators and Virtual Environments, (1996) 5:346-352

[3] Jang, D.P., Ku, J.H., Choi,Y.H., Wiederhold, B.K., Nam, S.W., Kim, I.Y. & Kim, S.I., The development of virtual reality therapy (VRT) system for the treatment of acrophobia and therapeutic case. IEEE Transactions on Information Technology in Biomedicine, (2002) 6 (3):213-217

[4] Rothbaum, B.O., Hodges, L.F., Kooper, R., Opdyke, D., Williford, J.S., North, M., Effectiveness of computer-generated (virtual reality) graded exposure in the treatment of acrophobia. American Journal of Psychiatry, (1995) 152 (4):626-628

[5] Emmelkamp, P.M.G., Krijn, M., Hulsbosch, A.M., de Vries, S., Schuemie, M.J. & van der Mast, C.A.P.G., Virtual reality treatment versus exposure in vivo: a comparative evaluation in acrophobia. Behaviour Research and Therapy, (2002) 40: 509-516.

[6] Juan, M.C., Botella, C., Alcañiz, M., Baños, R., Carrion, C., Melero, M., Lozano, J.A.: An augmented reality system for treating psychological disorders: Application to phobia to cockroaches. The Third IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'04) (2004): 256-257

[7] Chen, S.E.: QuickTime VR – An image-based approach to Virtual Environment Navigation. 22nd annual ACM conference on Computer Graphics (1995): 29-38

[8] Chiang, C.C., Huang, A., Wang, T.S., Chen, Y.Y., Hsieh, J.W., Chen, J.W., Cheng, T.: PanoVR SDK – A software development kit for integrating photo-realistic panoramic images and 3-D graphical objects into Virtual Worlds. ACM symposium on Virtual Reality Software and Technology (1997): 147-154

[9] Surround Video API Reference, Black Diamond Consulting Inc. (1996)

[10] American Psychiatric Association, "Diagnostic and statistical manual of mental disorders DSM-IV-TR", (4th ed., text revision). Washington, DC, APA, (2000)

[11] Kato, H., Billinghurst, M.: Marker tracking and HMD calibration for a video-based augmented reality. 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR'99), (1999) 85-94

a)

b)

Figure 5 Examples. View of a dam. a) The user is inside of the immersive photograph b) The user is partly inside and partly outside of the immersive photograph

Using Virtual Reality to Provide Nutritional Support to HIV+ Women

Sarah Brown
CVC Laboratory
University of Cape Town
sbrown@cs.uct.ac.za

David Nunez
Dept. of Psychology
University of Cape Town
dnunez@humanities.uct.ac.za

Edwin Blake
CVC Laboratory
University of Cape Town
edwin@cs.uct.ac.za

Abstract

South Africa has one of the highest rates of HIV infection in the world. Health sector resources are limited and novel approaches to providing timely, accurate information are vital. Nutritional support and care has been recognized as an important aspect of HIV care. We developed a prototype virtual environment (VE) to provide nutritional information, and tested it at a government clinic in Cape Town, South Africa. 9 HIV+ women explored the VE and were interviewed after their experience. Despite the sample's low computer literacy, the system was found to be usable and enjoyable. The information provided by the system was rated highly for quality by the test group, although the amount was found to be lacking. In general, it seems that the system will be a useful adjunct to peer support groups as a way to disseminate relevant medical knowledge.

1. Introduction

South Africa currently has the largest living population of HIV positive individuals in the world [1]. This, combined with the limited resources of a developing world economy, presents significant challenges for the South African health sector. In particular, the shortage of qualified medical personnel to deliver care and accurate information is of concern. In 1993, the World Health Organization [2] reported that South Africa had 1640 persons per physician (contrast that figure with 370 for Germany, 630 for the USA, 719 for Albania and 1001 for Mexico). Therefore, the development of methods to extend the reach of health care professionals is a major research interest.

Due to a number of socio-economical factors, computer literacy in many South African communities is low. We are interested in investigating whether virtual reality (VR) is an effective information dissemination technology for individuals with minimal (if any) prior computing experience.

This paper reports on a prototype system designed to provide primarily nutritional and some social support to HIV positive patients (specifically pregnant women). The prototype system was deployed on a desktop PC, and input to the system was through minimal keyboard and mouse usage.

We report on an initial trial of the system, which took place at a government clinic in Cape Town, South Africa. The purpose of this system is to provide a widely accessible source of accurate information, while reducing the need for fixed facilities and highly trained medical staff.

1.1. Previous work

The use of virtual environment (VE) systems in health care has had some success in the past. For example, *Breast Cancer Lighthouse* [3], designed to provide support for cancer patients, makes use of a spatial experience as a metaphor to structure its content. Similarly, Hamza *et al* [4] used a virtual woodland campfire storytelling space to structure social support information, which consisted of personal narratives based on experiences of receiving an HIV positive diagnosis. Unlike pure information or multimedia systems, such as CHESS [5], these systems attempt to deliver information while making the interface as transparent as possible. This is of course highly desirable for populations with low literacy and computer literacy rates. Our prototype follows this pattern of using a VE to deliver health care information and social support.

1.2. Focus on nutritional support and care for HIV+ pregnant women

An HIV diagnosis often implies changes in most aspects of a patient's life. For our prototype, we decided to provide a basic structure which could be expanded to include information on many aspects of everyday life, but focused on nutritional information for the present. This information can have a significant impact on HIV prognosis (see for example [6, 7, 8]); also, dietary habits can be relatively easily influenced by the presentation of new information, and thus present a good opportunity for the system to make an impact.

For this trial, we focused on pregnant women for a number of reasons. Firstly, HIV status, due to social stigma, is usually kept secret; therefore finding volunteers can be difficult. Many HIV positive pregnant women, however, attend support groups, and are thus an accessible population of HIV positive participants. Secondly, young pregnant women are highly over-represented in HIV positive populations (in South Africa, about 30% of all pregnant women are HIV positive [1]), making the development of support tools for this group an important goal.

2. Description of the Virtual Environment

The VE's design was based on a South African council house, placing emphasis on creating an environment that users could recognize find familiar. Table Mountain is a familiar site in most Cape Town suburbs; we therefore used panoramic photographs incorporating the mountain to texture the VE skybox (Figure 1).

Figure 1: The skybox outside the house shows Table Mountain, providing a familiar context for the VE.

The house contains four rooms, in which information is presented in its context in the home. In the proposed completed system, each room will contain information and interactions relevant to the activities that occur in it. For the prototype, two of these rooms contain virtual actors and points of possible interaction: the lounge and the kitchen. The actors communicate with speech (recorded from voice actors to represent a cross-section of local accents) and body gestures. Three of the voice actors were photographed to provide textures for the virtual actors. Background sounds were used outside the house (quiet neighborhood sounds such as chirping crickets and traffic hum), and in the kitchen (water running from the tap when applicable).

2.1. Introduction and Lounge Interactions

Figure 2: Andile welcomes the user to his home. This is the first actor the user sees.

The user starts her VR experience outside the house and is greeted by an actor named "Andile" (Figure 2). When the user has crossed a trigger positioned near the front door of the house, Andile walks inside and encourages the user to follow him.

When the user enters the house she meets two other actors. Andile joins the others in the lounge who are all meeting in a casual support group (Figure 3). The user can listen to their discussions about their HIV experiences.

This scene serves as an introduction to the house and to the characters, and by the sharing of personal HIV stories, provides some social support [4]. After the introduction in the lounge, the user is encouraged to move into the kitchen to learn about nutrition.

Figure 3: Each of the three actors in the lounge tells something about their HIV history. This serves both as an introduction and provides social support.

2.2. Kitchen Interactions

An actor in the kitchen ('Sandi') introduces herself and explains the room's different points of interaction. Two areas are presented: the first presents the concept of food groups, and the second, presents concepts of cleanliness and hygiene in the home. The user is then free to interact in one of the two areas and may move freely between the two at any time.

2.2.1. Food Groups interaction At this site Sandi introduces the concept of food groups, each arranged in a labelled wooden bowl. The groups presented are "Fats & Oils", "Sugars", "Herbs", "Milk", "Energy", "Vitamins", "Proteins", "Fruits" and "Vegetables". Sandi then encourages the user to select one of the food groups to learn more about it (Figure 4). The user can select a food group by means of a cross-hair cursor.

2.2.2. Cleanliness and hygiene interaction At this site, Sandi explains the importance of keeping items in the house clean and sterile to minimize the risk of opportunistic infections. Three concepts are presented: "Clean Food", "Clean Stomach" and "Clean Water" (Figure 5). By selecting an icon, more information on that concept is provided.

Information presented in the kitchen was compiled primarily from *Positive Health*, a booklet written by Neil M. Orr [9]. Documents from the South African Health Department [10] and United Nations Food and Agriculture Organization (FAO) [11] were also used.

3. Method

We set out to explore the impact of our prototype and its possible benefits as well as identify areas for future development. The system was therefore deployed for an initial pre-trial at a local government clinic. This research is part of a larger study on the long term impact of the system.

Figure 4: Sandi introduces food groups. Each of the nine different food groups is in a bowl. Selecting a bowl provides more information about the group.

Figure 5: Concept of cleanliness and food safety. From left to right "Clean water", "Clean Food" and "Clean Stomach" presented at the sink in the kitchen.

3.1. Initial Trial

The trial was conducted at a Midwife Obstetric Unit (MOU) at a public hospital in Cape Town, South Africa. Permission was obtained from the hospital's Matron-in-Charge and the University of Cape Town ethics committee. The hospital's head dietician approved the information content. Due to the stigmatization around HIV status, a great deal of care was taken in the recruitment of volunteers and confidentiality was of utmost importance.

3.1.1. Equipment & Venue The VE was displayed on a typical desktop PC. Subjects wore headphones and input was by means of mouse and keyboard. We anticipated that the participants' prior experience with computers would be limited (if any at all) and thus decided to use only one key for virtual walking. The mouse was used for directional control and to look around. The frame rates varied with scene complexity, ranging from 12 to 40 frames per second. A dedicated room, usually used by counsellors, was made available for the study. Additionally, a trained HIV counsellor was on-site and available to the participants for the entire duration of the study. During the participant's experience of the VE, a researcher was available to deal with any participant queries.

3.1.2. Participants 9 participants were recruited from a support group for HIV+ pregnant women, who attended the MOU weekly. The study was conducted over two weeks on four different afternoons. Generally, two participants were seen per afternoon at scheduled times.

3.1.3. Procedure Participants were first taken through a training VE, in which they became familiar with the controls. Prior computer experience of all participants was minimal, with most subjects never having used a computer before. The training VE thus served to make them more familiar with computers in general (i.e. the names of the peripherals), and make them comfortable with the controls and skills required to navigate in the VE. This training environment contained same basic house model used in the main VE and participants were given five tasks, similar to those given in the main VE. These tasks involved navigating through the house and selecting (using the mouse) specific items displayed in each of the four rooms.

When participants were finished training, they were asked whether they felt comfortable with the controls and were ready to experience the main VE. One participant chose to re-explore the training environment while all other subjects were comfortable enough to move onto the main VE. The training VE took 5 minutes on average and the mean time spent in the main VE was 28 minutes. All participants viewed all of the available information in the VE.

3.1.4. Interviews Directly after experiencing the VE, semi-structured interviews were conducted. These were voice-recorded and later transcribed. The interviews ranged from 12 to 15 minutes.

4. Results & Discussions

Our evaluation of the prototype system is based on the data acquired through the semi-structured interviews. From a qualitative content analysis of the interview transcripts, three main themes arose: quality of the content information, ease of use of the system, and accessibility of the information through the VE interface.

4.1. Quality of information

Almost all of the participants (one exception) said they felt excited about the amount and quality of information presented by the system. All agreed that the information was useful to them – a majority said they learnt something new. One participant in particular said that the system highlighted for her how much she still needed to learn. All participants agreed that the information in the system would be of great help, and should be presented soon after an HIV+ diagnosis. There was some disagreement as to what the time frame should be. Two (of nine subjects) would have wanted to see the system immediately after diagnosis; while the other seven women stated they would prefer to have reached some degree of acceptance of their status before the system should be introduced.

4.2. Ease of use of the system

Half of the participants had used a computer before this study. Of those five, only two had used the internet, and only one had used it to search for information on HIV. However, all participants reported that the system was easy to use, although most expressed some anxiety at using or possibly damaging the equipment during the early phases of their experience. The ease of use is corroborated by the overall impressions given of the system. All participants found the system useful, and about half stated that they found it exciting. About the same number mentioned that they enjoyed the freedom afforded by walking around the VE and being able to choose the information they wanted.

4.3. Accessibility of information

The ease of use of the system suggested that the information was highly accessible to our target population. We asked participants if they would return on their own to use the system if it were deployed in clinics. Without exception, all participants responded that they would return to use the system if it were available. Interestingly, when asked if the system could replace the function of a peer-support group, most participants expressed that the system would make a good adjunct to a peer-support group, for providing more specific information. When asked if this type of information was difficult for them to find, most agreed that it was easy to get though their counselors. This suggests that in areas where counseling support is limited, the system may be particularly useful. In general, we found that while the interface presents no problem, the volume of information should be vastly increased – in the average 28 minute session, each participant selected and listened to all of the available information, and more than half specifically mentioned that they would like to see the amount and variety of information in the system increased.

5. Conclusions

We are greatly encouraged by the results of this initial trial. We feel that while the prototype requires expansion in terms of the amount the information presented, VR is an extremely useful and promising medium for our target population. Off-the-shelf hardware, which is affordable to both government agencies and NGOs, was used. Therefore the deployment of such a VE system in a number of clinics or mobile clinics is economically viable. Also, our interviews indicate that patients want and are able to make use of the system. So it is unlikely that the systems once deployed will lie idle. Also, due to the simplicity of the system, peers could train each other in its use, even in communities with low literacy and computer literacy rates. Although computing experience was minimal, all participants mentioned that they found the system easy to use and enjoyed their VR experience. Additionally, training time was minimal (5 minutes). This provides some evidence that VR can, and would be used in communities with low levels of computer literacy.

Indeed, the information presented by our system could conceivably be adapted to the needs of a narrow community of users by periodically interviewing them about their needs, and adding content as required. Our general conclusion is that, used together with peer support groups, our system could provide a very useful stopgap measure in the face of limited counselling resources.

The outstanding feature of this system is its potential to empower communities to deal with the HIV epidemic. The current shortage of medical staff and knowledge pose a problem since many people have little means for coping with the disease. However, by making high-quality information easily and widely available, the emphasis can shift from looking to a scarce group of outsiders to provide solutions, towards looking to one’s peers to cope with the problems of living with HIV. Patients can not only exchange information, but, by teaching each other how to use the system, increase the flow of information from the medical community to the patient community.

Acknowledgements

The authors would like to thank the following: CAVES consortium and all CAVES technical staff for their involvement in VR Direct (the software used to create the VE). Mignon Coetzee, a clinical psychologist, for her advice on the VE content. Robert Hofmeyr and John Bilay (from CAMA) for their assistance with sound recordings. Mbulelo Grootboom, Jill Levenberg, Shaun Webb and Mbali Kgosidintsi for the time, talent and energy they invested in the voice recordings. Atomic Visual Effects for animation work.

Finally, we would like to thank Ayanda Dayimani, an HIV/AIDS counselor working at the MOU, whose assistance was invaluable for the completion of this study.

References

[1] L. Makubalo, P. Netshidzivhani, L. Mahlasela and R. du Plessis, *National HIV and syphilis antenatal sero-prevalence survey in South Africa.*, in D. o. Health, ed., 2003.

[2] *World Development Report*, in W. H. Organization, ed., 1993.

[3] C. Heeter and D. D. Greene, *Breast Cancer Lighthouse CD ROM: Using personal stories for patient and clinician education.*, in S. Weghorst, J. B. Sieburg and K. S. Morgan, eds., *Medicine meets virtual reality: Health care in the information age.*, IOS Press, Washington, DC., 1996, pp. 649-651.

[4] S. Hamza, M. Coetzee, E. H. Blake and D. Nunez, *A low cost VR support system for people living with HIV/AIDS*, *CyberTherapy 2004*, San Diego, CA., 2004.

[5] B. R. Shaw, F. McTavish, R. Hawkins, D. H. Gustafson and S. Pingree, *Experiences of women with breast cancer: Exchanging social support over the CHESS network.*, Journal of Health Communication, 5 (2000), pp. 135-159.

[6] D. M. Mwamburi, I. B. Wilson, D. L. Jacobson, D. Spiegelman, S. L. Gorbach, T. A. Knox and C. A. Wanke, *Understanding the role of HIV load in determining weight change in the era of highly active antiretroviral therapy.*, Clinical Infectuous Diseases, 40 (2005), pp. 167-73.

[7] J. Nerad, M. Romeyn, E. Silverman, J. Allen-Reid, D. Dieterich, J. Merchant, V. A. Pelletier, D. Tinnerello and M. Fenton, *General nutrition management in patients infected with human immunodeficiency virus*, Clinical Infectuous Diseases, 36 (2003), pp. 52-63.

[8] S. Grinspoon and K. Mulligan, *Weight loss and wasting in patients infected with human immunodeficiency virus.*, Clinical Infectuous Diseases, 36 (2003), pp. 69-78.

[9] N. M. Orr, *Positive Health*, Juta, Nelspruit, South Africa, 2001.

[10] R. Kennedy, *South African national guidelines on nutrition for people living with TB, HIV/AIDS and other chronic debilitating conditions*, Department of Health, South Africa, 2001.

[11] FAO, *Living well with HIV/AIDS. A manual on nutritional care and support for people living with HIV/AIDS*, 2002.

Session 12
23nd September, 2005

Gaming and Connected Presence

18.00-18.30 *Spatial Presence and Emotions during Video Game Playing: Does it Matter with Whom You Play?*
Niklas Ravaja[1], Timo Saari[1], Marko Turpeinen[2] Jari Laarni[1], Mikko Salminen[1] and Matias Kivikangas[1]
[1] M.I.N.D. Lab/CKIR, Helsinki School of Economics, Finland
[2] Helsinki Institute for Information Technology, Finland

18.30-18.45 *Learning, experience and cognitive factors in the presence experiences of gamers: An exploratory relational study*
David Nunez[1] and Edwin Blake[2]
[1] Department of Psychology, University of Cape Town, South Africa
[2] Department of Computer Science, University of Cape Town, South Africa

18.45-19.00 *Being there together and the future of Connected Presence*
Ralph Schroeder
Oxford Internet Institute, UK

Spatial Presence and Emotions during Video Game Playing: Does it Matter with Whom You Play?

Niklas Ravaja[1], Timo Saari[1], Marko Turpeinen[2], Jari Laarni[1], Mikko Salminen[1], & Matias Kivikangas[1]
[1]M.I.N.D. Lab/CKIR, Helsinki School of Economics, Finland, [2]Helsinki Institute for Information Technology, Finland
{ravaja@hkkk.fi, saari@hkkk.fi, marko.turpeinen@hiit.fi, laarni@hkkk.fi, misalmin@hkkk.fi, kivikangas@hkkk.fi}

Abstract

The authors examined whether the nature of the opponent (computer, friend, or stranger) influences Spatial Presence, emotional responses, and threat and challenge appraisals when playing video games. In a within-subjects design, participants played two different video games against a computer, a friend, and a stranger. In addition to self-report ratings, cardiac interbeat intervals (IBIs) were measured to index physiological arousal. When compared to playing against a computer, playing against another human elicited higher Spatial Presence, engagement, anticipated threat, post-game challenge appraisals, and physiological arousal, as well as more positively valenced emotional responses. In addition, playing against a friend elicited greater Spatial Presence, engagement, and self-reported and physiological arousal compared to playing against a stranger. The nature of the opponent influences Spatial Presence when playing video games, possibly through the mediating influence on arousal and attentional processes.

Keywords--- **Spatial Presence, emotions, social relationships, interbeat interval, video games**.

1. Introduction

1.1. Presence and its determinants

The experience of media users that they are personally and physically present in the displayed environment has been named “Presence,” or more specifically, “Spatial Presence” [1, 2]. In 1997, Lombard and Ditton defined presence as "the perceptual illusion of nonmediation" (i.e., an illusion that a mediated experience is not mediated) [1]. Although a complete sense of Presence may be elicited only by emerging technologies, such as virtual reality, more traditional media (e.g., television, video games) offer a lesser degree of Presence as well [1, 3].

Researchers have identified several formal characteristics of media as determinants of Presence: the number of human senses for which a medium provides stimulation (i.e., media sensory outputs), the consistency of sensory outputs, image quality, image size, motion, dimensionality, camera techniques, aural presentation characteristics, interactivity, obtrusiveness of a medium, and the number of people the user can (or must) encounter while using a medium [see 1, 4, 5]. In addition, content features of a medium, such as social realism, use of media conventions, and the nature of task or activity, may exert an influence on Presence [1]. People also differ in their ability to experience Presence [6], and the characteristics of the medium user (e.g., personality, willingness to suspend disbelief) may have a considerable impact on the sense of Presence [e.g., 7].

1.2. Emotions

Emotions can be defined as biologically based action dispositions that have an important role in the determination of behavior [e.g., 8]. According to a dimensional theory of emotion, all emotions can be located in a two-dimensional space, as coordinates of valence and arousal (or bodily activation) [e.g., 8, 9]. The valence dimension reflects the degree to which an affective experience is negative (unpleasant) or positive (pleasant). The arousal dimension indicates the level of activation associated with the emotional experience, and ranges from very excited or energized at one extreme to very calm or sleepy at the other.

It is generally agreed that emotions comprise three components: subjective experience (e.g., feeling joyous), expressive behavior (e.g., smiling), and the physiological component (e.g., sympathetic arousal) [10]. Heart rate (HR; or cardiac interbeat interval, IBI) is a frequently used psychophysiological index of arousal [11]. Emotional arousal is accompanied by increased sympathetic nervous system (SNS) activity that causes the heart to speed up. However, the heart is innervated also by the parasympathetic nervous system. Increased cardiac parasympathetic activity causes the heart to slow down and is associated with information intake and attention [11]. Although the dual innervation of the heart may entail interpretative difficulties associated with HR, HR appears to index primarily emotional arousal during video game playing [12]. Electrodermal activity (EDA; or skin conductance) has also frequently been used as a measure of arousal [11]. When emotional arousal increases, the accompanying activation of the SNS results in increased sweat gland activity and skin conductance.

Facial electromyography (EMG) is, in turn, the primary psychophysiological index of emotional valence

[11]. It is well established that increased activity at the zygomaticus major (cheek) and corrugator supercilii (brow) muscle regions is associated with positive emotions and negative emotions, respectively, during affective imagery and when viewing pictures or other media stimuli [11, 13].

1.3. Presence and emotions

Spatial Presence may exert an influence on both the valence and arousal dimensions of emotions. Media presentations that engender a greater sense of Presence have been suggested as often eliciting greater self-reported and physiological arousal (i.e., a component of emotional arousal) [1, see also 11]. For example, Meehan et al. found that a frightening (i.e., arousing) virtual environment (VE) depicting a pit room with an unguarded hole in the floor leading to a room 20 ft. below elicited greater EDA and HR acceleration (i.e., a decrease in IBIs) compared to a non-frightening virtual room [14]. In addition, EDA and HR changes correlated positively with self-reported Presence during exposure to the frightening virtual height situation. It should be recognized, however, that there is no reason to expect that EDA or HR would increase with increasing Presence when the content of the mediated environment is non-arousing (e.g., a deserted beach of a Caribbean island).

Media presentations, such as video games, engendering a strong sense of Spatial Presence have been suggested as eliciting higher overall enjoyment [5, 15]. Ravaja et al. also recently showed that a high sense of Spatial Presence was related to increased positive, and decreased negative, emotional responses to success in a video game as measured by facial EMG [16]. Thus, Spatial Presence is also related to the valence dimension of emotions. It should be emphasized that, although some authors have used EDA, HR, and facial EMG as measures of Presence, these psychophysiological measures are primarily measures of arousal, emotional valence, or attention rather than direct measures of Presence [see 11].

It is also of note that, although high Presence conditions may elicit increased arousal or more positive or negative emotional responses, it is also possible that emotions affect Presence experiences. For example, it is possible that emotional arousal elicited by arousing media content is accompanied by increased attentional engagement with the mediated environment [e.g., 17], thereby increasing Spatial Presence. That is, as suggested by the two-level-model of Spatial Presence, the focused allocation of attentional resources to the mediated environment contributes to the formation of Spatial Presence [2]. As a result of this increased attentional engagement with arousing media content, there are also less attentional resources left over for the processing of the cues signaling that the mediated environment is artificial.

1.4. The present study

The present study was designed to examine the influence of the nature of an opponent on the experience of Presence and emotional responses when playing video games. That is, we asked does it make a difference whether one plays against a computer, a friend, or a stranger?

Social interactions may be arousing owing to involvement and enthusiasm [e.g., 18]. Johnston, Anastasiades, and Wood also found that a two-person “soccer” video game elicited higher HR reactivity compared to a “squash practice” video game against a machine, suggesting that the social-competitive situation related to the former game results in increased arousal [12]. In contrast, research on the effects of the laboratory analogues of social support has shown that the presence of a supportive other attenuates physiological responses to behavioral challenges [for a review, see 19]. However, it may matter who offers the support. In the study of Christenfeld et al., when subjects gave a speech to a supportive friend or a supportive stranger, the supportive behaviors from a friend resulted in smaller cardiovascular responses than the same supportive behaviors offered by a stranger [20]. In general, a participant performing a task in front of an observer (friend or stranger) may be expected to experience increased sympathetic arousal compared with a participant performing the same task in the absence of an observer [19]. This is because the presence of another person who “observes” inevitably creates a situation laden with task performance evaluation potential [19]. It is of note, however, that the nature of the opponent (a friend vs. a stranger) might make a difference when playing a competitive video game.

Given the aforementioned considerations, we hypothesized that playing against another person would elicit greater anticipated threat prior to the game (Hypothesis 1), perceived challenge (Hypothesis 2), self-reported arousal (Hypothesis 3a), and physiological arousal as indexed by decreased cardiac IBIs (Hypothesis 4a) compared to playing against a computer. Given that (a) playing video games against another person involves high evaluation potential and (b) as opposed to a stranger, a friend may serve as a continuing reminder of task performance [19], we also expected that playing against a friend would elicit greater self-reported arousal (Hypothesis 3b) and physiological arousal (Hypothesis 4b) compared to playing against a stranger. Heightened arousal when playing against another person, and particularly against a friend, is likely to be accompanied by an increased desire to perform well and attentional engagement with the game. As suggested above, the allocation of attentional resources to the mediated environment may contribute to Spatial Presence experiences [see 2]. That being so, we hypothesized that playing against another person (friend or stranger) would elicit greater Spatial Presence (Hypothesis 5a) and Engagement (Hypothesis 6a) compared to playing against a computer. In addition, playing against a friend was hypothesized to elicit higher Spatial Presence (Hypothesis 5b) and Engagement (Hypothesis 6b) compared to playing against a stranger.

Given that humans are social beings who have an appetitive motivation for social interaction (social relationships are intrinsically rewarding) [21], we expected that playing against another human being would elicit more positively valenced emotional responses compared to playing against a computer (Hypothesis 7a). Likewise, playing against a friend might elicit more positive responses compared to playing against a stranger (Hypothesis 7b). The suggestion that Presence may result in greater enjoyment [5, 15] is also relevant to these predictions.

2. Methods

2.1. Participants

Participants were 99 (51 male and 48 female) Finnish undergraduates with varying majors, who ranged from 19 to 34 years of age (mean = 23.8 years). Of them, 61% played video games at least once a month. Participants participated in the experiment in groups of three same-sex persons. In each of the 33 groups, two of the participants were friends who knew each other before and one was a person unknown to the others (i.e., a stranger). Each participant received three movie tickets for participation. In the present study, we used only the self-report and physiological data collected from the 33 so-called main participants (see below).

2.2. Design

A 2 (Game) × 3 (Opponent) within-subjects design was employed.

2.3. Video games

In the present study, we used two video games: Super Monkey Ball Jr. (Sega Corporation, Tokyo, Japan) and Duke Nukem Advance (Take 2 Interactive, Berkshire, UK). The games were played with the Nintendo Game Boy Advance console (Nintendo Co., Ltd., Kyoto, Japan). In the two-player condition, two Game Boy Advance consoles were connected with a Game Boy Advance Game Link Cable (Nintendo Co., Ltd., Kyoto, Japan).

Super Monkey Ball Jr. takes place in a surrealistic world with bright colors and includes a game board hanging in the air and a cute little monkey trapped in a transparent ball. The game view is from behind the monkey. In the single-player mode, the player's task is to tilt the board to roll the ball towards a particular goal without falling off the edge of the board into the depths. The player needs to avoid obstacles and pick bananas as the monkey rolls around the stages. The aim was to clear each stage with as high a score as possible. The player had 30 to 60 s to clear each stage and earned extra points for clearing the stage in half this time. The player earned extra points also by picking bananas. The practice session and the actual play sessions were played with the Normal Mode, Beginner difficulty level, and AiAi character. In the two-player mode (Monkey Duel), both players steered their own monkey characters and raced through the stage trying to reach the goal first. There were no bananas to collect and the player had to start from the beginning of the stage if he or she fell or was pushed off the edge. Also in the two-player mode, the player had 30 to 60 s to clear each stage. In general, Super Monkey Ball Jr. is a relatively nonviolent game with happy background music and atmosphere. It requires fast reflexes and some strategy.

Duke Nukem Advance is a version of the classic first-person shooting game. In the single-player mode, the player controls Duke Nukem character and tries to stop the alien scientists from taking over the world. The Game starts in a military base and the player has to clear each stage by finding a specific item or completing some task. To complete the tasks, the player has to kill alien monsters who roam over the base. The player has to solve some easy puzzles to proceed in the game and he or she can pick up more powerful weapons, armor, and first aid kits from around the base. In the single-player mode, the game was played at the Let's Rock difficulty level. The two-player mode was a death match game located in a similar military base environment. Each player controlled one Duke character and tried to hunt down and kill the other character. There were no other opponents, and after death the player could start again from a random place in the game environment. In general, the game is rather violent.

2.4. Procedure

When arriving to the laboratory, the three participants returned a number of questionnaires that had been sent to them beforehand. From the two participants who were friends, one was randomly chosen as the main participant. After a brief description of the experiment, the participant filled out an informed consent form. Electrodes were then attached to the main participant and he or she was seated in a comfortable armchair. Next all three participants practiced both games for 5 min in the single-player mode. This was followed by a rest period of 7 min, during which baseline physiological measurements were performed on the main participant. The main participant played each of the two games for 8 min against a computer (single-player mode of the game), a friend, and a stranger. The order of these six game sessions was randomized for each (main) participant. The main participant and opponent sat next to each other during game playing. After playing all games, the electrodes were removed, and the participants were debriefed and thanked for their participation.

2.5. Self-report measures

All self-report scales were presented on a computer screen.

2.5.1. Presence. The sense of presence of the participants was measured after each game with the ITC-Sense of Presence Inventory (ITC-SOPI), a 44-item self-report instrument [22]. Previous work with the ITC-SOPI has identified four separate factors: (a) Spatial Presence (19 items; e.g., "I had a sense of being in the game scenes," "I felt I was visiting the game world"), (b) Engagement (13 items; e.g., "I felt involved [in the game environment]," "My experience was intense"), (c) Ecological Validity/Naturalness (5 items; e.g., "The content of the game seemed believable to me," "The game environment seemed natural"), and (d) Negative Effects (6 items; e.g., "I felt dizzy," "I felt nauseous"). In the present study, we used only the 37 items addressing the first three factors. The wording of some of the items was slightly altered to adapt the instrument specifically for use with video games. Each of the items was rated on a 5-point scale, ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). The psychometric properties of the instrument have been shown to be acceptable.

2.5.2. Valence and arousal. Participants rated their emotional reactions in terms of valence and arousal to each of the games using 9-point pictorial scales. The valence scale consists of 9 graphic depictions of human faces in expressions ranging from a severe frown (most negative) to a broad smile (most positive). Similarly, for arousal ratings, there are 9 graphical characters varying from a state of low visceral agitation to that of high visceral agitation. The ratings are made by selecting a radio button below an appropriate picture. These scales resemble P. J. Lang's Self-Assessment Manikin [23].

2.5.3. Threat and challenge appraisals. Before each game, the degree of perceived threat that the game provided (i.e., anticipated threat) was assessed by asking participants, "How threatening do you expect the upcoming game to be?" [cf. 24]. After each game, subjective experience of challenge was assessed by asking, "How challenging was the game you just played?" Both items were rated on a 7-point scale, ranging from 1 (*not at all*) to 7 (*extremely*).

2.6. Physiological data collection

Electrocardiogram (ECG) was recorded from the main participant using the Psylab Model BIO2 isolated AC amplifier (Contact Precision Instruments, London, UK), together with three EKG leads in a modified Lead 2 placement. IBIs (ms) were measured with the Psylab Interval Timer.

The digital data collection was controlled by Psylab7 software, and the signal was sampled at a rate of 500 Hz.

2.7. Data reduction and analysis

Mean values for IBI were derived for each of the sixteen 30-s epochs during the games.

All data were analyzed by the General Linear Model (GLM) Repeated Measures procedure in SPSS. Analyses of the ratings measures data included two within-subjects factors, i.e., game (2 levels: Super Monkey Ball Jr. and Duke Nukem) and opponent (three levels: computer, friend, stranger). When analyzing IBI data, a third within-subjects factor was included, i.e., time (16 levels: the sixteen 30-s epochs). Two orthogonal contrasts were used to compare the appropriate levels of the opponent within-subjects factor: (a) computer vs. friend and stranger (Contrast 1) and (b) friend vs. stranger (Contrast 2).

3. Results

3.1. Threat and challenge appraisals

As hypothesized (Hypothesis 1), Contrast 1 indicated that playing against a friend or a stranger elicited higher anticipated threat compared to playing against a computer, *F*(1, 32) = 7.55, *p* = .010, η^2 = .19 (Figure 1, top panel).

In testing Hypothesis 2, post-game challenge ratings tended to be higher for the friend and stranger conditions compared to the computer condition, although Contrast 1 narrowly failed to reach statistical significance, *F*(1, 32) = 3.54, *p* = .069, η^2 = .10 (Figure 1, bottom panel).

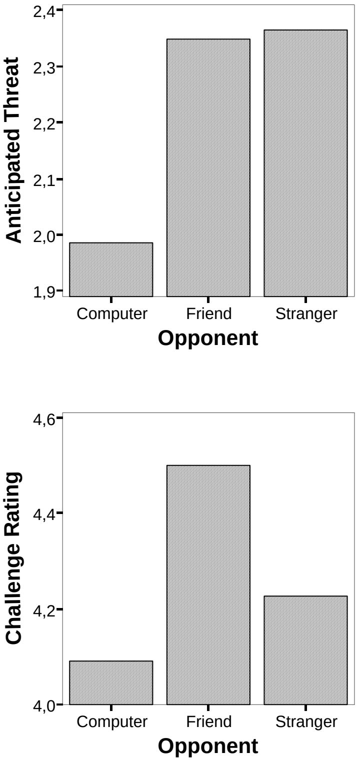


Figure 1 Anticipated threat (top panel) and challenge ratings (bottom panel) as a function of the opponent

3.2. Presence ratings

In agreement with Hypothesis 5a, Contrast 1 indicated that Spatial Presence was higher when playing with another human being (i.e., a stranger or a friend) compared to playing against the computer, *F*(1, 32) = 5.22, *p* = .029, η^2 = .14 (Figure 2, top panel). In addition, in testing Hypothesis 5b, Contrast 2 showed that playing with a friend elicited higher Spatial Presence compared to playing with a stranger, $F(1, 32) = 5.97$, $p = .020$, $\eta^2 = .16$.

In addressing Hypothesis 6a, Contrast 1 showed that playing with a human elicited higher Engagement than playing with a computer, $F(1, 32) = 17.83$, $p < .001$, $\eta^2 = .36$ (Figure 2, bottom panel). In addition, playing with a friend elicited higher Engagement than playing with a stranger (Hypothesis 6b), $F(1, 32) = 12.34$, $p = .001$, $\eta^2 = .28$.

We also tested the differences in Ecological Validity/Naturalness ratings between the two games. Ecological Validity/Naturalness was higher for Duke Nukem compared to Super Monkey Ball Jr., $F(1, 32) = 5.53$, $p = .025$, $\eta^2 = .15$.

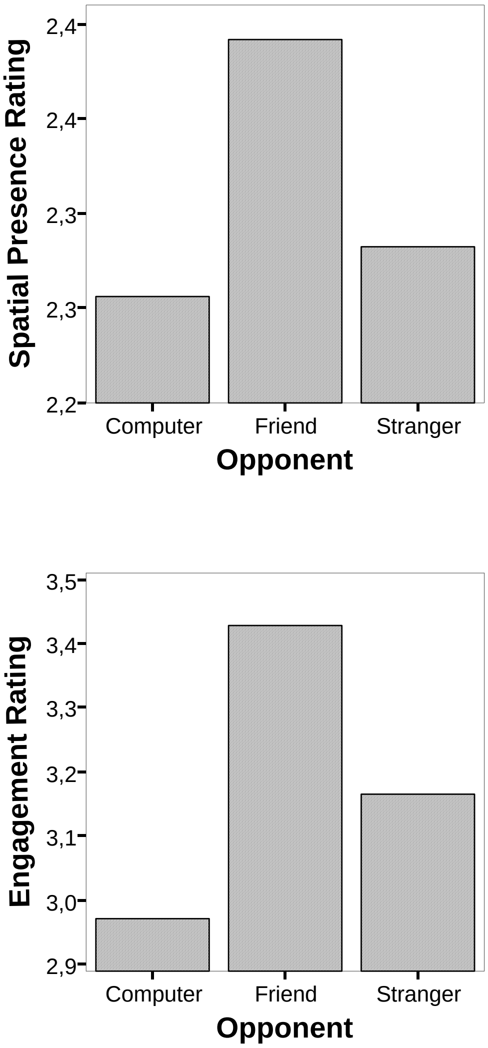


Figure 2 Spatial Presence (top panel) and Engagement (bottom panel) as a function of the opponent

3.3. Valence and arousal ratings

As suggested by Hypothesis 7a, playing against a human elicited a more positive emotional response compared to playing against a computer, for Contrast 1, $F(1, 32) = 24.19$, $p < .001$, $\eta^2 = .43$ (Figure 3, left panel). In testing Hypothesis 7b, although playing with a friend tended to elicit a more positive emotional response compared to playing with a stranger, Contrast 2 narrowly failed to reach statistical significance, $F(1, 32) = 3.53$, $p = .076$, $\eta^2 = .10$.

The Hypothesis 3a suggestion that self-reported arousal would be higher when playing against a human compared to playing against a computer was not supported by Contrast 1, $p = .191$; this was apparently due to low arousal ratings elicited by playing with a stranger. However, in agreement with Hypothesis 3b, Contrast 2 was significant indicating that playing with a friend elicited higher self-reported arousal than playing with a stranger, $F(1, 32) = 9.26$, $p = .005$, $\eta^2 = .22$ (Figure 3, middle panel).

3.4. Cardiac interbeat intervals

In agreement with Hypothesis 4a, Contrast 1 showed that cardiac IBIs were shorter (i.e., higher HR) when playing with a human compared to playing with a computer, $F(1, 31) = 27.20$, $p < .001$, $\eta^2 = .47$ (Figure 3, right panel). As hypothesized (Hypothesis 4b), Contrast 2 showed that playing with a friend elicited shorter IBIs (i.e., higher HR) compared to playing with a stranger, $F(1, 31) = 10.75$, $p = .003$, $\eta^2 = .26$.

4. Conclusions

In the present investigation, we investigated how the nature of the opponent (i.e., computer, friend, or stranger) influences Spatial Presence, emotional responses, and threat and challenge appraisals when playing video games. As hypothesized, the results showed that arousal ratings and physiological arousal as indexed by cardiac IBIs were higher when playing against another person (friend or stranger) than when playing against a computer (self-reported arousal was low when playing against a stranger, however). Apparently, the social-competitive situation related to playing against another person evokes increased arousal [cf. 12]. The presence of another person who "observes" inevitably creates a situation that involves high task performance evaluation potential, thereby increasing arousal [19]. This suggestion is also supported by the findings that playing against another person elicited higher anticipated threat prior to the game and higher post-game challenge ratings compared to playing against a computer. Threat appraisals have previously been associated with increased sympathetic arousal [25].

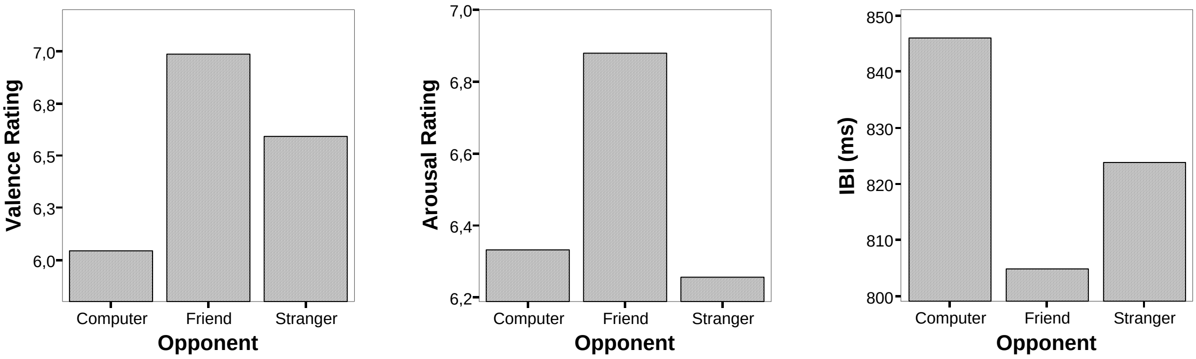


Figure 3 Valence ratings (left panel), arousal ratings (middle panel), and cardiac interbeat intervals (IBIs; right panel) as a function of the opponent

We also found that playing against a friend elicited greater self-reported arousal and shorter cardiac IBIs (i.e., higher physiological arousal) compared to playing against a stranger. This is likely because, as opposed to a stranger, a friend may serve as a continuing reminder of task performance [19]. Thus, it may be more important for a player to perform well when playing against a friend, which may result in increased arousal.

An important finding was that playing against another human being elicited higher Spatial Presence and Engagement as measured by the ITC-SOPI compared to playing against a computer. In addition, we found that playing against a friend elicited higher Spatial Presence and Engagement compared to playing against a stranger. Given that these differences in Presence were paralleled by arousal differences (see above), arousal may be a mediating factor. The two-level-model of Spatial Presence suggests that the focused allocation of attentional resources to the mediated environment contributes to Spatial Presence experiences [2]. Thus, given that arousal increases attention [17], increased arousal when playing against another person, and particularly when playing against a friend, may have contributed to increased Spatial Presence. It is also of note that, as a result of an increased attentional engagement with the game in the stranger and friend conditions, there may be less attentional resources left over for the processing of the signs that the game environment is artificial. Relevant to this, we have previously found that games played with a higher difficulty level elicit higher Spatial Presence (and arousal) compared to easier games (an exceedingly high difficulty level may, however, decrease Presence) [26]. A higher difficulty level is also likely to tax the cognitive resources, thereby diminishing attention paid to cues signaling that the game environment is not real.

We also found that playing against another human being elicited more positively valenced emotional responses compared to playing against a computer. This was expected, given the appetitive motivation of humans for social interaction [21]. This finding is also in line with the suggestion that high Presence conditions result in greater enjoyment [5, 15].

An apparent limitation of the present study was that the participants had to fulfill the relatively long questionnaire six times, which may have influenced the results (although we counterbalanced the six conditions). In the present study, both players were in the same room. However, video games are increasingly played over Internet or LAN, so that the players do not see each other. Thus, future studies should examine whether the present results replicate when the players are in different rooms, but have the knowledge with whom they are playing. Future studies might also examine the potential moderating effect of personality (e.g., sociability) on the associations found in the present study.

In sum, the present study showed that playing against another human elicited higher Spatial Presence, engagement, anticipated threat, post-game challenge appraisals, and physiological arousal compared to playing against a computer. In addition, playing against a friend elicited greater Spatial Presence, engagement, and arousal compared to playing against a stranger. The nature of the opponent influences Spatial Presence when playing video games, potentially through the mediating influence on arousal and attentional processes.

References

[1] Lombard, M., & Ditton, T. (1997). At the heart of it all: The concept of presence. *Journal of Computer Mediated Communication [Online]*, *3*(2). http://www.ascusc.org/jcmc/vol3/issue2/lombard.html.

[2] Wirth, W., Hartmann, T., Böcking, S., Vorderer, P., Klimmt, C., Schramm, H., Saari, T., Laarni, J., Ravaja, N., Gouveia, F.R., Biocca, F., Sacau, A., Jäncke, L., Baumgartner, T., & Jäncke, P. (2004). *Constructing presence: A two-level model of the formation of spatial presence experiences*. Manuscript submitted for publication.

[3] Steuer, J. (1995). Defining virtual reality: Dimensions determining telepresence. In F. Biocca & M. Levy (Eds.), *Communication in the age of virtual reality* (pp. 33-56). Hillsdale, NJ: Lawrence Erlbaum and Associates.

[4] Heeter, C. (1992). Being There: The subjective experience of presence. *Presence, 1*(2), 262-271.

[5] Lombard, M., Reich, R.D., Grabe, M.E., Bracken, C.C., & Ditton, T.B. (2000). Presence and television: The role of screen size. Human Communication Research, 26, 75-98.

[6] Witmer, B. & Singer, M. (1998). Measuring presence in virtual environments: A Presence Questionnaire. Presence: Teleoperators and Virtual Environments, 7, 225-240.

[7] Laarni, J., Ravaja, N., Saari, T., & Hartmann, T. (2004). Personality-related differences in subjective presence. In M. Alcañiz Raya & B. Rey Solaz (Eds.), Proceedings of the PRESENCE 2004 (pp. 88-95). Valencia, Spain: Editorial de la UPV.

[8] Lang, P.J. (1995). The emotion probe. Studies of motivation and attention. *American Psychologist, 50,* 372-385.

[9] Larsen, R.J., & Diener, E. (1992). Promises and problems with the circumplex model of emotion. In M. Clark (Ed.), *Review of personality and social psychology* (Vol. 13). Sage, Newbury Park, CA, pp.25-59.

[10] Scherer, K.R. (1993). Neuroscience projections to current debates in emotion psychology. *Cognition and Emotion, 7,* 1-41.

[11] Ravaja, N. (2004). Contributions of psychophysiology to media research: Review and recommendations. *Media Psychology, 6,* 193-235.

[12] Johnston, D.W., Anastasiades, P., & Wood, C. (1990). The relationship between cardiovascular responses in the laboratory and in the field. *Psychophysiology, 27,* 34-44.

[13] Ravaja, N., Saari, T., Kallinen, K., & Laarni, J. (in press). The role of mood in the processing of media messages from a small screen: Effects on subjective and physiological responses. *Media Psychology*.

[14] Meehan, M., Insko, B. Whitton, M., & Brooks Jr., F. P. (2002). Physiological Measures of Presence in Stressful Virtual Environments. *ACM Transactions on Graphics, Proceedings of ACM SIGGRAPH 2002, 21(3),* 645-653.

[15] Heeter, C. (1995). Communication research on consumer VR. In F. Biocca & M.R. Levy (Eds.), Communication in the age of virtual reality. Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 191-218.

[16] Ravaja, N., Laarni, J., Saari, T., Kallinen, K., Salminen, M., Holopainen, J., & Järvinen, A. (2004). Spatial presence and emotional responses to success in a video game: A psychophysiological study. In M. Alcañiz Raya & B. Rey Solaz (Eds.), Proceedings of the PRESENCE 2004 (pp. 112-116). Valencia, Spain: Editorial de la UPV.

[17] Gatchel, R.J., Gaas, E., King, J.M., & McKinney, M.E. (1977). Effects of arousal level and below-zero habituation training on the spontaneous recovery and dishabituation of the orienting response. *Physiological Psychology, 5,* 257-260.

[18] Warner, R.M., & Strowman, S.R. (1995). Cardiovascular reactivity and positive/negative affect during conversations. *Journal of Behavioral Medicine, 18,* 141-159.

[19] Thorsteinsson, E.B., & James, J.E. (1999). A meta-analysis of the effects of experimental manipulations of social support during laboratory stress. *Psychology and Health, 14,* 869-886.

[20] Christenfeld, N., Gerin, W., Linden, W., Sanders, M., Mathur, J., Deich, J.D., & Pickering, T.G. (1997). Social support effects on cardiovascular reactivity: is a stranger as effective as a friend? *Psychosomatic Medicine, 59,* 388-398.

[21] Aitken, K.J., & Trevarthen, C. (1997). Self/other organization in human psychological development. *Development and Psychopathology, 9,* 653-677.

[22] Lessiter, J., Freeman, J., Keogh, E., & Davidoff, J. (2001). A cross-media presence questionnaire: The ITC-Sense of Presence Inventory. *Presence: Teleoperators and Virtual Environments, 10,* 282-297.

[23] Lang, P. J. Behavioral treatment and bio-behavioral assessment: Computer applications. In J.B. Sidowski, J.H. Johnson, & T.A. Williams (Eds.), *Technology in mental health care delivery systems* (pp. 119-137). Norwood, NJ: Ablex, 1980

[24] Ravaja, N., Keltikangas-Järvinen, L., & Kettunen, J. (in press). Cloninger's temperament dimensions and threat, stress, and performance appraisals during different challenges among young adults. *Journal of Personality*.

[25] Tomaka, J., Blascovich, J., Kelsey, R.M., & Leitten, C.L. (1993). Subjective, physiological, and behavioral effects of threat and challenge appraisal. *Journal of Personality and Social Psychology, 65,* 248–260.

[26] Ravaja, N., Saari, T., Salminen, M., Laarni, J., Holopainen, J., & Järvinen, A. (2004). Emotional response patterns and sense of presence during video games: Potential criterion variables for game design. In A. Hyrskykari (Ed.), *Proceedings of the NordiCHI 2004* (pp. 339-347). New York: ACM Press.

Learning, experience and cognitive factors in the presence experiences of gamers: An exploratory relational study

David Nunez
Department of Psychology
University of Cape Town
dnunez@humanities.uct.ac.za

Edwin Blake
Collaborative Visual Computing Lab
University of Cape Town
edwin@cs.uct.ac.za

Abstract

This paper presents a large scale (N=101) exploratory relational study of computer gamers' experiences and habits related to presence. The study posited and examined the effect of two presence maximization strategies (controlling distracters and maintaining updated computer hardware) and two hypothetical cognitive styles (thematic inertia and capacity to integrate non-diagetic information) on gamers' rating of the importance of presence in games. The data show that frequency of game playing, but not game playing experience, affect self-rated presence importance, and that presence importance does not decline with experience. The data also suggest that presence maximization strategies are erratically effective in improving gaming experiences, and that the capacity to integrate non-diagetic information (but not thematic inertia) is a reliable predictor of self-rated presence importance.

1. Introduction

The role of cognition in presence has attracted some research attention in the past few years. The focus of this of research has been on isolating particular causal variables through experimental methods (for instance [1] or [2]). While this strategy has been fruitful, it is of limited use in producing new models for further investigation. This type of exploratory investigation is best done by examining natural relationships between large numbers of factors in large samples. This paper presents an exploratory relational study which was designed to examine the role of learning, experience and cognitive strategies to maximize presence in habitual users of virtual environments (VEs). For this study, we used computer gamers as the population, as we felt that this represented a large group of habitual VE users with a wide degree of variance in both VE usage experience and quality of presence experiences.

1.1 Time, experience and self-rated importance of presence

A number of experimental studies have found relationships between presence and various time and experience related factors such as age (e.g. [3]), game playing experience (e.g. [4]) and previous exposure to virtual environments (e.g. [5]). While such experimental studies have provided valuable insights, we feel that their ecological validity is limited (as is indeed the case for experiments in general [6]) by their use of a limited number of VE conditions from which they draw presence scores. This study thus aims to examine an average degree of presence during game playing in a general sense, in an effort to increase the generalizability of our findings. Rather than asking participants to report on any one recent presence experience, we asked them to rate how important they consider presence to be to their gaming (we call this *self-rated presence importance*). We reasoned that if gamers have enjoyable or compelling presence experiences while gaming, their ratings of presence importance will be higher. We measured presence importance by means of self-report items such as “A game should make me feel as if I am transported to inside the game world.” and “I prefer games which create a sense of being in a place”.

1.2 Presence maximization strategies

We conjectured that if it is possible to manipulate one's gaming environment and habits to maximize presence, then it is highly probable that gamers would have discovered and evolved these techniques on their own. Furthermore, if gamers have developed presence maximization strategies, then the use of such strategies probably varies with time-related factors (length of time playing, age, etc). We identified two possible presence maximization strategies from the literature which the average gamer could easily implement on a regular basis: minimizing attention distracters [7, 8], and improving display fidelity by maintaining up-to-date computer equipment [9, 10].

1.3 Cognitive styles and presence

We were interested in finding evidence of particular cognitive styles which affect presence. We hypothesized two possible factors: thematic inertia and the capacity to integrate non-diagetic information.

Thematic inertia is the term we use to describe the tendency of subjects to engage in thematically similar activities – for instance, after watching a film with a certain theme, a high thematic inertia subject might read a book with the same theme. As thematic inertia can be linked to schemata activation [11], we theorize that individuals in whom schemata activation degrades slowly will tend to show a higher degree of thematic inertia. Furthermore, if presence is associated with schemata activation (as argued in [1]), then it is reasonable to suggest that thematic inertia might be a correlate of presence.

The second cognitive factor is the capacity to integrate non-diagetic information. In film, the term 'non-diagetic' refers to information which does not emanate from the story world (e.g. background music or narration). According to constructionist models of presence (e.g. [7] or [12]), presence is a function of how information from various sources is integrated into a coherent whole. Although non-diagetic information reduces the fidelity or realism of a system, it is reasonable that, if it is cognitively integrated correctly, it could contribute to presence.

2. Exploratory study

The study was advertised as a 'computer gaming habits survey' to various computer science classes, and the survey itself was posted on-line. A total of 101 responses were collected over a one-week period. Only 3 respondents were women (2.97%). The mean age was 22.13 years (s=3.23), with a minimum of 17 and a maximum of 34.

2.1 Method

We created a 40 item instrument measuring 10 factors. 6 of these were time and learning related factors (see Table 1), and the other 4 were cognitive and experiential factors (see Table 2). For most items, a Likert-type response format was used.

Factor	No. items	Example Item
Length of time playing presence games	3	How long have you been playing first person shooters?
Frequency of playing presence games	3	How often do you play simulators?
Frequency of playing non-presence games	3	How often do you play fighting games?
Knowledge of computers	1	How much knowledge do you have about how computers work?
Knowledge of games	1	How much knowledge do you have about how computer games works?
Age	1	Your age:

Table 1: Time-related and learning factors.

2.2 Categorization of game types

For this study, we broadly divided computer games into two categories: those which aim to produce presence ('presence games') and those which do not ('non-presence games'). Presence games include among others simulators, role-playing games and first-person shooters, while non-presence games include real-time strategy, abstract puzzles and fighting games.

3. Results

3.1 Learning and experience effects

We conjectured that how important a player considers presence might be a function of learning or experience. We tested a linear regression model to predict the self-rated importance of presence in games using all six time-related factors as predictors (F=2.78; df=6, 66; $p<0.017$; $R^2=0.202$). By examining the partial regressions to control for inter-variable dependencies, we found the only significant predictor to be frequency of presence game playing (partial r=0.351; t(66)=3.04; $p<0.0033$). When we examined each of the six items composing the self-rated importance of presence factor, we found that one item ("The quality of a game's sounds are very important for my game experience.") was also inversely predicted by frequency of non-presence game playing (partial r=-0.25; t(66)=-2.11; $p<0.037$).

Factor	No. Items	Example Item
Integration of non-diagetic information	5	Inappropriate music in a game can ruin the game experience for me.
Self-rated importance of presence	6	A game should make me feel as if I am transported to inside the game world.
Thematic inertia	6	After watching a TV program or film, I often feel like playing a game that is similar to the film or program.
Presence maximization	6	When I play, I turn off the lights and try to keep the room dark.

Table 2: Cognitive and experiential factors

Only one item ("For me, the most important aspect of game playing is the ability to explore other worlds."), was not predicted by time-related factors at all. The lack of time or learning effect on this item is probably attributable to the wording of the item. Although some players may enjoy exploring game worlds (a high-presence activity), most games make exploration a secondary activity – the player's primary goals (winning a fight, solving a puzzle, etc.) are often non-presence activities.

3.2 Learning to maximize presence

We examined the role of time and learning related factors in players' presence maximization strategies. Again, a multiple regression analysis was computed with all six time related factors as predictors for presence maximization strategies (F=2.83; df=6,66; $p<0.016$; $R^2=0.204$). Only knowledge of the workings of computer games was a significant predictor (partial r=-0.311; t(66)=-2.66; *p*<0.0097). Interestingly, the partial correlation shows that higher knowledge of game workings is associated with reduced efforts to manage presence.

This finding suggests that gamers who understand games more (and presumably the reliance of modern games on specialized computer hardware) would at least make an effort to keep their equipment up to date. We hypothesized that maintaining updated computer equipment may be beyond the economic reach of our sample of university students, and this would thus confound the finding. We found evidence of this when comparing the two items "As far as I can afford it, I make sure my computer has the best hardware for playing games." and "I will consider upgrading my computer to play a particular game." How long the players had been playing presence games was indeed a significant predictor for this second item (partial r=0.28; t(66)=2.39; *p*<0.019), but not for the first.

For the distraction related items, there were good indications that time-related factors play a role. For the item "If I am disturbed while I am playing, it ruins the experience for me.", both age (partial r=0.25; t(66)=2.057; *p*<0.043) and how long the player had been playing presence games (partial r=0.29; t(66)=2.52; *p*<0.014) were significant predictors.

3.3 Effectiveness of presence maximization strategies

We were interested in the extent to which players' strategies for maximizing their presence were effective. Again, we used a multiple regression analysis with the presence maximization factors as a predictor for self-rated presence importance. The subsequent model was significant, although it explained only a small amount of the dependent variable's variance (F=18.87; df=1, 99; *p*<0.0005; $R^2 = 0.15$).

An item-by-item investigation of the self-rated presence importance factor showed that only two of the six items in the factor failed to show this pattern. The items "The quality of a game's sounds are very important for my game experience." and "I prefer games which create a sense of being in a place" were not predicted by presence maximization strategies.

3.4 Cognitive factors and time/learning effects

We first examined relationships between our two cognitive factors (thematic inertia and the capacity to integrate non-diagetic information) and the six time-related factors. As the rate of schemata activation and decay is probably set at an early age and changes little over time [11] we expected no time effects on thematic inertia. Using a multiple regression analysis with the six time factors as predictors, we indeed found no significant effect on thematic inertia (F=0.89; df=6, 66; *p*<0.505).

For the integration of non-diagetic information factor, the picture is theoretically more complex. Some theorists propose that this integration task is not innate, but learned as one becomes more literate in decoding the medium [8, 13]. If this is true, then we expect to see learning effects. We did indeed find a significant effect. A multiple regression on capacity to integrate non-diagetic information with time factors as predictors was significant (F=2.42; df=6, 66; *p*<0.036; R^2=0.18). Of the six time factors, only length of time playing presence games is significant (partial r=0.365; t(66)=3.191; *p*<0.002). Although this result can be interpreted as supporting a 'learning to decode' hypothesis, it is also possible that those subjects who are better able to integrate non-diagetic information tend to have a more enjoyable presence experience during gaming and thus keep playing this type of game for longer periods.

3.5 Cognitive factors as predictors of self-rated presence importance

The two cognitive factors (thematic inertia and the capacity to integrate non-diagetic information) show a significant correlation with each other (r=0.36; n=101; p < 0.01). This supports the notion that they share some common cognitive basis. To determine if these cognitive factors are related to presence, we used them as predictors for self-rated presence importance in a multiple regression analysis. This gives a significant model (F=12.49; df=2, 98; *p*<0.0001; R^2=.202). In this model, only integration of non-diagetic information is a significant predictor (partial r = 0.34; t(98)=3.56; *p*<0.0005). When we examined the effect of thematic inertia on self-rated presence importance on an item-by-item basis (controlling for the integration of non-diagetic information), we found it to be a significant predictor of two items: "I prefer games which create a sense of being in a place." (partial r=0.29; t(78)=2.45; *p*<0.016; R^2=0.16) and "For me, the most important aspect of game playing is the ability to explore other worlds." (partial r=0.25; t(78)=2.32; *p*<0.022; R^2=0.16).

4. Discussion

4.1 Learning and experience in presence

Although this is only an exploratory study and cannot show causation, the data show some interesting trends with regard to experience in VEs and cognition in presence. Firstly, it seems that the most reliable time or learning related predictor of how important players consider presence in gaming to be, is the proportion of their gaming time spent playing presence games. It seems that presence displays a slow-decay effect: one presence experience leads to the desire to have another (this is supported in part by the positive relationship between thematic inertia and some of the presence importance items). Then, if no gaming occurs for a period, the benefit decays (this is indicated by the fact that while frequency of presence game playing is positively associated with presence, length of time having played presence games does not). The data does not seem to indicate that users become desensitized to the presence experience over time. This is inferred from the general lack of effect of the length of time playing presence games.

Indeed, the opposite may be true, as age has a weak positive effect on self-rated presence importance.

4.2 Presence maximization strategies

With regards to presence maximization, the data suggest that gamers do successfully engage in strategies to maximize their presence. Interestingly, these efforts generally vary (inversely) with knowledge of how games work. We propose two explanations for this phenomenon: one is that as gamers' knowledge about the technical aspects of the game interferes with their ability to suspend their disbelief during play. The other is that all gamers have naïve theories of how presence 'works', but more experienced gamers (who probably obtain most of their knowledge from gaming websites and gaming magazines) believe the common game marketing line that the software is largely responsible for presence, and thus make no effort to control their own environment. We would need to explicitly tap into these naïve theories to validate this hypothesis. Regardless of what gamers believe about the causality of their presence experiences, it seems that the presence maximization techniques do have some effect, although with very little consistency.

These findings may be partly obscured by economic factors which we did not take into account. One of the two presence maximization strategies we measured was the maintenance of up-to-date computer hardware. It is likely that the gamers in our sample would like to buy the newest hardware, but as almost all were university students, their economic realities would interfere. Evidence for this comes from the comparison of the item which measures real money expenditure (in which no time effect was found), with the item which measured hypothetical expenditure (for which length of time playing presence games was a predictor). This implies that long-time players of presence games recognize the importance of maintaining updated hardware, but may not always be capable of doing so in practice.

4.3 Cognitive styles in presence

We found some evidence of cognitive styles associated with self-rated presence importance, although it is not clear if these develop through playing presence games, or if their prior existence leads to an increase in playing presence games. Of the two cognitive factors we examined, the ability to construct coherent presence experiences from both diagetic and non-diagetic information sources seemed to be the most important to presence experiences. As this capacity improves with presence game playing experience, it seems that in general presence game experience (while controlling for age) leads to more presence. This corresponds to some extent with the positive age/presence relationship reported by [3].

4.4 The role of sound in presence experiences

An interesting finding arises from the data about the importance of sound which is worth mentioning. Players' ratings of the importance of sound to the game experience were not linked to presence management strategies (while ratings of the importance of graphics were). However, the importance of sound was strongly linked to frequency of presence game playing. This may imply that the integration of sound into the presence experience is not affected by a player's efforts, but does improve with repeated exposure. This may suggest that the contribution of sound to presence is processed separately from other modalities, as suggested by [8].

References

[1] D. Nunez and E. H. Blake, *Conceptual priming as a determinant of presence in virtual environments.*, *2nd International Conference on Computer Graphics, Virtual Reality, Visualisation and Interaction in Africa (AFRIGRAPH 2003)*, Cape Town, South Africa, 2003, pp. 101-108.

[2] S. Brown, I. Ladeira and C. Winterbottom, *The Effects of Mediation in a Storytelling Virtual Environment.*, *International conference on storytelling 2003*, 2003, pp. 102-111.

[3] C. Youngblut and B. T. Perrin, *Investigating the relationship between presence and task performance in virtual environments.*, *IMAGE 2002 Conference*, Scottsdale, AZ., 2002.

[4] M. K. A. Usoh, M. Whitton, R. Bastos, A. Steed, M. Slater and F. P. Brooks, *Walking > Walking-in-place > flying, in virtual environments.*, *Computer Graphics Annual Conference (SIGGRAPH99)*, Los Angeles, CA., 1999, pp. 359-364.

[5] M. Meehan, B. Insko, M. Whitton and F. P. Brooks, *Physiological measures of presence in stressful virtual environments.*, *SIGGRAPH 2002*, San Antonio, USA, 2002, pp. 645-652.

[6] R. Rosenthal and R. L. Rosnow, *Essentials of behaviour research: Methods and data analysis.*, McGraw Hill, New York, NY, 1991.

[7] W. Wirth, P. Vorderer, T. Hartmann, C. Klimmt, H. Schramm and S. Böcking, *Constructing presence: Towards a two-level model of the formation of spatial presence experiences*, *6th Annual International Workshop on Presence*, Aalborg, Denmark, 2003.

[8] D. Nunez, *Working memory and presence: Reconsidering the role of attention in presence.*, *7th International workshop on presence (PRESENCE 2004)*, Valencia, Spain, 2004, pp. 44-47.

[9] B. G. Witmer and M. J. Singer, *Measuring presence in virtual environments: A presence questionnaire.*, Presence: Teleoperators and Virtual Environments., 7 (1998), pp. 78-89.

[10] W. A. IJsselsteijn, H. De Ridder, J. Freeman and S. E. Avons, *Presence: Concept, determinants and measurement.*, *Proceedings of the SPIE*, 2000, pp. 520-529.

[11] D. E. Rumelhart and A. Ortony, *The representation of knowledge in memory.*, in R. C. Anderson, R. J. Spiro and W. E. Montage, eds., *Schooling and the Acquisition of Knowledge*, Lawrence Erlbaum Associates, 1977.

[12] D. Nunez, *A constructionist cognitive model of presence.*, *7th International workshop on presence. (PRESENCE 2004)*, Valencia, Spain, 2004, pp. 313-316.

[13] D. Bordwell, *A case for cognitivism.*, Iris, 9 (1989), pp. 11-40.

Being There Together and the Future of Connected Presence

Ralph Schroeder
Oxford Internet Institute, University of Oxford
ralph.schroeder@oii.ox.ac.uk

Abstract

Research on virtual environments has provided insights into the experience of presence (or being there) and copresence (being there together). Several dimensions of this experience, including the realism of the environment and of the avatar embodiment, have been investigated. At the same time, research on a number of new media has begun to use concepts that are similar to copresence, such as mutual awareness, connected presence, and engagement. Since digital environments can be reconfigured and combined easily, and since an increasing number of such environments are used to connect people in their everyday lives, it is useful to think about the various modalities of connected presence as a continuum – with shared virtual environments in which people are fully immersed as an end-state. This paper proposes a model for the different modalities of connected presence whereby research on shared virtual environments can be integrated with research on other new media - and vice versa. It is argued that this model can improve our understanding both of the uses of shared virtual environments and of their future development among a variety of media for 'being there together'.

Keywords: virtual environments, presence, copresence, computer-mediated communication.

1. Shared Virtual Environments as an End-state

Shared virtual environments (SVEs) have made it possible for people to experience 'being there together' in the same computer-generated space. The experiences of 'presence' in a virtual environment and 'copresence' with other people have been explicated in a number of studies. At the same time, a number of studies of new media technologies have begun to use concepts of 'presence' and 'copresence' and related concepts such as 'awareness', 'engagement' and the like. These media include mobile telephones, instant messaging, and online games. The main aim of this paper is to relate research on virtual environments to research on new media and to ask, what can we learn about SVEs learn from other new media, and vice versa?

A useful way to do this is to think of SVEs as an end-state – a purely mediated relationship in which the user of SVE technology experiences copresence with others in a fully immersive environment. Various technologies are now available whereby users and environments are represented to each other in fully immersive displays, either in the form of computer-generated embodiments and scenes, or in the form of the 3D video capture of people and scenes. Despite current technical limitations, these immersive displays represent an end-state in the sense that – barring direct sensory input into the brain (in the manner of science fiction novels such as William Gibson's *Neuromancer* and Neal Stephenson's *Snow Crash*), synthetic environments for 'being there together' that are displayed to the users' senses cannot be developed further than fully immersive VEs. Nevertheless, even fully immersive SVEs will, like other new media, have certain possibilities and constraints. It is argued here that relating these possibilities and constraints of SVEs to other media will provide us with a bettter understanding of technologies for being there together and their potential future uses.

It is proposed here that SVEs and other new media should be seen as varying on three dimensions: presence, copresence, and the extent of one's connected presence (the term 'connected presence' was coined by Licoppe [1]; this concept will be explained in the following section). The third dimension, as we shall see, captures a number of different elements, but the main reason for this dimension is that we not only want to know about presence and copresence in abstract terms (the experiential state of the user at a particular point in time), but also in terms of the actual *extent* to which our relationships are mediated in this way. This yields a 'connected presence' cube (see figure 1 at the end of the paper).

The next section of this paper will elaborate the connected presence cube. A longer version of the paper will give an overview of the relevant findings about 'presence', 'copresence' and connected presence, and also compare SVEs with other media in relation to these three dimensions. The concluding section spells out the lessons we can learn from an integrated model of connected presence and how these can inform the design of SVEs.

2. Presence, Copresence, and Connected Presence

Research on VEs has produced a range of studies about 'presence' and to a lesser extent about 'copresence'. There are still debates about how to define and measure presence and copresence. Here it is not necessary to go into these debates in detail (for overviews, see [2, 3]). It is, however, important to provide a precise definition of SVEs which will allow us to compare them with other media: virtual environments provide ‘the user(s) with the sensory experience of being in a place other than the one you are physically in, and being able to interact with that place’, or simply ‘being there’ [4,5]. Copresence can then be defined as ‘being there together’.

Shared VEs have three dimensions (x,y,z), which can be represented as being related to each other. On all three dimensions, we can take the individual’s presence in a real physical environment and a face-to-face encounter as our starting point (0,0,0). On the first dimension, being in physical world is at one end of the y axis and having a sense of being there (alone) in a purely media-generated place is at the other end of the end (0,1,0). This dimension is discussed in virtual environments research under the rubric of ‘presence’ or ‘being there’. On this dimension, highly immersive environments such as Cave-type [6] environments are at the top end of the y axis (0,1,0), but simulators and IMAX screens also provide the user with the experience of ‘being there’(though with limited possibilities for interacting with the environment).

On the second dimension, again with our point of departure face-to-face encounters in the physical world at one end, mediated relations with persons whom we encounter only virtually are at the other end (1,0,0). In virtual environments research, this is called ‘copresence’, but it could equally be called ‘being there together’. Telephones minimally provide us with this sense, though they lack the spatial compoment (not entirely, as we shall see), with instant messaging providing more of a spatial sense of copresence. So these two technologies are somewhere along the continuum of copresence, with the telephone providing some experience of copresence (>1,0,0) and instant messaging a somewhat spatial experience of copresence (>1,>1,0).

‘*Completely*’ mediated relationships then constitute a third dimension (the z axis). This is the *extent* to which one’s relationships are mediated through environments in which presence and copresence are experienced. This dimension has several subcomponents: the ‘affordances’ or ‘constraints’ of the mediation, the extent to which one’s relationships with others are *exclusively* mediated in this way, and third and finally the extent of time spent in these mediated encounters compared with one’s face-to-face relationships. Together these constitute ‘connected presence’ or the extent to which ‘being there together’ is mediated. Once we add this third dimension, some everyday technologies like the telephone will receive a much higher value for this dimension (0,>1,>1) than SVE systems which typically have a low value for this dimension.

2.1. The End-State of SVEs and the Third Dimension

These three dimensions allow us to picture SVEs with completely immersive networked VR systems - systems in which the user exclusively has a sense of being there with others - as an end-state. This end-state is one in which in which users would live entirely inside immersive virtual worlds (1,1,1), and this allows us to plot all experiences of connected presence as approximations towards this end-state (see figure 2 at the end of this paper).

Before we elaborate and compare these experiences further, however, three points need to be made about figure 2: Of course it is true that all forms of mediated environments only complement – and do not replace - physical, face-to-face environments and relationships. Here, however, the focus is on *mediated* relationships. The balance between mediated relationships and face-to-face relationships in the physical world will be discussed below. The point of envisioning living together in virtual worlds is that – as we shall see – this will provide a useful model to think about and study SVEs and other media.

Another problem is that this plotting exercise is highly imperfect: the extent to which people experience a sense of being there with others in, say, telephone conversations, online chat rooms, and different types of virtual reality systems will vary considerably according to context. As long as we bear this variation in mind - an easy solution would be to plot areas of various sizes rather than points on the three axes – these three dimensions will allow us to make useful comparisons.

Being there together in different SVEs will vary considerably on the first two dimensions. One reason to go beyond these two dimensions and add comparisons on the third dimension is that the end-state of the first two dimensions (remembering that this is a single point in time) will be influenced by the third; in other words, presence and copresence will be affected by the extent of experience with the medium.

Some brief examples can illustrate this point: One is that users must learn to cope with the other person’s avatar - sometimes it is easy to walk through another person, at other times users will maintain interpersonal distance to a similar extent as in face-to-face encounters. This depends on the type of SVE system used (see the comparison of three systems in [7]) but also, in immersive SVEs, on the stage of the task people are in, or how habituated to interacting with an avatar they have become [8]. Note that presence and copresence are inescapably affected by ‘connected presence’ – whether one walks through or maintains a conventional face-to-face distance from another avatar is bound to influence the experience of being in the environment and interacting with an avatar.

Another example from the same immersive SVE trial is that users point out objects to the other person with an untracked arm or they ‘lean’ to hear the person even when there is no spatial sound; yet at other times, they use the devices appropriately [8]. Again, this depends on the amount of time they have spent on the task and how ‘used to’ the system they have become.

Similar phenomena can be identified for other new media. For example, people can treat places at the other end of a mobile phone conversation as if they were sharing the remote space – as when they gesture to the other person (even though the gesture cannot be seen) [9]. Or again, instant messaging (IM) can, with routine use, create the sense of the other person’s copresence in the sense that people will treat IM as a shared space in which people can step and out of each other's awareness.

Another example is when, in networked immersive projection technology (IPT) systems, people use their bodies as reference point in interacting with objects, using verbal and non-verbal communication to do spatial tasks together. They need more verbal communication in networked desktop systems for the same task because they need to describe in words where they would otherwise have used gestures and their bodies [10]. Again, this takes getting used to in both cases. Notice again that people also do this in mobile phone conversations, for example giving an indication of their location to let their partner know how they are coping with the space around them [9]. Or, to take a non-spatial example, the absence of eye gaze to indicate who one is speaking to can be compensated for in both telephone and SVE situations by means of words (or in SVE's also by gestures, see [11]).

2.2. Two End States of Being there Together

SVE technologies range from immersive projection technology systems or IPTs (also known as Cave-type displays) and head-mounted displays to desktop systems. Two types of technologies currently occupy the furthest points on the dimensions of presence and copresence (1,1,0): Networked IPT systems that display computer-generated avatars and spaces, and environments that allow users to share the same 3D video space with video avatars (blue-c is currently the only example of the latter, see [12]).

The difference between video- 3D environments (essentially holographic videoconferencing systems) versus computer-generated 3D environments is important for the discussion to follow and therefore deserves to be spelled out: Both are end-states of people completely immersed in mediated communication environments interacting with each other, but they have quite different capabilities: video environments *capture* the appearance of real users and real places, while virtual environments *generate* user representations (avatars) and virtual places or spaces. The two technologies also allow the user to *do* different things: video environments are realistic and are constrained by this realism, virtual environments allow manipulation but they do not capture real scenes. The two environments therefore represent two quite *different* end-states – even though *both* are on the same top right hand corner in figure 1 terms of presence and copresence (1,1,0).

To appreciate the difference between these two immersive VEs, picture your body (and those of others), as well as the real place around you, captured by cameras and reproduced in full - and now add the fact that, although this capturing has been done digitally, the digital environment of 3D video images is designed such that objects (including people) can only behave according to the laws of the physical world. In other words, this is a 3D videoconferencing scenario in which the space around the users is included.

Now picture, by contrast, your body controlling a computer-generated avatar along with other such avatars in a computer-generated environment - the appearance and behaviours of which are unconstrained by real-world laws (for example, flying around together). Note that the difference between the two scenarios is not just 'realism', but also what control is exercised over one's body – is it captured or tracked? – and over the environment – are objects captured or can they be manipulated? The Rubik's cube task, for example, which involves collaboratively putting together cubes that a suspended in space and that snap together (described in [8]), would be impossible to implement in a video-captured environment. (In fact, the two endstate scenarios may be mixed in practice – for example, capturing the user on video but putting them into a computer-generated environment, or putting a computer-generated avatar into a video-captured environment - but in their pure forms they are quite different.)

If they are fully realized in the way described here, they are also, as mentioned earlier, the furthest possible extensions of technologies for 'being there together' or of shared synthetic environments - since no conceivable system could go beyond providing a more fully immersive experience of being there together (perhaps, again, neuro-physiological 'mind-melting' is conceivable, but this falls outside the definition of displays for the senses). Mixed or augmented reality devices, where the user is partly inside a VE and partly engages with the physical world, will constitute approximations to these two ideal end-states.

It is important to emphasize that the experience of presence is a sensory one – primarily visual and also audio (and sometimes haptic). This is important because there are debates about whether media which do not afford sensory experiences of another place or person – a book, say, or a text-based MUD – can be discussed in the same context as VEs (see the discussion in [13]). This is ruled out by the definition of VEs given earlier: unless the experience is a sensory one, one based on perception of a place or person via our sensory apparatus, the experience 'mediated' by books and the like is excluded. Thus a complete end-state will provide an environment for being there together for *all* the senses, but since sensory inputs and outputs apart from vision, sound and haptics (such as smell and taste) are rather remote, we can concentrate on the audio-visual environments that are currently available.

3. Shared Virtual Environments, the Multiple Modes of Connected Presence, and the Future of Mediated Relationships

SVEs can be compared to other environments for 'being there together' which raise issues pertaining to the immersiveness and interactivity of graphical plus audio environments (again, interfaces for the other senses could be mentioned here, but interactive and immersive graphics with audio is the most common type of VE system and environment). And they allow us to compare an end-state of full and constant immersiveness with various other conditions of connected presence. SVEs can thus be used to investigate a range of communications conditions along the presence, copresence and connected presence dimensions. The end-state of SVEs represent a valuable research tool for the study of the role of (computer-)

mediated communication in society. In addition, this end-state can be used to advance social science research, with experiments in SVEs that are difficult or impossible in face-to-face situations because various conditions of presence and copresence can be manipulated [14,15]. ('Manipulating conditions' may bring to mind social psychology, but it needs to be remembered that all kinds of conditions can be manipulated in SVEs, such as the means by which users can contact each other, how they can shape the built environment, etc.). In short, they offer a laboratory for studying face-to-face encounters and other media by allowing an array of conditions towards an end-state.

What brings all the issues around the different types of presence together into a coherent whole, from the point of view of taking mediated relationships rather than face-to-face encounters in the physical world as the baseline, is the focus of attention *inside* the environment (exclusively, away from the physical world and its face-to-face encounters) – which consists of the forms of attention on the other person(s) or mutual focus on one side - and on the environment on the other. And this focus can be on seeing or hearing – the environment and the other person(s). But the focus can also be on what you can *do* in the environment, and *do* there together – how one can interact with each other and with the environment [5].

This notion of interaction, however, is too passive for gauging connected presence. What is also needed is a more active notion of how relations can be maintained – or how they are enabled and constrained – in different media. Apart from the control over the immediate activity or what holds ones' attention, we could ask about the extent to which people have control over the environment in different media or mediated environments - how much they can be modified, what control over their appearance users have, what level of interactivity the displays and tools provide, and the like (all these have already been mentioned in passing.) And we should add the nature of the relationships – their depth, which encompasses the extent in time and the immediacy or exclusivity - that these media afford for 'being there together' and for making the environment one's own.

Debates about our 'mediated' relationships with others have arisen previously in relation to new media. Recently, the debate has been about whether the internet contributes to fewer offline relationships and the like [16]. If we think of these debates in terms of copresence and connected presence, they can be put into perspective: it is not that purely mediated interpersonal relations should be seen as causing loneliness or being inferior to face-to-face relations and the like; rather, different media provide different possibilities for being there together in the changing landscape of interpersonal connected presence.

Relationships are thus shaped not only by the 'medium', but by its 'affordances'. And, these affordances apply not just to the relationship with people, but also relationships to the environment and our control over it. Even if, as mentioned earlier, our relations in these media technologies should be described in terms of areas rather than as points on the three axes in the two figures, certain technologies and their uses nevertheless remain clustered in particular areas in relation to each other. This is an obvious point, but one that is not often made (Hutchby [17] is an exception): different technologies provide different constraints and possibilities for 'being there together', and if we put these on our three axes, we can begin to see what the futures of different media might look like.

This leads to what is perhaps the most comprehensive question that can be raised in relation to the intersection between the three dimensions of presence: Given that our relationship to the world mediated is by information and communication technology, what affordances, physical and social, do the various technologies for 'being there together' provide? This is the question to which the end-state presented here can begin to give some interesting answers. The end-state of SVEs points to a particular form of the mediation of our physical and social worlds and particular forms of living in immersive virtual worlds. If, however, we do not take face-to-face relationships as a baseline but approximations to this end-state, then we can ask: what do SVEs, in contrast with other less immersive relations, 'afford'? How do the levels of immersiveness and togetherness compare – with each other, rather than compared with face-to-face relations in the physical world?

Many SVEs provide a rich modality for 'being there together' compared to other media and they offer more control. Yet, as can be seen in studies of related media [1,18], other media also provide a strong sense of mutual awareness and availability on an everyday basis. With the changing landscape of mediated relationships and new media technologies, the line between SVEs and other new media technologies (which often include images and sounds of the other person and of the environment) that are shared over interpersonal networks are becoming increasingly blurred. Hence a research programme will be required which takes SVEs beyond the laboratory and early uses, and beyond online gaming and social spaces, and put 'being there together' into the context of our multiple modes of connectedness in everyday settings.

The connected presence cube allows us to do this; to see individuals connected to others via various communication and interaction modalities, with face-to-face communication only one among other possibilities. People are either immersed in the physical world or in the virtual world, stepping in and out of these constantly, and sometimes participating in several such worlds, limited only by the fact that sensory attention needs to be focused on a limited set of people and features of the environment, which makes multiple simultaneous channels (communicative multitasking) difficult. Increasing communication means that we are continuously connected to others who are aware of our presence and copresence to a greater or lesser extent. If we think of the multiple devices for connected presence that we use constantly throughout the day, it is possible to see that we need to manage our accessibility, mutual awareness and focus of attention continuously with different affordances (or constraints and possibilities) in different technologies for mediated interaction. The design of SVEs should therefore be informed by how best to combine different levels of presence, copresence and connected presence in our everyday lives.

References

[1] Licoppe, C. ‘Connected’ presence: the emergence of a new repertoire for managing social relationships in a changing communication technoscape. In *Environment and Planning D: Society and Space*, vol. 22, 135-156. 2004.

[2] Scheumie, M.J., van der Straaten, P., Krijn, M, van der Mast, C. Research on Presence in Virtual Reality: A Survey. In *Cyberpsychology and Behaviour*, 4(2): 183-201, 2001.

[3] Biocca, F.; Harms, C. and Burgoon, J.K. Toward a More Robust Theory and Measure of Social Presence: Review and Suggested Criteria. In *Presence: Journal of Teleoperators and Virtual Environments*, vol.12, no.5, 456-80, 2003.

[4] Ellis, S. Origins and Elements of Virtual Environments. In W. Barfield and T. Furness (eds.) *Virtual Environments and Advanced Interface Design*. New York: Oxford University Press, 14-57. 1995.

[5] Schroeder, R. Social Interaction in Virtual Environments: Key Issues, Common Themes, and a Framework for Research. In R. Schroeder (ed.), *The Social Life of Avatars: Presence and Interaction in Shared Virtual Environments.* London: Springer, 1-18, 2002.

[6] Cruz-Neira, C., Sandin, D., DeFanti, T. Surround-screen projection-based virtual reality: the design and implementation of the CAVE. *Computer Graphics Proceedings, Annual Conference Series 1993*. ACM SIGGRAPH, New York: ACM, 135-42, 1993.

[7] Becker, B. and Mark, G. Social Conventions in Computer-mediated Communication: A Comparison of Three Online Shared Virtual Environments. In R. Schroeder (ed.), *The Social Life of Avatars: Presence and Interaction in Shared Virtual Environments*. London: Springer, 19-39, 2002.

[8] Steed, A. et. al. Strangers and Friends in Caves: An Exploratory Study of Collaboration in Networked IPT Systems. *ACM SIGGRAPH 2003 Proceedings on Interactive 3D Graphics*. New York: ACM Press, 51-54, 2003.

[9] Ling, R. *The Mobile Connection: The Cell Phone’s Impact on Society*. Morgan Kaufmann. 2004.

[10] Heldal, I., Steed, A., Spante, M., Schroeder, R., Bengtsson, S. & Partanan, M. Successes and failures in copresent situations. In *Presence: Teleoperators and Virtual Environments*, 14, 5, forthcoming, 2005.

[11] Brown, B., Bell, M. Play and Sociability in There: Some Lessons from Online Games for Collaborative Virtual Environments. In R. Schroeder and A.-S. Axelsson (eds.). *Avatars at Work and Play: Collaboration and Interaction in Shared Virtual Environments.* London: Springer, 2005.

[12] Gross, M. et al. blue-C: A Spatially Immersive Display and 3D Video Portal for Telepresence. In *Proceedings of ACM SIGGRAPH*, July, 819-827, 2003.

[13] Klimmt, C. & Vorderer, P. Media Psychology ‘is not yet there’: Introducing Theories on Media Entertainment to the Presence Debate. In *Presence: Journal of Teleoperators and Virtual Environments.* 12 (4), 346-359. 2003.

[14] Blascovich, J. Social Influence within Immersive Virtual Environments, in R. Schroeder (ed.), *The Social Life of Avatars: Presence and Interaction in Shared Virtual Environments.* London: Springer, 127-145, 2002.

[15] Bailenson, J. and Beall, A. Transformed Social Interaction: Exploring the Digital Plasticity of Avatars. In R. Schroeder and A.-S. Axelsson (eds.). *Avatars at Work and Play: Collaboration and Interaction in Shared Virtual Environments.* London: Springer, 2005.

[16] Baym, N. Interpersonal Life Online. In L. Lievrouw and S. Livingstone (Eds.), *The Handbook of New Media.* London: Sage, 62-76. 2002.

[17] Hutchby, I. *Conversation and Technology: From the Telephone to the Internet.* Cambridge: Polity, 2001.

[18] Nardi, B., Whitaker, S. and Bradner, E. Interaction and Outeraction: Instant Messaging in Action. In *Proceedings of CSCW’00*, Dec.2-6. Philadelphia PA.: ACM, 79-88, 2000.

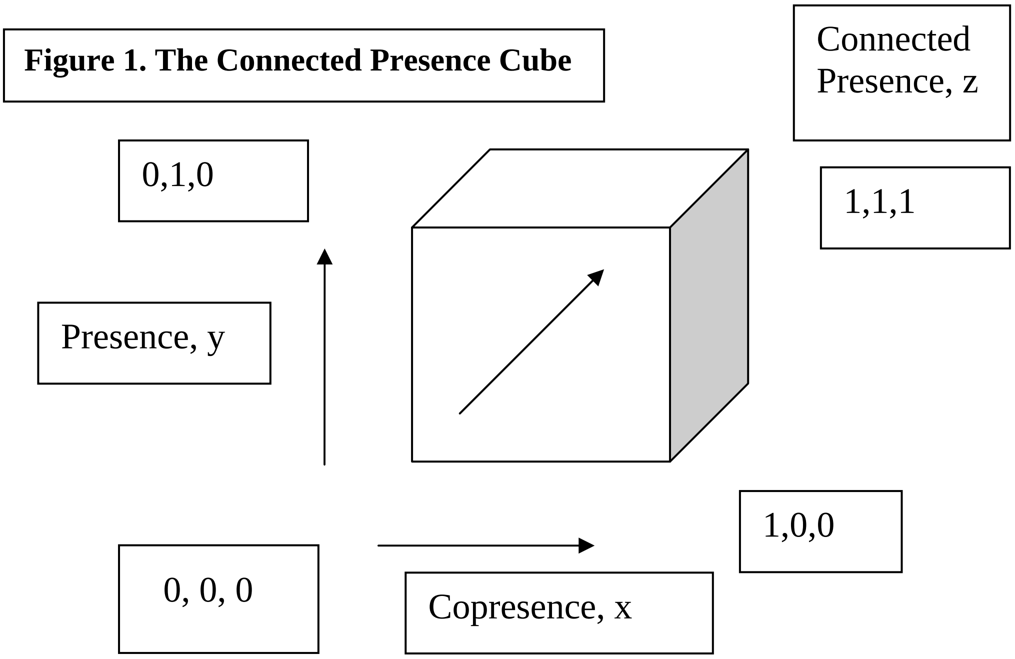


Figure 1. The Connected Presence Cube

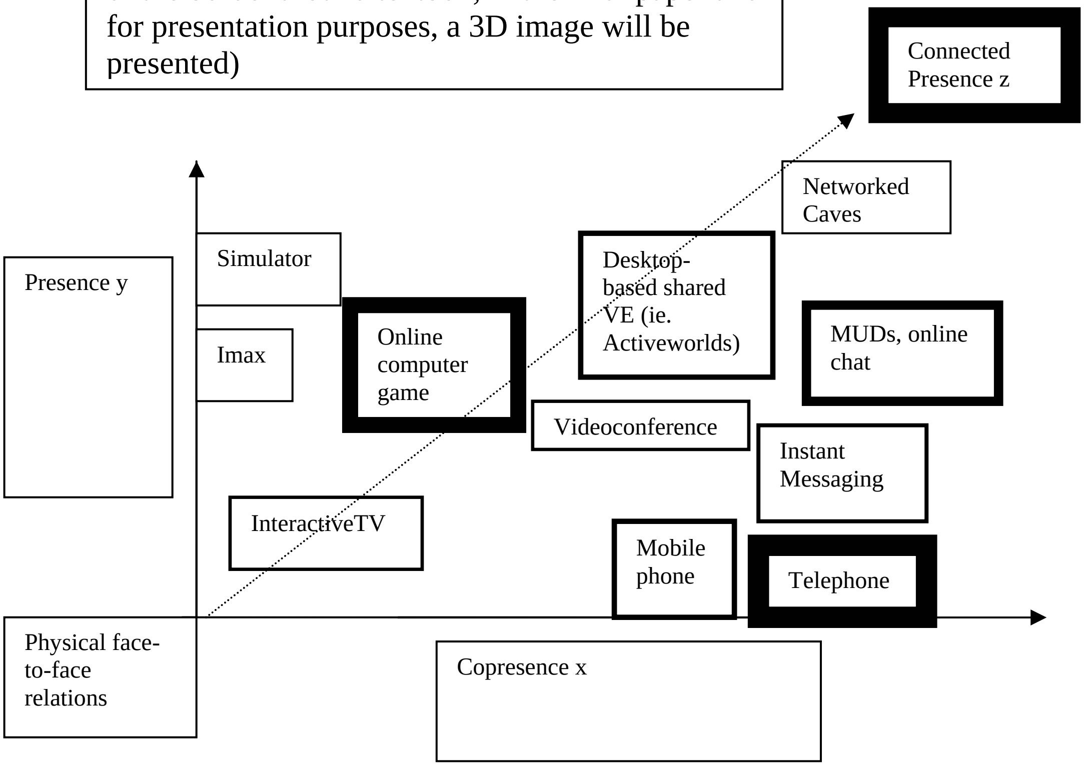


Figure 2. Presence, Copresence and Connected Presence in Different Media for Being There Together (the z axis is represented by the strength of the border around textbox, in the final paper and for presentation purposes, a 3D image will be presented)

Sketches

The following sketches are accompanied by posters that are displayed during the conference.

Integration of a brain-computer interface into Virtual Environments

Guger C. [1], Edlinger G.[1], Leeb R.[2], Antley A.[3], Friedman D.[3], Slater M.[3]

[1]g.tec - Guger Technologies OEG, Herbersteinstrasse 60, A-8020 Graz, Austria, www.gtec.at
[2] Laboratory of Brain-Computer Interfaces, Graz University of Technology, Austria.
[3] Department of Computer Science, University College London, United Kingdom
guger@gtec.at

Abstract

Electroencephalogram (EEG)-based brain-computer interface (BCI) systems convert brain activity into control signals and have been initially developed for people with severe disabilities. In the last few years BCI systems were also used in Virtual Environments (VE) for the control of experiments. Specially for the application in VE the BCI system has to satisfy specific demands. The key advantage of a Pocket PC based BCI approach is its small dimension and battery supply. Hence a mobile BCI system can be worn by a human subject during experiments in VE.

1. Introduction

Electroencephalogram (EEG)-based brain-computer interface (BCI) systems have been developed for people with severe disabilities in order to improve their quality of life. A BCI measures brain activity and transforms specific thoughts into control signals. Applications of BCI systems comprise the restoration of movements, communication and environmental control [1]. However, recently BCI systems have been also used in other research areas such as in the field of virtual reality (VR) [2, 3].

Parameters generally used to quantify the performance of BCI systems are the accuracy and speed. Furthermore, a BCI approach should ensure that the users learn to control the system within a few training sessions. The level of control should be stable after an initial learning phase and should improve over time [1, 4, 5].

Different strategies are used for the control of a BCI. The user can e.g. perform a real hand or foot movement or can just imagine the movement. Depending on the strategy it is important to measure the EEG activity exactly over the corresponding brain regions.

Thereafter, feature extraction and classification of EEG data is performed resulting in the control signal. After some training sessions the BCI accuracy enhances to a certain degree, meaning the BCI system and the subject have adapted to each other for a better general system performance [1, 4].

However, for the portable use in Virtual Environment (VE) the BCI system must be as small as possible and easy to use.

2. Mobile system

For portable applications like in VE an embedded solution including the processor and DAQ board without mechanical disks and extra display is required. Size, robustness and usability are major considerations. The hardware must be fully portable and supplied by a battery [5].

Fig. 2. Components of a Pocket PC based BCI system.

For the embedded BCI a standard Pocket PC is used as a portable host. The Pocket PC is connected via a serial cable to an embedded target computer system g.MOBIlab (see Fig. 2). The embedded system consists of a µC operating at 12 MHz to optimize the power consumption. A 16 bit analog to digital converter samples 8 analog channels. Each channel is sampled at 256 Hz. The amplifier module is equipped with 4 EEG type channels, 2 ECG type channels and 2 analog inputs for external sensors. Two digital inputs and 2 digital outputs allow controlling different external devices. Two batteries of type “AA” power the embedded system.

The Pocket PC operating system is Windows Mobile and the BCI system is programmed in eMbedded Visual C++. The integrated Wireless LAN (WLAN) module of the Pocket PC can be used for wireless data transmission. Data are stored on the internal 64 MByte storage or streamed to a Compact Flash card for later analysis. An application programming interface allows accessing the hardware components and data buffers. Hence BCI applications can be adapted to optimally meet user specific needs or novel applications can be developed.

The following paragraphs give one example for a typical BCI experiment based on oscillatory brain activity measured over electrode positions C3 and C4 [6].

3. Training phase

For BCI training two different experimental paradigms are implemented. In order to acquire EEG data in the training phase the first experiment is performed without feedback. Therefore, an arrow pointing to the left or right side of the computer monitor is shown (Fig. 2). Depending on the direction of the arrow the subject has to imagine a specific kind of movement. If the arrow is pointing to the left hand side the subject should imagine a left hand movement, if the arrow is pointing to the right side the subject should imagine a right hand movement.

EEG data for a total of 160 trials (80 right and 80 left hand movement imageries) are acquired. Specific EEG parameters are then extracted from the data and the trials are classified into two classes yielding a subject specific classifier.

4. Application phase

After computing the classifier the application phase can be started. The classifier weights the extracted features calculated from the EEG data in such a way that the thoughts are converted in real-time into bar movements (Fig. 3).

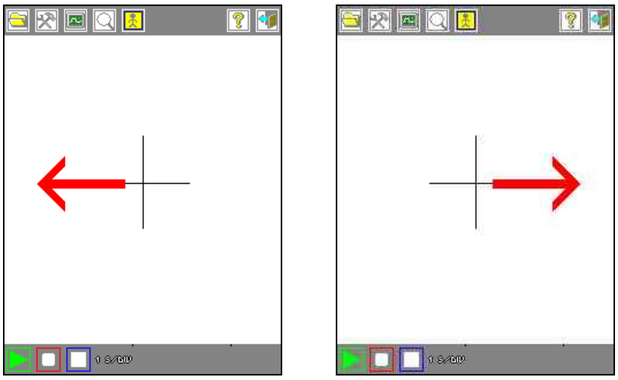


Fig. 2. Training phase displayed on the Pocket PC: Red arrows indicate that the subject should image a left hand (left panel) or right hand (right panel) movement.

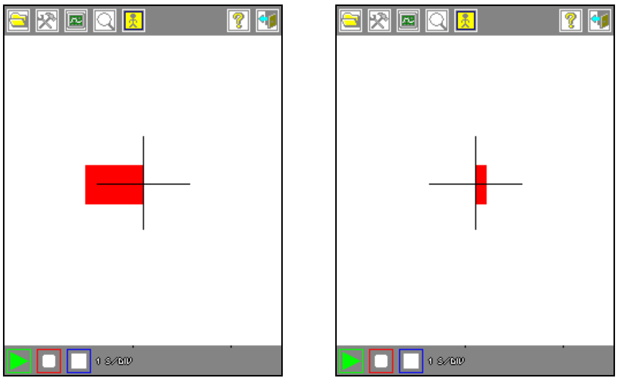


Fig. 3. Application phase displayed on the Pocket PC: The direction of the bar indicates the classification in either the left hand movement class (left panel) or the right hand movement class (right panel).

A classification result of a right hand movement extends the bar to the right side. A classification result of a left hand movement class extends the bar to the left side. This cursor movement was translated into a navigation signal in a CAVE environment. In this way the subject was able to move forward or backward depending on the imagination [2].

5. Discussion

The embedded BCI system with its compact dimension allows the usage of the BCI inside VE and (as Pocket PC CPUs are getting more and more powerful) also for implementing sophisticated applications. The system can be easily worn by the human subject and is fully battery powered. Therefore the subject is more flexible and can move through the virtual world. A big advantage is that the Pocket PC based BCI operates immediately after switching it on without booting of the operating system.

The combination of BCI systems with Virtual Reality allows people to accomplish tasks within a virtual environment simply by having the appropriate 'thoughts'. The purpose of this is to explore completely new paradigms for operating with computers, to give people an experience that is unlike anything possible in real life - thus exploiting the power of virtual reality to deliver entirely new experiences, and finally has obvious practical applications for people with disabilities.

Acknowledgment

This project is funded by the European Union project PRESENCIA, IST-2001-37927 and by the Austrian FFF.

References

[1] G. Pfurtscheller, C. Guger, G. Müller, G. Krausz, and C. Neuper,. Brain oscillations control hand orthosis in a tetraplegic. *Neuroscience Letters*, vol. 292, 211-214, 2000.

[2] R. Leeb R. Scherer, C. Keinrath, C. Guger, G. Pfurtscheller. Exploring Virtual Environments with an EEG-based BCI through motor imagery. Biomed. Technik, vol. 52, 86-91, 2005.

[3] J.D. Bayliss. Use of evoked potential P3 component for control in virtual apartment. IEEE Trans. Neural. Sys. Rehab. Eng., vol. 11(2): 113-116, 2003.

[4] C. Guger. Real-time data processing under Windows for an EEG-based brain-computer interface. *Dissertation*, University of Technology Graz, 1999.

[5] T.M. Vaughan, J.R. Wolpaw, and E. Donchin. EEG-based communication: prospects and problems. *IEEE Trans. Rehab. Engng*, vol. 4, 25-43, 1996.

[6] C. Guger, et al. How many people are able to operate an EEG-based brain-computer interface (BCI) ?. *IEEE Trans. Rehab. Engng*,, vol. 11, 145-147, 2003.

“Being there” in New Media Art: Evaluating the feeling of Presence during Mediated Art Experiences

Jeanne Figueira Grossetti, Cheryl Campanella Bracken
Cleveland State University, Cleveland State University
{j.grossetti@csuohio.edu, c.bracken@csuohio.edu}

Abstract

This paper examines the possibility for presence to inform the development and study of new media art. The paper links the concepts of art spaces, new media art, and presence. The authors argue that presence has already been incorporated into new media art and that such a union offers social scientists as well as artists a unique opportunity to study the interaction of these areas. The paper concludes by posing several research questions and has the potential to be a rich area of research.

1. Introduction

Art of the late twentieth and early twenty-first centuries increasingly engages with questions of consciousness, presence, agency, interactivity, and time {1-3]. A hallmark of this art is its self-consciousness and intentional engagement, even play, with media theory and concept. Our goal in writing this paper is to articulate the relationship between new media art and the concept of presence; to explore the potential of presence research methods to quantify viewers’ experience of presence during a mediated art experience (MAE), and to suggest strategies for future presence research of MAEs in and out of art spaces.

2. Aesthetic Encounters and Art Spaces

The significance of the aesthetic encounters cannot be overstated. During this experience, artist and audience engage, to varying degrees, with the raw material of human consciousness: dream images and echos of primordial pain and pleasure [4]. This material is mediated through music, written or spoken word, painting, sculptural installation, interactive computer-generated projection, and other materials and means. According to Nietzsche, it is during the encounter with the symbolic matrix of dreams, and that which lies behind dreams, that a person transcends illusion and encounters his or her “very self” [4, p. 11].

Set apart for aesthetic encounter, the art space, like the sacred space, intentionally extends consciousness, creating an environment in which it is collectively understood that participants may engage meaningfully with, and through, works of art. The iconic and transcendent functions of art are therefore more effective when they happen in art spaces.

2.1. Context: New Media Art

New media art and MAEs have been an increasing feature of the cultural landscape since the mid-twentieth century. The Fluxus movement, beginning in the early 1960s, included an influential group of artists, musicians, writers and performers from the U.S., Europe and Japan. Among them was the Korean-born artist Nam June Paik. Through video installations, electronic collages, video sculptures, and other work, Paik has directly engaged with global communications theory [5]. Exemplified by his Video Buddhas (which combine sculpture, monitor, and camera), Paik’s work asks such questions as: who is looking?

Since the beginnings of Fluxus, new media artists around the world have looked into the mediated environment with much the same curiosity which social scientists bring to it: a desire to penetrate into the workings of the human mind.

Art by its nature is interactive: an interface among artist, viewer, and artifact. This is true of a drawing, painting, or sculpture, whether representational or non-representational, as well as of a photograph, film, or VR experience [6]. The interaction can be introspective or relational. However, new media art allows for new and continuous interactions allowing both the artist and the “viewer” to have a dialogue, critique another work, collaborate on projects, etc…

Clearly, then, the relationship between new media art and communications theory is not a new one. While artists respond to the questions posed by communications research through their artistic creations, communications researchers probe the human engagement with works of during MAEs to test and further develop the communications theory of presence.

3. Presence and NMA

Presence has been conceptualized in a comprehensive overview as the “illusion of non-mediation” [7]. Lombard and Ditton also provided a variety of useful guidelines for testing sensations in virtual environments, particularly for the realms of the visual (image size, distance and quality; motion and color; dimensionality; and imaging techniques) and auditory (sound quality, frequency range, signal to noise ratio and dimensionality.

Presence researchers are interested in new media arts because of their potential to illuminate the human experience of presence in mediated environments. Lombard and Ditton [7] provide a useful model for researching presence in MAEs; elaborate models of presence that take into consideration physiological, psychological, cultural and social aspects of human consciousness [8,9,10]. Each has ramifications for research into (and creation of) new media art, MAEs and VEs. The Illusion of Being was created to test perceived presence and duration [11]; analyzed works by contemporary artists in art spaces to study presence and interactivity. With the exception of Leggett, each of these articles has focused on VR, immersive experiences, and/or interactivity, although these are only the most complex forms of mediated experience. The models have much to offer to researchers working with other forms of MAE, including video projection, video installation, computer-enhanced installation, and other lower-immersion mediated environments.

We propose the following types of questions to presence researchers interested in new media art:

- In the early twenty-first century, what can we know about the experience of viewing new media art in (actual) art spaces?
- Can presence research provide insight into the nature of the art experience and how might this insight be used to augment the encounter between visitors to art spaces and works of new media art, particularly lower-immersion MAEs?
- Are there differences in the types of presence experiences in identified art space (either physical or mediated) compared to generic public spaces?
- How does the feeling of presence vary with method of presentation: projection onto a screen versus display on a monitor? Does the size of the image/art matter to visitors to art spaces?
- Does the artistic content impact presence and/or the emotional states of the viewers?
- How does a gallery visitor engage with a projected work that juxtaposes the real and the virtual; With a room that becomes a stream-of-consciousness visual poem of color and motion via computer manipulation; With a computer-generated environment that changes in response to the visitor's movement or breath?
- Does the feeling of presence change in relation to solitary versus communal participation in MAEs?
- What does it mean for a visitor to an art space to see sounds, to hear colors, to feel emotions through the skin (e.g., haptics)?
- As new media artists increasingly move along these paths, what will these altered experiences mean to the consciousness of their viewers/participants?

We acknowledge that some these types of questions have been investigated by [15]. There are others who are interested in exploring the phenomenon of presence with in new media art. However, these questions (and others) offer the possibility for a rich research agenda. There are many directions in which this area of research could pursue. We encourage others to begin to answer these questions and to develop there own.

Conclusion

The potential for presence research to shed light on new media art experiences is clearly worth further study. Art in the twenty-first century is often concerned with matters of human consciousness. As a mediated experience, art is of direct relevance to social science research that seeks to penetrate into the workings of the human mind in mediated experiences – presence research.

While presence researchers are asking questions about artful mediated experiences, their research most often takes place in non-art spaces. Leggett has suggested that presence researchers look at new media art created to be experienced as art in art spaces. The authors hope that presence researchers will become more aware of the interesting and fertile area of research new media presents.

References

[1] Greene, R. (2004). *Internet Art.* New York: Thames & Hudson world of art.

[2] Paul, C. (2003). *Digital Art.* New York: Thames & Hudson world of art.

[3] Rush, M. (1999). *New Media in Late 20th-Century Art.* New York: Thames & Hudson world of art.

[4] Nietzsche, F. (1998) *Selections from "The Birth of Tragedy," Dayton, E., ed. Art and Interpretation: An Anthology of Readings in Aesthetics and the Philosophy of Art.* Orchard Park, NY: Broadview Press.

[5] Paik, N. J. (2005). *Electronic Arts Interface .* Website available at http://www.eai.org/eai/artist.jsp?artistID=481. Retrieved 2 May 2005.

[6] Dayton, E. (1998*). Art and Interpretation: An Anthology of Readings in Aesthetics and the Philosophy of Art.* Orchard Park, NY: Broadview Press.

[7] Lombard, M., & Ditton, T.B. (1997). "At the heart of it all": The concept of presence. *Journal of Computer-Mediated Communication, 13*(3). Available at www.ascusc.org/jcmc/vol3/issue2/lombard.html

[8] Waterworth, J.A. (1999). *Creativity and Sensation: The Case for Synaesthetic Media. Leonardo*, 30(4), pp. 327-330.

[9] Waterworth, E. and Waterworth, J.A. (2001). Focus, Locus, and Sensus: The Three Dimensions of Virtual Experience. *CyberPsychology & Behavior, 4*(2), 203-213.

[10] Riva, G., Waerworth, J.A., and Waterworth, E. (2004). The Layers of Presence: A Bio-cultural Approach to Understanding Presence in Natural and Mediated Environments. *CyberPsychology & Behavior, 7*(4), 402-416.

[11] Waterworth, J.A. and Waterworth, E. (2003a). Being and Time: Judged Presence and Duration as a Function of Media Form. *Presence: Teleoperators and Virtual Environments, 12*(5)495-511.

[15] Waterworth, E., & Waterworth, J.A. (2003b). The illusion of being creative. In G. Riva, F. Davide, & W. IJsselsteijn (Eds), *Being there: Concepts, effects, and measurement of presence in synthetic environments.*

Using Visual Augmentations to Influence Spatial Perception in Virtual Representations of Real Scenes

Claus B. Madsen, Lisbeth W. Lorentzen
Laboratory of Computer Vision and Media Technology, Aalborg University, Denmark
{cbm@cvmt.aau.dk, lwl@cs.aue.aau.dk}

1. Introduction

This paper presents an approach to enhancing the perception of depth in Virtual Environments based on Image-Based Rendering. We propose that by augmenting scenes with virtual objects and structures we can stimulate the user's desire to perform visual exploration and thus heighten the user's sense of presence.

In traditional 3D model-based Virtual Environments (VEs) users are free to navigate within the bounds of the model. This allows the user to visually explore the scene to get a sense of its spatial layout and composition, exactly as in the real, physical world. The main problem with such model-based VEs is that it is difficult, bordering on impossible, to photo-realistically recreate complicated, real-world locations due to: 1) the complexity of constructing a 3D model of the scene geometry, and 2) the computational complexity of rendering realistic illumination phenomena in real-time.

Image-Based Rendering (IBR) attacks both these problems. IBR synthesizes the user's current view of a scene from a set of pre-recorded images of a real scene. By using IBR users can move around and visually explore a visual recreation of a real scene, and since the visualization is based solely on images the scene can be arbitrarily complex (i.e., there is no 3D modeling involved), and all illumination phenomena are naturally recreated correctly. IBR is by definition photo-realism. But IBR suffers from a catch: with current computer technology it can only provide limited movability in scenes. In our present system the user can only move freely inside a circle with a radius of about 2 feet. Such limited movability makes visual exploration of scenes a little trivial and the amount of motion parallax that can arise from such small movement is limited.

In static scenes motion parallax is loosely speaking the difference in how points in the scene move across the retina as the observer moves [1] [2] Motion parallax is an extra-ordinarily important cue for perceiving the 3D structure of a scene, more important than stereoscopic vision for distances of more than a few feet, and rivaled only by high level information such as a priori knowledge of the natural sizes of recognizable objects.

This paper explores the use of visual augmentations, i.e., the addition of virtual objects to the scene in order to create stronger motion parallax for small ego movements. We propose that scenes can be augmented in two different ways resulting in different explorative behaviors and different perception of the scene.

2. Background

This study has its background in a research project which uses IBR to enable people to visually explore real world places without actually being there. Furthermore our system allows us to augment virtual objects into scenes.

Image-Based Rendering (IBR) is an alternative paradigm to traditional 3D model-based computer graphics. In IBR views of a scene from arbitrary viewpoints are synthesized from data in a large set of images acquired at some location. In our IBR system approximately 400 images are acquired by moving the acquisition camera along the circumference of a circle, with the camera lens point outward of the circle. Currently our setup allows us to acquire images in circle with a radius of 60 centimeters.

From the acquired set of images we can synthesize views of the scene from any point inside the circle. We call this area Region Of Exploration (REX), because it is within this area the user can explore the scene in all directions (full view sphere). The position and viewing direction of the user is tracked with a commercial tracking system, and correct views are presented to the user in stereo at more than 20 frames per second in either a Head Mounted Display (HMD) in a six sided CAVE. In case of the HMD the system runs on a single standard PC, whereas for the CAVE version a PC per projection surface is used.

As computers can hold more and more memory larger and larger REXs are feasible at no extra computational cost, but IBR will always entail some REX concept, i.e., some finite area within which the user can move freely, but outside which the scene cannot be rendered. IBR's biggest advantage is that no modeling whatsoever is involved. We just set up the acquisition system, scan the scene, and afterwards the images can be used directly for photo-realistic visual exploration of the scanned location.

Since IBR is based directly on recorded images the scenes that can be visualized with this technique have to be static. Moreover the visualization approach cannot handle if there are real world objects inside the acquisition region, and thus inside the REX.

In order to get dynamics, interactivity and/or objects inside REX we need to insert virtual objects (similarly to augmented reality). Augmented objects are visualized using a traditional model-based rendering paradigm, that is, virtual objects are modeled and textured in a commercial modeling package such as 3D Studio Max, saved in a VRML file and loaded into our system at start up time. To get scenario consistent illumination of the virtual objects we model the real scene illumination conditions [3] .

3. Experiences with the system

The system has been tested on hundreds of test persons experiencing one or more of about a dozen scanned real world locations ranging from wide open outdoor scenes over an indoor sub-tropical botanical garden to small office spaces. Generally our experience is that people are impressed with the IBR based approach and its ability to realistically recreate complex real world places in 3D stereo. Yet, our main impression is that people tend not to fully exploit the potential for visually exploring the displayed locations. People do look around in all directions but they do typically not perform much sideways head movement or shift their position perpendicularly to the viewing direction. If no perpendicular movement is performed it is impossible to appreciate that the displayed environment is in fact a full 3D environment with objects and structures at different depths, because then no motion parallax is generated and the only cue to depth differences is the stereo disparities.

The question then is: what space characteristics motivate visual exploration? We conjecture that spaces which generate strong motion parallax are interesting to explore. For the types of motion we are talking about this means scenes that have a lot of vertical structure at different depths, ranging from the center of the REX to far away.

4. Motivating exploration with augmentation

The main reason for inserting augmentations is to animate the user to visually explore the scenario and to engage in movements which generate visual parallax.

For the purpose of this study we have decided to operate with two types of augmentations, and the main hypothesis of this study is that these two types have fundamentally different effects on users' spatial perception of a scene. An example of each type is shown in Figure 1.

Outside-in augmentations. This type of augmentation basically occupies the center of the observer's area of move ability (REX). With these augmentations the observer is pushed out to the border of the REX and is stimulated to circulate the augmentation and thus to visually explore the scene relative to the augmentation. The main visual focus of the observer will tend to be on the augmentation.

Inside-out augmentations. This type of augmentation essentially surrounds the observer. The observer is animated to perform movements perpendicularly to the viewing direction in order to see past the augmentation, and the main visual focus may be evenly distributed between the augmentation and the rest of the scene.

Conclusions

We have proposed the use of visual augmentation for Virtual Environments based on Image-Based Rendering in order to stimulate observers to engage in more active visual exploration of the scenes. Specifically we have argued for categorizing augmentations into two types: outside-in and inside-out, which we believe will invoke different exploration behaviors.

Finally we believe that these two types of augmentation can be applied in similar manner to normal model-based Virtual Environments, which when displayed in either HMD or CAVE also restrict users' area of body movement due to the range of tracking equipment and, in the case of CAVE, the size of the CAVE. That is, all VEs where navigation is performed with normal body movements rather than with interaction devices can exploit the use of special objects placed in the scene at or around the center of the exploration area to shape the manner in which people explore the scene.

Figure 1: Top: solar clock augmented into an outdoor scene (outside-in augmentation). Bottom: pavilion inserted into a botanical garden scene (inside-out augmentation).

Acknowledgements

This research is funded by the BENOGO project under the European Commision (IST-2001-39184), and by the CoSPE project (26-04-0171) under the Danish Research Agency.

References

[1] V. Bruce, P. R. Green. Visual Perception: Physiology, Psychology and Ecology. 2nd edition. Lawrence Erlbaum Associates, publishers. 1990.

[2] E. Trucco, A. Verri. Introductory Techniques for 3-D Computer Vision. Prentice Hall. 1998.

[3] C.B. Madsen, M.K.D. Sørensen, M. Vittrup. Estimating Positions and Radiances of a Small Number of Light Sources for Real-Time Image-Based Lighting. *Proceedings of EUROGRAPHICS 2003,* 37 – 44, September 2003.

Self-induced Footsteps Sounds in Virtual Reality: Latency, Recognition, Quality and Presence

Rolf Nordahl
Medialogy, Aalborg University Copenhagen
2750 Ballerup, Denmark
rn@media.aau.dk

Abstract

In this paper we describe the results of experiments whose goal is to investigate the effect of enhancing a virtual reality experience with the sound of synthetic footsteps. Results show that the sense of presence is enhanced when the sound of one's own motion is added. Furthermore, the experiments show that the threshold for detection of latency between motion and sound is raised when visual stimuli is introduced.

1. Introduction

Through the recent years some presence studies have focused on whether the addition of auditory cues in Virtual Environments (VE) and Virtual Reality (VR) could lead to measurable enhancements in participators feeling of presence. Most of the results of previous research have been focusing of sound delivery methods [1,2,4], sound quantity [3,4] and quality of visual versus auditory information [3,5]. To our knowledge, the effect of self induced sounds to enhance sense of presence has not been investigated yet. In this paper, we are interested in investigating if enhancing the VR experience with the sound of the subjects' own footsteps enhances sense of presence.

We designed a real-time footstep synthesizer, controlled by the subjects by using a set of sandals embedded with pressure sensitive sandals. By navigating in the environment, the user controls the synthetic sounds.

2. Designing synthetic footsteps

The footstep sound synthesizer works in real-time under the Max/MSP platform[1]. Footsteps recorded on seven different surfaces were obtained from the Hollywood Edge Sound Effects library[2]. The surfaces used were metal, wood, grass, bricks, tiles, gravel and snow.

The sounds were analyzed using the Fast Fourier Transform (FFT), and the main resonances were extracted from the spectrum. .Such resonances were used to build a modal synthesizer [6, 7]. To trigger the synthetic footsteps, the users were asked to wear a pair of sandals embedded with pressure sensitive sensors placed one in each heel.

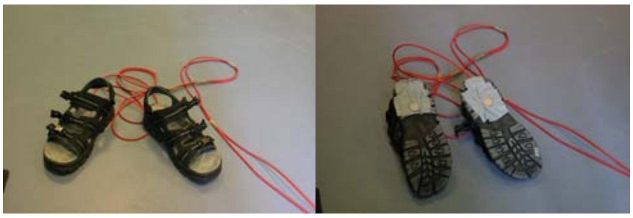


Figure 1: The shoe controller developed for the experiments.

When the subject walked around, the heel of the sandal would come into contact with the floor, thereby trigging the sensors. Through the use of a microprocessor, the corresponding pressure value was converted into an input parameter which was read by the real-time sound synthesizer Max/MSP. The sensors were connected to the microprocessor through wires, as shown in Figure 1, and the microprocessor was connected to a laptop PC.

3. Experimental setup

The main goal of the experiments was to test if the sound of one's own footsteps enhanced the sense of presence. Two different experiments were performed.

In the first experiment only the auditory feedback was provided. Subjects were asked to wear the sandals enhanced with sensors and a pair of headphones without being told the purpose of either. In this experiment latency-perception and auditory recognition of the floor surface were tested.

In the second experiment subjects were exposed to a VE provided through a HMD. The subjects were divided into 2 groups, one only exposed to visual feedback while the other was exposed to bimodal (audio-visual) feedback. The visual material was a reconstruction of the Prague technical museum developed as a part of the BENOGO-project. 16 subjects participated in both experiments.

4. Sound recognition, quality and evaluation

In order to test the shoe controller on different synthesized surfaces, to understand how their quality and appropriateness were perceived, a testing scenario was designed.

The seven different synthesized surfaces were played in random order, and subjects were asked to recognize the surface and judge the quality of the sound in a scale from 1 to 5 (unimodal case) or judge the appropriateness of the sound in the displayed scenario (bimodal case).

---

[1] www.cycling74.com

[2] www.hollywoodedge.com

‘	**Unimodal**		**Bimodal**	
	Recognition	**Quality**	**Recognition**	**Appropriate**
Metal	93,3%	3,9	70%	2,05
Wood	37,5%	3,7	60%	2,95
Grass	18,8%	3,25	25%	1,63
Bricks	37,5%	3,81	70%	4
Tiles	6,25%	3,78	60%	3,8
Gravel	93,75%	3,78	100%	1,6
Snow	37,5%	2,53	35%	1,47

Table 1. Result of the sound identification test, in the unimodal and bimodal condition

Table 1 shows the results of the sound identification test. Notice how subjects could easily recognize the metallic surface, more likely in the unimodal (93,3 %) than in the bimodal case (70%). Notice also the high recognition factor of gravel (93,75% in unimodal and 100% in bimodal). Hard surfaces such as wood, tiles and bricks were harder to identify, and often confused among each others. Consistent with the fact that the floor of the technical museum was made of bricks, this sound was considered most appropriate. Notice also how the recognition of such surface significantly increases when visual feedback is provided (37,5%) versus (70%).

5. Latency perception

The goal of this experiment was to test the level of acceptance of latency in VR. Subjects were asked to walk around and inform the facilitator when they perceived a delay between the step and the corresponding sound. While the subjects were walking, the facilitator was increasing the delay between the steps and the corresponding sounds, by a factor of 5 ms. The test was performed both without and with visual feedback.

	Uni-modal		**Bi-modal**			
	M	SD	M	SD	T	p
Latency perceived	41.7ms	5.8ms	60.9ms	20.7ms	-3.5	0.002

Table 2: Result of the latency test in the unimodal and bimodal condition.

The results of the latency perception test are shown in Table 2. Notice how the perceived latency is significantly higher in the audio-visual condition (M= 60.9 ms, SD=20.7 ms) rather than in the auditory only condition (M=41.7 ms, SD=5.8 ms). This result is most likely due to the fact that the attention of the subjects in the bimodal condition was mostly focused on the visual rather that on the auditory feedback, which was clearly not the case in the unimodal condition.

6. Presence test

In order to measure the subjective feeling of presence in unimodal (visual) and bimodal (audio-visual) case, the Swedish Viewer User Presence questionnaire [8] was chosen. 20 participants were randomly assigned either the visual (n 9, one female), or the audio-visual condition (n 10, one female). In both conditions subjects were asked to wear the HMD, headphones and sandals. In order to facilitate the self-motion, subjects were asked to count the number of airplanes and cars they could identify in the virtual space.

	Unim		**Bim**		**T**	**P**
	M	**SD**	**M**	**SD**		
PRESENCE	4.64	0.63	5.35	0.39	-2.88	0.012
ENJOYMENT	5.44	1.01	6.01	1.22	-1.27	0.2
EXTERNAL AWARENESS	-1.1667	2.97	0.2	2.33	-1.1	0.28
SIMULATOR SICKNESS	1.35	0.77	1.31	0.29	0.1	0.9

Table 3: Results of the SVUP Presence Questionnaire

Results displayed in Table 3 show that the sense of presence is significantly higher in the bimodal (M=5.35, STD=0.39) than in the unimodal case (M=4.64, SD=0.63)(P=0.012, t=-2.88).

7. Conclusions

Results obtained show that the sense of presence is significantly enhanced when self sound is added to the VR environment. However it should be noted that no condition with other kinds of sound was tested since the original real scenario did not contain any sounds, other than a very distant noise from a fan. In future tests such conditions will be added. Furthermore, tests involving both conditions with HMD and CAVE setups will be used.

References

[1] R.L. Storms and M.J. Zyda. Interactions in perceived quality of auditory-visual displays. Presence -Teleoperators an Virtual Environments, vol. 9, n. 6, pp. 557-580, 2000.

[2] Sanders, Richard D. and Scorgie, Jr. Mark A., “The Effect of Sound Delivery Methods on A User’s Sense of Presence in A Virtual Environment”, Thesis, Naval Postgraduate School, Monterey, California, 2002

[3] P. Chueng and P. Marsden. Designing Auditory Spaces: The Role of Expectation, HCI International, 2003.

[4] S. Serafin and G. Serafin. Sound design to enhance presence in photorealistic virtual reality. Proc. ICAD 2004.

[5] J. Freeman and J. Lessiter. Here, there and everywhere: The. effect of multichannel audio on presence. Proc. ICAD,. 2001

[6] L. Savioja et al. Creating Interactive Virtual Acoustic Environments. Journal of the Acoustical Engineering Society, vol. 47, no.9, pp.675-705, 1999.

[7] Cook. Real sound synthesis for interactive applications. AK Peters, 2002

[8] P. Larsson et al. The actor-observer effect in virtual reality presentations. Cyberpsychol Behav. 2001 Apr;4(2):239-46.

Phasic heart rate response in virtual environments

Gert Pfurtscheller[1,2], Robert Leeb[1], Christoph Guger[3], Mel Slater[4]

[1] Laboratory of Brain-Computer Interfaces, Graz University of Technology, A-8010 Graz, Austria
[2] g.tec - Guger Technologies OEG, A-8020 Graz, Austria
[3] Department of Computer Science, University College London, United Kingdom

Abstract

Heart rate responses induced by motor imagery were investigated in 4 subjects in a series of experiments with a Brain-Computer Interface (BCI). The goal of the BCI experiment was either to control a bar on a PC monitor or to move forward within a virtual environment (VE). In the first case all subjects displayed a HR decrease during motor imagery in the order of 2 – 5 %. The thought-based control of VE resulted in a heart rate increase in 2 subjects and a heart rate decrease in the other 2 subjects. The heart rate acceleration in the VE is interpreted as effect of mental effort and motivation.

Keywords — ***Heart rate, Motor imagery, Virtual environment, Brain-Computer Interface***.

1. Introduction

Preparation or planning of a self-paced movement is accompanied by a deceleration of the heart rate [1, 2]. Because similar neuronal structures are involved in motor execution and motor imagery [3] it is of interest to investigate heart rate (HR) changes in the Brain-Computer Interface (BCI) experiments with motor imagery as mental strategy and different types of feedback (FB). The objective of the study is to investigate HR changes while the BCI is used to control (i) a simple bar on a PC monitor and (ii) a virtual environment (VE) with the goal to move e.g. forward in a virtual street as far as possible.

2. Materials and Methods

The study was performed on 4 healthy student volunteers aged 23 – 30 years (mean age 26.7 years). The subjects took part in a series of BCI experiments over some months with the goal of achieving control over their brain activity [4] for mental control of a virtual environment (VE) [5]. Two mental strategies were used: either imagination of right hand vs. left hand or right hand vs. foot/leg movements. In the majority of these experiments, in addition to the EEG, the electrocardiogram (ECG) was also acquired from the thorax which was sampled at 250 Hz, stored and used for off-line processing.

2.1. Experimental paradigm

Each subject took part in a number of BCI training runs in Graz. Thereafter, runs were performed in London in a multi-projection based stereo and head-tracked VE system commonly known as a "CAVE" [6] and finally, control runs were made in Graz again. In each run, the subject had to imagine movements in response to an auditory cue stimulus, given either as single beep or as double beeps. Each trial lasted about 8 seconds, during which at second 3 the cue-stimulus appeared. The subject was instructed to imagine the indicated movement over the next 4 seconds while feedback was given during that time.

The data of VE sessions in the Cave in London and the final control sessions on the PC monitor in Graz are reported in this paper. The EEG trials were used for the discrimination of the 2 mental states of motor imagery [4]. Details can be found elsewhere [5].

2.2. Electrocardiogram processing and calculation of HR changes

After the detection of the QRS-complexes in the ECG a R-R time series was extracted. From this time series the instantaneous heart rate was calculated by linear interpolation between consecutive RR-interval samples and resampling with 4 Hz. After selecting of 8-s instantaneous HR trials with 3s prior to the cue-stimulus, averaging was performed across the 40 trials of each run. The result is an event-related HR-time course together with the sample-by-sample intertrial standard deviation (SD).

3. Results

The results obtained in all 4 subjects are summarized in Table 1. All subjects displayed a HR decrease (increase of the R-R intervals) during motor imagery under control condition with FB on a PC monitor. A characteristic example from one subject is displayed in Fig. 1A. In the VE 2 subjects displayed also a HR decrease, but 2 subjects a HR increase. A detailed analysis of one BCI experiment in the CAVE with a HR increase revealed the following: A correct classification of foot motor imagery in the EEG data resulted in a forward moving in the VE, while a false classification of hand motor imagery resulted in a backward moving. In the former case the positive FB was accompanied by a weak HR change (Fig. 1B). In the latter case, when the subject was not successful to move forward and disappointed, the HR displayed an increase of about 4.5 bpm (beats-per-minute) (see Fig. 1C).

subject	*condition*	RRref [ms]		RRresp [ms]		change	p
		mean	SE	mean	SE	[%]	
S1	*control*	716	2,31	740	2,43	3,32	<0,001
	VE	708	2,77	723	2,97	2,12	<0,001
S2	*control*	976	4,76	1004	4,45	2,87	<0,001
	VE	812	2,93	846	2,62	4,19	<0,001
S3	*control*	865	3,88	901	3,51	4,12	<0,001
	VE	954	4,91	904	7,20	-5,19	<0,001
S4	*control*	796	4,49	841	4,84	5,65	<0,001
	VE	765	7,56	753	7,54	-1,63	<0,05

Table 1: Mean R-R interval and standard error (in ms) in the reference interval before the cue-presentation (RRref) and during motor imagery (RRresp) for all subjects and conditions. In addition changes (in %) between the RRref and RRresp and their significances are displayed.

4. Discussion

Motor imagery is accompanied by a slight but significant heart rate deceleration in the order of 2 – 5 %, when simple feedback is given on a computer monitor. This is not surprising, because motor imagery involves similar cortical structures to those activated during preparation of voluntary movement [7] and heart rate decelerates when a subject prepares a self-paced movement [2]. A logical consequence of increased motor cortex excitability is that it should propagate down to the brain stems spinal cord and motor neuron levels. It is also known that motor imagery activates not only neural structures in primary motor cortex [7] but is also accompanied by an increased corticospinal and spinal reflex pathway excitability [8]. In the brain stem the parasympathetic system is activated which results in a slowing of the heart rate.

A surprising result is the heart rate acceleration in 2 subjects observed in the VE experiments. Thought-based forward moving in a VE is a great challenge for a subject and therefore the motivation is higher as in standard BCI experiments with a PC. We can hypothesize, that a negative FB in the CAVE (e.g. by moving backwards instead of forwards) increases the mental effort and the motivation of the subject with the goal to reach a change of the classification result and reveal thereby a positive FB, e.g. by moving forward in the VE. As a result of this increased mental effort (in close connection with the motivation to move forward as far as possible) the sympathetic system becomes activated and the HR increases.

The study suggests that neocortical structures involved in motor imagery impinge upon brain stem cardiovascular nuclei and modify the heart rate. In general, motor imagery is associated with an HR deceleration. Subcortical structures related to motivational and other psychological processes activate the sympathetic system and reveal a HR acceleration.

Acknowledgements

The work was funded by the European PRESENCIA project (IST-2001-37927).

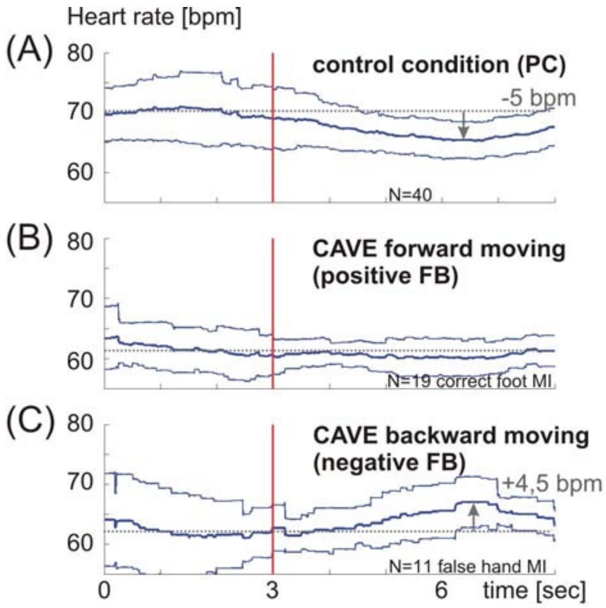


Figure 1: Heart rate (HR) changes (±SD) in bpm for a control experiment (A) without CAVE and an experiment within the CAVE (B and C). HR changes during positive feedback (FB) and forward moving (correct classification of foot motor imagery) are displayed in (B) and HR changes during negative FB and backward moving (false classification of hand motor imagery) are displayed in (C).

References

[1] J. R. Jennings, M. W. van der Molen, R. J. Somsen, et al., "On the shift from anticipatory heart rate deceleration to acceleratory recovery: revisiting the role of response factors," *Psychophysiology*, vol. 27, pp. 385-95, 1990.

[2] G. Florian, A. Stancak, and G. Pfurtscheller, "Cardiac response induced by voluntary self-paced finger movement," *Int J Psychophysiol*, vol. 28, pp. 273-83, 1998.

[3] M. Lotze, P. Montoya, M. Erb, et al., "Activation of cortical and cerebellar motor areas during executed and imagined hand movements: an fMRI study," *J Cogn Neurosci*, vol. 11, pp. 491-501, 1999.

[4] G. Pfurtscheller, C. Neuper, and N. Birbaumer, "Human Brain-Computer Interface," in *Motor cortex in voluntary movements: a distributed system for distributed functions. Series: Methods and New Frontiers in Neuroscience*, E. Vaadia and A. Riehle, Eds.: CRC Press, 2005, pp. 367-401.

[5] R. Leeb and G. Pfurtscheller, "Walking through a Virtual City by Thought," *Proc. of the 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society - EMBC 2004, San Francisco, USA*, pp. 4503-6, 2004.

[6] C. Cruz-Neira, D. J. Sandin, and T. A. DeFanti, "Surround-screen projection-based virtual reality: the design and implementation of the CAVE," in *Proceedings of the 20th annual conference on Computer graphics and interactive techniques*: ACM Press, 1993, pp. 135-142.

[7] C. A. Porro, M. P. Francescato, V. Cettolo, et al., "Primary motor and sensory cortex activation during motor performance and motor imagery: a functional magnetic resonance imaging study," *J Neurosci*, vol. 16, pp. 7688-98, 1996.

[8] S. Clark, F. Tremblay, and D. Ste-Marie, "Differential modulation of corticospinal excitability during observation, mental imagery and imitation of hand actions," *Neuropsychologia*, vol. 42, pp. 105-12, 2004.

Social Presence as Presentation of Self

Ruth Rettie
Kingston University
{http://www.kingston.ac.uk/~ku03468}

Abstract

This paper distinguishes between two different social presence concepts: social presence as projection of the other and social presence as experience. The concept is further explicated by relating it to presentation of self [1]. This is illustrated by excerpts from ongoing qualitative research.

Keywords ---**social presence, copresence, mediated communication, mobile phone.**

1. Two senses of social presence

Definitions of social presence are wide-ranging. However, analysis of numerous definitions suggests there is an underlying confusion between two closely related, but different usages of ‘social presence’. The difference is subtle. Firstly, and in line with Short et al., [2] social presence is used for the sense or perception of another, that is, the projected presence of a person. Secondly, social presence is used for the experience of being present, that is, socially present in an environment which includes another. The former is the ‘sense **of** the other’, the latter, the ‘sense of **being** with the other’. At first sight these appear to be simply two sides of the same thing, but this is incorrect. The former is intersubjective, and refers to social presence as projection or presentation; the latter is subjective, the phenomenological experience of being present socially. The first is closer to Short et al., the second is closer to copresence, and is about being with others. This explains how copresence may be conflated with social presence.

Short et al. [2, p. 65] introduced their *“hypothetical construct”* social presence in the first, projected, sense. It is *“the degree of salience of the other person”*, and is related to the transmission of cues through the medium. Although, Short et al. introduced social presence from the perspective of the receiver in the interaction, the term is also used in a more active sense, from the perspective of the sender. This first person perspective is typically found in learning theory literature [e.g. 3] changing social presence from a passive transmission of cues to an active accomplishment of the sender, a presentation of self.

The second sense of social presence is an experience, the sense of **being** with another rather than the sense **of** another. For example, Sallnas: *“Social presence refers to the feeling of being socially present with another person at a remote location”*, [4, p. 22]. Figure 1 shows the difference between the two senses of social presence.

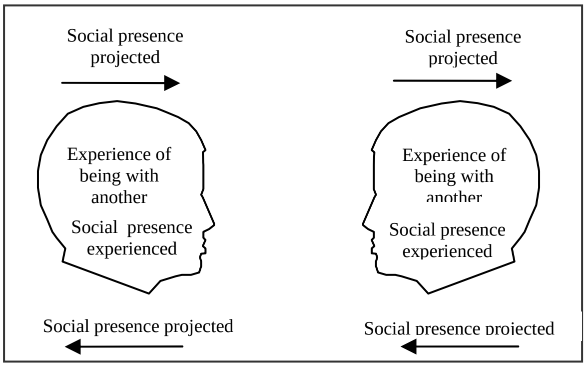


Figure 1: Two Senses of Social Presence

Garau [5] notes that the concepts of social presence and copresence are confounded; this occurs when social presence is used in the second sense of being socially present with another. One way to distinguish social presence and copresence is to use the distinction between the first and second person. Taking experienced social presence, there is a difference between my subjective experience and your subjective experience. These distinctions are relevant to social presence, but less relevant to copresence. Whereas I might, unilaterally, experience being socially present with another, copresence is a **mutual,** symmetrical relationship. On this definition, one-way media enable social presence, but not copresence; the unilateral use of web cam with instant messenger would increase social presence, but only bilateral use would create copresence. Copresence refers to the **mutual** awareness of each other by the participants in an interaction.

The remainder of the paper is concerned with social presence as defined by Short et al. [2], i.e. as a projection.

2. Social presence and presentation of self

Short’s concept of social presence is strikingly similar to Goffman’s [1] presentation of self. Introducing the concept of social presence, Short et al. [2, p. 64] relate it directly to the presentation of self, commenting that on the telephone *“negotiators are likely to be less concerned with the presentation of the self”*. On the telephone there is a reduced capacity to transmit cues such as *“facial expression, direction of looking, posture, dress and non-verbal cues”*; this reduces the social presence of the medium. This also reflects the similarity between social presence and presentation of self, for Goffman these cues are expressions given off in the presentation of self.

This paper suggests that the presentation of self helps to illuminate the concept of social presence. Unlike social presence, which assumes that there is a holistic combination of cues that is perceived as projected, albeit depleted by mediated channels, Goffman recognised that both are constructions. The social presence of an individual is not invariant, but is a performance created through expressions and directed at the maintenance of a particular perspective. Presentation of self is adapted to the situation, including roles, context and social norms. Furthermore, the self projected depends not only on the various cues, but on their classification by the recipient as intentional or unintentional. These complexities mean that the projected self perceived by the other participant may be different from the intended presentation of self.

The complexity of presentation of self challenges the simplicity of social presence, which is treated as a straightforward construct, sent and received either directly in face-to-face interaction or with some loss through mediated channels; the element of joint construction and collaboration in interaction is ignored. Goffman also claims that both the projection and the reception of cues are **deliberately** adapted for different channels. Although most research on social presence has been done with strangers, the salience of the other is likely to be less dependent on cues when people already know one another. In this case, cues may serve more as reminders and less like the 'building blocks' of a holistic perception of the other.

1. Research

This section briefly describes ongoing research which illustrates how an exploration of expressions given off can provide a more sensitive analysis than social presence. The research explores the **perceived** affect of medium on communication, focused on two media which coincide in a single device, mobile phone calls and texts. The research consists of 32 ethnographic interviews. Respondents were all regular users of mobile phones, over 21, equally split between men and women, and included different social classes and work statuses. A number of techniques were used to elicit the personal constructs used to categorise communication channels, for more details see [6].

Although, respondents did not seem to have any holistic conception corresponding to social presence, the items of the original Short et al. [2] scale were frequently used spontaneously. For instance, the warm–cold, sensitive – insensitive dimensions were sometimes used to explain choice between channels. 'Social' was also used, but it was contrasted with work rather than unsocial; personal was frequently used but rarely contrasted with impersonal. The research also suggests that these are not simple concepts suitable for linear scales. For instance, personal was used in at least three different ways. It usually meant 'intimate' but was also used to mean 'private' and 'characteristic of a person'. The contents of mobile phones are personal in the sense of being private. Phone calls are also personal in the sense that voices are **personally** distinctive, and specific to a person (like handwriting in a letter). However, respondents frequently used personal in the sense of 'intimate'. Even in this specific usage, respondents were often unable to categorise text messages as more or less personal than phone calls, because the two media are intimate in different ways. The lack of copresence meant that people could safely say more intimate things in a text message; on the other hand, phone calls are intimate because response is concurrent and spontaneous. This illustrates how linear comparison of media on these dimensions is inappropriate.

The research also explores how control over cues given off varies in the two media, and how these are **deliberately** used in the presentation of self and construction of social presence, for instance, in the exact timing of a text message.

3. Conclusions

This paper distinguishes two different concepts that are conflated under the term 'social presence', social presence projected and social presence as experience. Projected social presence is related to Goffman's [1] presentation of self. The concept of presentation of self suggests that social interaction is more complex and multi-dimensional than that assumed by social presence theory, and recognises the active involvement of an individual in the projection of self.

The paper also briefly describes ongoing research on mobile phone calls and text messages. This research suggests that although respondents do not have any holistic conception corresponding to social presence, the dimensions of 'social presence' **are** relevant to the differentiation of media. However, they are complex constructs and consequently linear scales are inappropriate.

Qualitative research can be used to explore how expressions given and given off are used in the projection and interpretation of self. This approach improves understanding of the interactional effects of mediation and facilitates comparison of communication channels.

References

[1] E. Goffman. The Presentation of Self in Everyday Life. Penguin Books Ltd. 1959.

[2] J. Short, W. Williams, B. Christie. The Social Psychology of Telecommunications. Wiley & Sons. 1976.

[3] L. Rourke, T. Anderson, D.R. Garrison, W. Archer. Assessing social presence in asynchronous text-based computer conferencing. *Journal of Distance Education/Revue de l'enseignement à distance*. 14. 51-70. 1999.

[4] E.L. Sallnas. The effect of modality on social presence, presence and performance in collaborative virtual environments. KTH Numerical Analysis & Computer Science. URL: http://www.nada.kth.se/utbildning/forsk.utb/avhandlingar/dokt/040324sallnas.pdf

[5] M. Garau. The impact of avatar fidelity on social interaction in virtual environments. University College London. URL: http://www.cs.ucl.ac.uk/staff/m.garau/papers/garau_phd.pdf

[6] R. Rettie. Social Presence as Presentation of Self. Kingston University. URL: www.kingston.ac.uk/~ku03468

Interaction with haptic feedback and co-location in virtual reality

David Swapp, Vijay Pawar and Céline Loscos
University College London
d.swapp@cs.ucl.ac.uk, vijaympawar@hotmail.com, c.loscos@cs.ucl.ac.uk

Abstract

This paper outlines a study into the effects of co-location[1] *of haptic and visual sensory modes in VR simulations. The study hypothesis is that co-location of these sensory modes will lead to improved task performance and enhanced sense of presence within a VR environment. Experiments were conducted to evaluate the effect on user performance of co-located haptic feedback. Results show that co-location is an important factor, and when coupled with haptic feedback the performance of the user is greatly improved.*

Keywords--- **Haptic interaction, virtual environment systems, visual-haptic co-location**

1. Introduction

Presence is likely to be enhanced by multi-modal input: in a VR environment, the addition of sensory modes should consolidate our sense of presence, although conflicting sensory cues are liable to degrade the sense of presence. At the moment, research in VR is dominated by simulation for the visual and audio sensory modes. In many application areas it is likely that touch can also be a compelling factor in presence [1] [2] , and other studies show that the addition of haptics can lead to improved task performance[3] [4] .

Precise co-location of haptics is technically hard to achieve. A commonly-implemented compromise is the use of visual markers to represent the haptic contact points. Because the markers are visually rendered by the same graphics system as the virtual environment, spatial correspondence is guaranteed. In the current study, such a setup is referred to as non-colocated haptics.

1.1 Implementation issues for co-location

- **Occlusion:** For screen-projection systems (as opposed to HMDs), occlusion problems arise when we reach behind a displayed graphical object: instead of our hand being occluded by the object, the reverse is the case.
- **Accommodation:** Accommodation (focus) of the eyes on a virtual object is determined by the distance from the eyes to the projection surface. However, if we are trying to view a real object (e.g. the haptic contact point) that is co-located in space with a virtual object, this gives rise to a perceptual dissonance –we can feel the object at our fingertip via haptic feedback, but we cannot visually focus on both virtual object and fingertip simultaneously.
- **Calibration:** The co-ordinate systems for both visual and haptic rendering must be aligned. Discrepancies between haptic positioning (which typically can be calibrated to a very high degree of accuracy) and head tracking will lead to a decoupling of the visual and haptic renderings. Additionally, CRT nonlinearities can distort stereo disparities and disrupt co-location.

2. Design of experiments

In order to evaluate the effect of co-location on user performance, we designed 3 experiments to test users' interaction accuracy, ease of manipulation, and agility. The experiments were run on a PC with NVidia Quadro FX1100 graphics, displayed on a CRT monitor. The user wore shutter glasses for stereo viewing. Haptic interaction was provided with a Phantom Desktop from Sensable technology[5] The Phantom was positioned to allow co-location and the full workspace of the device. The interaction workspace was between the screen and the user, the support being on the right hand side of the user.

For each task there are 2 independent variables: co-location and haptic feedback. For co-location, the Phantom is carefully positioned such that the point of interaction on the Phantom coincides visually with the point of contact in the 3D scene. For non-co-location, visual markers indicate this point of contact. When haptic feedback is turned off, the Phantom is used as a 3D joystick. Thus there are 4 classes of interaction:

- co-located haptics
- non-colocated haptics
- co-location with no haptic feedback
- non-colocation, no haptic feedback

2.1 Task design

The first task tests spatial accuracy. The user needs to touch, one by one in a given sequence, a set of objects distributed in 3D space. A screenshot is shown in Figure 1.

Figure 1 Spatial accuracy test

---

[1] The term 'co-location' is used throughout to refer to the co-location of haptic and visual sensory modes, except where otherwise specified.

The second task tests spatial manipulation. It involves manipulating a ball through an environment consisting of a sequence of objects, akin to moving it through a maze.

The third task tests spatial response. Gravity is simulated and the user must juggle objects in the environment. The task stops when an object drops.

For all tasks, there are 3 levels of difficulty, with increasing numbers of objects, more complex spatial arrangement, and decreasing object size. For each trial, the time taken to complete the task is measured**.**

2.2 Experiment procedure

A within-groups design was employed on a set of 6 users. Each user was given a description of the tasks, after which the system was calibrated for stereo adaptation and co-location. Users were asked to keep their head as still as possible to maintain correct stereo and co-location. A training period of a few minutes followed. The tasks were then presented in the following order: spatial accuracy, spatial manipulation, then spatial response. Each task was performed using the 4 interaction classes in order: co-located haptics; non-colocated haptics; co-located with no haptic feedback; non-co-located with no haptic feedback.

3. Results

All users completed the set of tasks and times were recorded. The results are shown in Figures 2, 3 and 4. For Figures 2 and 3 shorter time indicates better performance. For Figure 4, longer time indicates better performance.

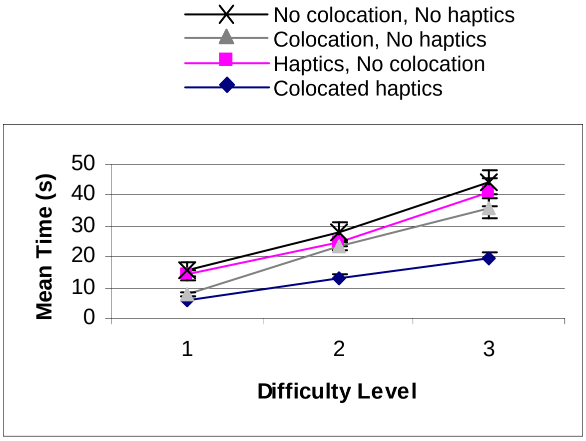


Figure 2 Results for spatial accuracy.

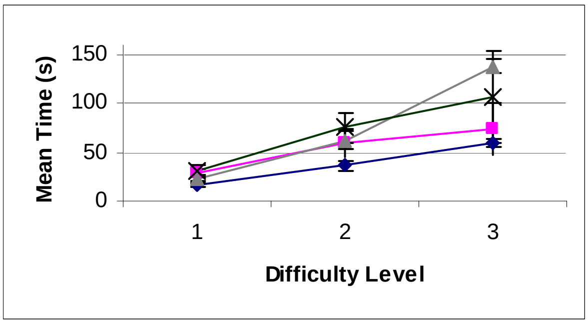


Figure 3 Results for spatial manipulation.

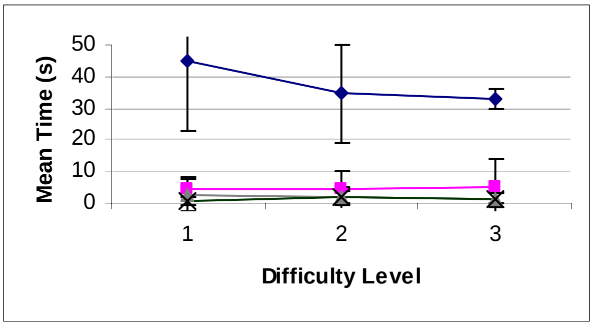


Figure 4 Results for spatial response.

The most salient results are summarised below:

- Interaction with co-located haptic feedback leads to better user performance for all tasks.
- For the spatial accuracy task, co-location is of greater benefit than haptic feedback in task performance.
- The spatial response task is almost impossible to perform without co-located haptic feedback.
- Users' comments reflect the quantitative findings, with preferences for both haptic feedback and co-location.

4. Conclusions

This study indicates not only that haptic feedback assists interaction performance in a 3D environment, but also that co-location is a significant factor. The next step for this research is to extend it to a fully immersive VE system equipped with a larger haptic device[1] [2] [6] Head-tracking and a larger haptic workspace will allow us to investigate more fully some of the implementation problems described earlier. A more immersive system will also enable a broader investigation of the impact of multi-sensory co-location on presence.

References

[1] A. Frisoli, G. Jansson, M. Bergamasco, C. Loscos – Evaluation of the Pure-Form Haptic Displays Used for Exploration of Works of Art at Museums, *World haptics conference, Pisa, Italy*, March 18-20, 2005.C.

[2] C. Loscos, F. Tecchia, M. Carrozino, A. Frisoli, H. Ritter Widenfeld, D. Swapp, M. Bergamasco, The Museum of Pure Form: touching real statues in a virtual museum, *VAST 2004, 5th International Symposium on Virtual Reality, Archaeology and Cultural Heritage, Brussels*, Dec. 2004.

[3] Y. Hurmuzlu , A. Ephanov, D. Stoianovici, Effect of a Pneumatically Driven Haptic Interface on the Perceptional Capabilities of Human Operators, *Presence: Teleoperators & Virtual Environments, 7 (3), , pp. 290-307,* 1998

[4] E. Sallnäs, K. Rassmus-Gröhn, C. Sjöström, Supporting Presence in Collaborative Environments by Haptic Force Feedback, *ACM Transactions on Computer-Human Interaction, 7 (4) pp461-476*. 2000

[5] Sensable Technologies. http://www.sensable.com/.

[6] A. Dettori, C.A.Avizzano, S. Marcheschi, M. Angerilli, M. Bergamasco, C.Loscos, A.Guerraz, Art touch with CREATE haptic Interface, *ICAR 2003, 11th International Conference on Advanced Robotics, University of Coimbra, Portugal*, June 30 - July 3, 2003.

Creating a Virtual Window using Image Based Rendering

Jonas M. Weikop, Rasmus Stenholt, Rune E. Laursen, and Erik Granum
Computer Vision and Media Technology Laboratory, Aalborg University, Denmark
{weikop,rs,rul,eg}@cvmt.dk

Abstract

This paper describes the creation of a virtual window using image based rendering (IBR). The virtual window is an illusion of a real window created by a large screen plasma TV, a position tracker, and a database of systematically acquired photographs. Using IBR opens the possibility of providing a head tracked user with the impression of looking through a window to an entirely different, real place. The IBR technology used for the virtual window is provided by the Benogo [1] presence research project. During the process of designing and testing the window, several difficulties have been encountered and interesting observations have been made. These issues discovered in connection with the virtual window are discussed in this paper.

1. Introduction

The motivation behind creating a virtual window is the ambition to enable people to realistically experience distant places without travelling. Using a virtual window for this purpose provides a metaphor which is well-known and easily recognizable. Furthermore, an off-the-shelf large screen display can be used for the purpose of emulating the window pane itself, so the physical part of the window also has the advantage of being easy to acquire compared to more advanced VR display systems such as a CAVE [2] or a state-of-the-art head mounted display (HMD).

The possibility of creating a virtual window displaying real places is opened by the emergence of real-time IBR technology combined with modern PCs and motion tracking equipment. With IBR technology [3,4], artificial images are generated from real photographs as opposed to a detailed 3-D model. The IBR technique used, X-slits projection [3], enables IBR to display real, complex objects and places with motion parallax and stereo given a small set of images. The images used for a virtual window are acquired by a camera with a fish-eye lens moving along a line. At the time of acquisition, the camera travels sideways along the line, taking pictures at fixed space intervals. This is shown on Figure 1. Other attempts at creating a virtual window are described in [5] using a single, static environment map and [6] using a 3-D model of a room.

Once the images are acquired, images from new viewpoints in front of the acquisition line can be created. The illusion of a window is created when these viewpoints correspond to the position of a person walking in front of the line. This means that a user walking in front of the virtual window is presented with an image corresponding to what would have been seen from his/her current position through a window placed at the original acquisition line.

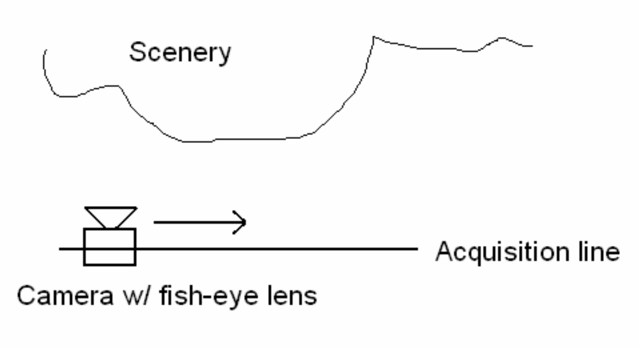


Figure 1: The acquisition of images occurs along a straight line.

2. Technical setup and tracking

The technical setup used for the window consists of a box-shaped tower containing a PC and tracking equipment. One tower wall has a hole for mounting a 42” plasma TV which is viewed from outside the tower. The only visible part of the TV is the screen. Schematically the setup works as shown on Figure 2.

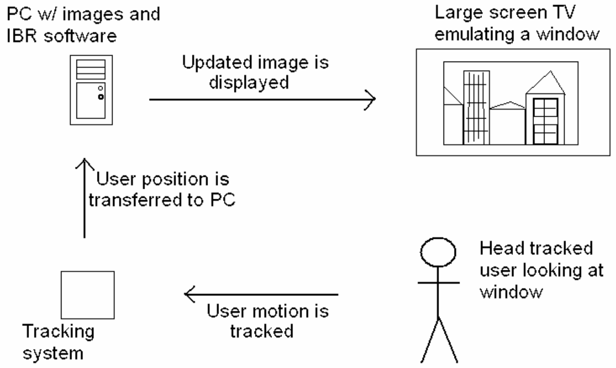


Figure 2: Schematic overview of the system.

Tracking the user is done using a Polhemus FasTrak electromagnetic (EM) tracking system which tracks 6 degrees of freedom. Only positional tracking is needed for a window without stereo, though. Using magnetic tracking for the virtual window setup has turned out to cause problems with robustness as EM tracking is sensitive to metallic objects and EM fields in the nearby environment. The TV casing is made from metal and user head movement close the window requires the most precise tracking, but it is unfortunately the region of movement that is most affected by noisy tracking. Therefore, another tracking technology would be desirable.

3. Window experiences

One of the issues with the virtual window regards the physical context of the experimental setup. When a person standing in front of the virtual window has to believe that he/she is standing in front of a real window, the experience is described as confusing. However, if the same scenario is set up in the CAVE, the experience is quite believable even though the technical set up is the same. This observation is interesting, and we have not yet come up with the definitive explanation to this phenomenon. One of the explanations, however, could be that a virtual window integrated into the tower in our visualization laboratory is far out of its original context: A small tower is unlikely to contain large scenery, e.g. a museum hall.

If people cannot relate to the physical surroundings as being ‘a place with a window’, e.g. the tower in the lab, they cannot believe that they are standing in front of a window. The CAVE setup is only different with respect to the physical context. In the CAVE you can not as easily imagine or see, what really is on the other side of the walls or the virtual window, and this may well be an important factor for maintaining the window illusion. Consequently it may be easier to believe the illusion of looking out on the world through a window from inside a room, than looking into a large world inside a relatively small box such as the tower. This is illustrated on Figure 3.

A small scene may be more acceptable for the outside-in case, since people are accustomed to looking at such scenes from an outside-in perspective, e.g. museum displays. Conversely, large scenes like the view of a city or a large room is more commonly viewed inside-out through a window. Future tests will investigate this matter further.

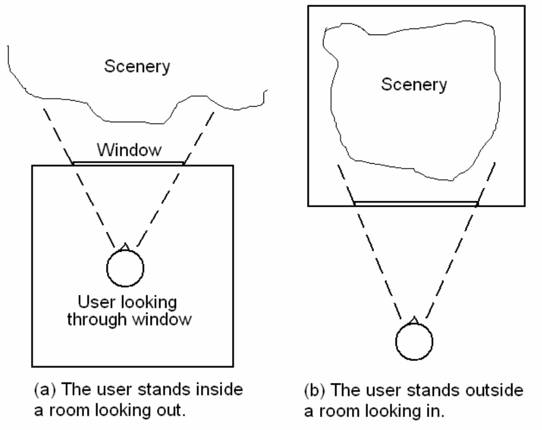
Ego motion is another important aspect of the user experience in a virtual environment (VE). During the work process with the virtual window we have made an interesting empirical observation. It seems easier to perceive the virtual window as a real window if we watch it on a video recording, than when standing in front of the actual window. Our preliminary experiments have found that people who are exposed to the recording of the virtual window quite vividly perceives the filmed window as being real. The reason for this difference has not been further investigated yet. A possible explanation, however, is that when standing in front of the virtual window, the entire body is used in the experience. When a tracked video camera records the window, the experience has changed to another medium. Some of the flaws of the system, e.g. the tracking lag, are no longer detectable. Therefore, we may perceive the recorded scenario to be more realistic.

5. Conclusions

In this paper we described some of our experiences with building a virtual window using the image based rendering technology provided by the Benogo project. The window can give an acceptable illusion of reality, when viewed from a distance that keeps the TV screen's influence on the tracker minimal, and the visualization from becoming too pixelated. The illusion of a window by means of IBR, however, still has some way to go, before it becomes entirely believable by a moving, human observer.

The solution to the persisting problems seems to require further experiments with the window. Especially experiments where the surroundings of the window are changed from the outside-in type setup of a box shaped tower to an inside-out outlook on a scene.

One of the most interesting empirical observations made during work on the virtual window is the fact that the window seems more believable when viewed on a film recorded by a tracked video-camera. This gives rise to the question why this is so, and indicates that there may be a perceptual difference between watching a film of someone else’s viewpoint in a virtual environment and experiencing that same VE first-hand.



(a) The user stands inside a room looking out.

(b) The user stands outside a room looking in.

Figure 3: The difference between looking through the virtual window (a) inside-out or (b) outside-in.

Acknowledgments

This research is funded in part by the BENOGO project under the European Commission IST programme (IST-2001-39184). The funding is gratefully acknowledged.

References

[1] URL: The Benogo project website, http://www.benogo.dk, 2005

[2] Cruz-Neira, C. et al, The CAVE, Audio Visual Experience Automatic Virtual Environment. Communications of the ACM, June 1992. pp. 64-72

[3] Zomet, A. et al, Mosaicing new views: the Crossed-Slits projection, IEEE Transactions on Pattern Analysis and Machine Intelligence, June 2003, Vol. 25, Issue 6, pp. 741-754, ISSN: 0162-8828

[4] Levoy, M. Hanrahan, P., Light Field Rendering, In Proc. SIGGRAPH 1996, pp- 31-42, August 1996

[5] Radikovic, A. S., A True Virtual Window, Master Thesis, Texas A&M University, December 2004

[6] Williams, N. et al, Monticello trough the Window, 4th International Symposium on Virtual Reality, Archaeology and Intelligent Cultural Heritage, 2003

Virtual Presence for the Web

Heiner Wolf
Jabber Virtual Presence Project
{wolf@bluehands.de}

Abstract

This short paper introduces the Web as a virtual space. Applying the location metaphor to the Web leads to the notion of virtual presence on the World Wide Web. We describe requirements for a virtual presence service, which can cope with the scale of the World Wide Web.

Keywords--- **WWW, virtual space, web awareness, virtual presence, voice over IP, Jabber**.

1. Virtual Spaces

Today there are many virtual worlds and virtual spaces. From chat over games to educational worlds. There are small ones, like virtual shopping malls and large worlds like online RPG with millions of users. There are old ones like the early MUDs (text based so called Multi-User Dungeons) and new ones with advanced 3D environments.

A common characteristic of most online virtual spaces is that they feature virtual presence. This means that users are not just in the world. They are at a certain place inside the world. They see each other and interact with their virtual neighbors. While many worlds offer communication media for separated users, there are usually more ways to interact if people are virtually close to each other. Interaction happens in many different ways. From chat, over fighting and trade to more complex actions like virtual weddings an parties.

People like to meet others in virtual worlds. This has been true since early MUDs. It can be seen million fold every day in current MMORPG. The notion of being close seems to be almost as important in virtual worlds, as it is in the real world. But there is one virtual space where millions go every day, without seeing each other: the Web.

1.1. The Web

The Web is not just a collection of linked documents. The Web is a virtual space of millions of virtual places. Each Web site is a virtual place. The Web has no global spatial dimensions, only pages are spatial, but as we surf the Web, we use the location metaphor and much less the document metaphor. We go to a Web site rather than opening a hypertext document. While we are there at a virtual location, there are other people present at the same location at the same time. They are also jumping from page to page, and reading the same content. But we cannot see them. Presence, awareness of other people and synchronous communication do not exist on the Web.

We use other tools for synchronous communication. The two most important being text chat and VoIP. VoIP is usually completely unrelated to the Web. Text chat is available on web sites in the form of chat channels. The deficiency of chat channels is that they are limited to certain pages and the Web site administrator must actively enable them. There is no synchronous communication across all Web pages.

1.2. Presence for the Web

The question is, if synchronous communication on the Web is desirable at all. We answer this question positively, because there are so many signs of and synchronous communication on Web pages like user online counters, so called shout boxes, and live chat support tools. They are isolated islands of synchronous communication.

After creating the problem, we are of course going to solve it. We present the concept of ubiquitous virtual presence for the Web. It makes people aware of each other who are at the same Web location at the same time. Virtual presence is an enabling concept for synchronous interaction on the Web.

The goal of this system is to show users as avatars directly on Web pages. They should be able to move around so that they can approach each other. They see other avatars as long as they are on a Web page. Each Web page or site is a place where people can meet and talk. Pages are the places and streets of the Web.

was macht denn die ganze meute unter der woche? arbeiten alle so viel?...

die meute vermisst dich sehr, harald dreht durch, er ohne dich nicht ausgel...

Da gibts noch das Chat fenster. Doppelklick auf User und fenster mit der ganzen Geschichte

Figure 1 Avatars on Web pages

2. A Virtual Presence System

There have been earlier virtual presence (VP) projects. Most used the term Web awareness. There were early systems, like Virtual Places [1], CoBrow (Collaborative Browsing) [2], and WebPlaces [3]. Later came centralized proprietary systems, from companies like Hypernix, NovaWiz, and Cyland.

The major difference of the system we are proposing is that it is designed to be ubiquitous and distributed, and that it will be based on open internet standards.

2.1. Requirements

We design a system that will cope with the size of the Web in terms of the number of users and the number of virtual places. Scalability is an important requirement. This leads to a distributed architecture, if we want to avoid a single large provider and a single point of control.

We want to show users directly on the Web page as avatars. Users shall be able to move their avatars and virtually approach other people. The user interface must be adapted to the location metaphor.

Since we make people aware of each other there will be a component in the network that can associate people by their virtual location. The system works with virtual locations which are derived from URLs. But still the privacy of the user must be protected. We require, that there will no URLs of users be sent over the network.

An important requirement is the use of existing standards and tools. The acceptance to install and operate VP components is much higher, if the protocols are standardized and if the software components are publicly available and already widely in use.

2.2. Implementation

Fortunately there do exist server networks, which provide all we need. Almost any distributed chat network is suitable as a VP network. Chat channels would be used as locations and channel names become locations identifiers. The participant list of a chat channel serves as the virtual presence list. And all chat systems offer a native chat capability.

Other desirable features are a distributed per user storage facility for arbitrary data to store avatars and extensibility to add virtual presence related features.

We concentrated on large chat networks with active open source communities. From these we chose the Jabber network (XMPP) [4] over IRC, because it provides additional services, like server storage.

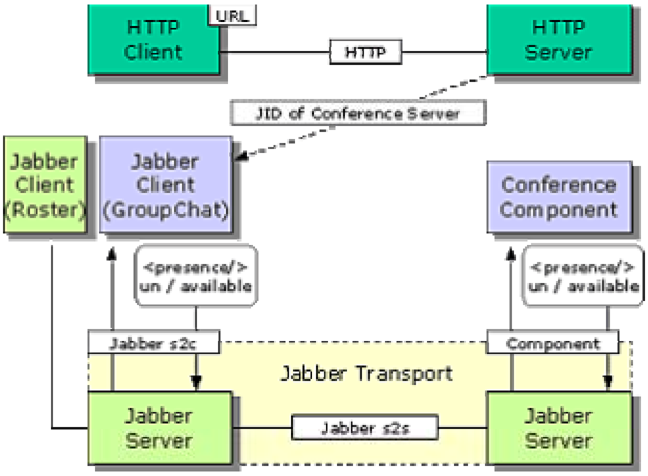


Figure 2 The Jabber network as transport and processing infrastructure

Jabber is an XML based instant message and presence network. The Jabber community developed a variety of protocols. There are open source and commercial implementations of clients and servers. The Jabber network is structured so that a client connects to a server and servers interconnect with each other. The VP network becomes an overlay network on the Jabber network. Many servers share the load of client connections and other servers share the virtual presence load and chat conversations.

2.3. A Client

From the technical point of view, we implemented a Jabber group chat client with a graphical user interface, which automatically enters and leaves Jabber chat rooms while the user is browsing the Web. From the user's point of view we developed a program, which runs in the background, and as soon as you go to a Web site, it shows your avatar on the page and the avatars of other users, who are at the same Web page at the same time.

In addition to the basic functionality of showing avatars and chat, it provides a small video for each user. It supports private chat for one-to-one conversations. Avatars can be animated figures which walk like small game characters.

2.4. Voice Rooms on Web Pages

We could even automatically add users to VoIP channels. This turns Web pages into VoIP enabled room. Users can just open a Web page and talk into the microphone. Anyone on the same page would hear them without dialing or setting up conference calls. Actually, this is the single most requested feature. A VoIP component is in development.

Conclusions

We introduced the concept of virtual presence on the Web as a solution to the apparent lack of ubiquitous synchronous communication in the virtual space of the Web. We described requirements for and the implementation of a distributed virtual presence system.

Acknowledgements

The project has been funded by SURFnet bv and bluehands GmbH & Co.mmunication KG.

References

[1] Mass, Y. Virtual Places - Adding people to the web, Proceedings of the Third International World-Wide Web Conference 1995

[2] Sidler, G., Scott, A. et al (1997) Collaborative Browsing in the World Wide Web, Proceedings of the 8th Joint European Networking Conference, Edinburgh, May 12.-15

[3] Maglio, P.P., Barrett, R., WebPlaces: Adding People to the Web, Poster Proceedings of the Eighth International World Wide Web Conference 1999, <http://www.almaden.ibm.com/cs/wbi/papers/www8/wwwplaces-abstract.html>

[4] Adams, D J (2002) Programming Jabber, O'Reilly.

Contents by First Author