# TABLE OF CONTENTS

## THEORY


Andre Gysbers, Christoph Klimmt, Tilo Hartmann, Andreas Nosperi & Peter Vorderer. *Exploring the Book Problem: Text Design, Mental Representations of Space, and Spatial Presence in Readers* ................................................................. 13

Sungkil Lee, Gerard J. Kim, Albert Rizzo, Hyungjin Park. *Formation of Spatial Presence: By Form or Content?* ................................................................. 20

Matthew Lombard, Matthew T. Jones. *Presence and Sexuality* ................................................................. 28

Blair MacIntyre, Jay David Bolter, Maribeth Gandy. *Presence and the Aura of Meaningful Places* ................................................................. 36

David Nunez. *Working Memory and Presence: Reconsidering the Role of Attention in Presence* ................................................................. 44


## EXPLORING THE SENSE OF PRESENCE

Gary Bente, Sabine Rüggenberg, Nicole, C. Krämer. *Social Presence and Interpersonal Trust in Avatar-Based, Collaborative Net-Communications* ................. 54

Masaki Emoto, Kenichiro Masaoka, Masayuki Sugawara, Fumio Okano. *Quantitative evaluation of sensation of presence in viewing the "Super Hi-Vision" 4000-scanning-line wide-field video system* ................................................................. 62

Freeman, J., Lessiter, J., Keogh, E., Bond, F.W., & Chapman, K. *Relaxation Island: virtual, and really relaxing* ................................................................. 67

Anne-Cécile Jeandrain. *Why and How Do the Telepresence Dimensions Influence Persuasive Outcome ?* ................................................................. 73

Younbo Jung, Kwan Min Lee. *Effects of Physical Embodiment on Social Presence of Social Robots* ................................................................. 80

Jari Laarni, Niklas Ravaja, Timo Saari and Tilo Hartmann. *Personality-related differences in subjective presence* ................................................................. 88

Annie Luciani. *Dynamics as a common criterion to enhance the sense of Presence in Virtual environments* ................................................................. 96

Shaleph O’Neill, Rod McCall, Michael Smyth, David Benyon. *Probing the Sense of Place* ................................................................. 104

Niklas Ravaja, Jari Laarni, Kari Kallinen, Timo Saari, Mikko Salminen, Jussi Holopainen, Aki Järvinen. *Spatial Presence and Emotional Responses to Success in a Video Game: A Psychophysiological Study* ................................................................. 112

Ruth Rettie. *Using Goffman’s Frameworks to Explain Presence and Reality* ................................................................. 117

Bernhard E. Riecke, Jörg Schulte-Pelkum, Marios N. Avraamides, and Heinrich H. Bülthoff *Enhancing the Visually Induced Self-Motion Illusion (Vection) under Natural Viewing Conditions in Virtual Reality* ................................................................. 125

Timo Saari, Jari Laarni, Niklas Ravaja, Kari Kallinen, Marko Turpeinen. *Virtual Ba and Presence in Facilitating Learning from Technology Mediated Organizational Information Flows* ................................................................. 133
Aleksander Väljamäe, Pontus Larsson, Daniel Västfjäll and Mendel Kleiner. 
Auditory Presence, Individualized Head-Related Transfer Functions, and Illusory 
Ego-Motion in Virtual Environments ……………………………………………………….. 141
Vinoba Vinayagamoorthy, Andrea Brogni, Marco Gillies, Mel Slater, Anthony 
Realism ……………………………………………………………………………………………. 148

APPLICATIONS

Rosa Baños, Cristina Botella, Víctor Liaño, Belén Guerrero, Beatriz Rey, 
Mariano Alcañiz. Sense of Presence in Emotional Virtual Environments …………… 156
Doron Friedman, Robert Leeb, Angus Antley, Maia Garau, Christoph Guger, 
Claudia Keinrath, Anthony Steed, Gert Pfurtscheller, Mel Slater. Navigating 
Virtual Reality by Thought: First Steps ………………………………………………….. 160
Gary Marsden, Shih-min Yang. Eliminating Design and Execute Modes from 
Virtual Environment Authoring Systems ………………………………………………… 168
Beatriz Rey, Mariano Alcañiz, José Antonio Lozano, Rosa Baños, Cristina Botella, 
Nuria Lasso, Jorge Osma. Emotionally controlled Virtual Environments: A New 
Tool for Enhancing Presence through Media Content Manipulation …………………. 174
Timo Saari, Niklas Ravaja, Jari Laarni, Kari Kallinen, Marko Turpeinen. Towards 
Emotionally Adapted Games …………………………………………………………………. 182
Maria Spante, Ralph Schroeder, Ann-Sofie Axelsson, Michael Christie. How 
Putting Yourself into the Other Person’s Virtual Shoes Enhances Collaboration … 190
Anne-Mie Sponselee, Yvonne de Kort, & Anneloes Meijnders. Healing Media: 
The moderating role of presence in restoring from stress in a mediated 
environment ………………………………………………………………………………………. 197
Arthur Tang, Frank Biocca, Lynette Lim. Comparing Differences in Presence 
during Social Interaction in Augmented Reality versus Virtual Reality 
Environments: An Exploratory Study …………………………………………………….. 204
Shigeru Wesugi, Yoshiyuki Miwa. Facilitating Interconnectedness between Body 
and Space for Full-bodied Presence- Utilization of “Lazy Susan” video projection 
communication system - …………………………………………………………………….. 208

MEASUREMENT

Jeremy N. Bailenson, Eyal Aharoni, Andrew C. Beall, Rosanna E. Guadagno, 
Measures of Embodied Agents’ Social Presence in Immersive Virtual 
Environments ………………………………………………………………………………………. 216
Saskia Böcking, Andre Gysbers, Werner Wirth, Christoph Klimmt, Tilo Hartmann, 
Holger Schramm, Jari Laarni, Ana Sacau, Peter Vorderer. Theoretical and 
Empirical Support for Distinctions Between Components and Conditions of Spatial 
Presence …………………………………………………………………………………………… 224
Maia Garau, Hila Ritter Widenfeld, Angus Antley, Doron Friedman, Andrea 
Brogni, Mel Slater. Temporal and Spatial Variations in Presence: A Qualitative 
Analysis ……………………………………………………………………………………………… 232
Guger C., Edlinger G., Leeb R., Pfurtscheller G., Antley, A., Garau, M., Brogni A., 
Friedman D., Slater M. Heart-Rate Variability and Event-Related ECG in Virtual 
Environments ……………………………………………………………………………………… 240
Chad Harms, Frank Biocca. Internal Consistency and Reliability of the Networked Minds Social Presence Measure .......................................................... 246
Tassos A. Mikropoulos, Evangelos Tzimas, Georgios El. Dimou. Objective Presence Measures through Electric Brain Activity .............................................. 259
Espen Nystad, Angelia Sebok. A Comparison of Two Presence Measures Based on Experimental Results ................................................................. 266
Jari Takatalo, Jukka Häkkinen, Jeppe Komulainen, Heikki Särkelä, Göte Nyman. The Experiential Dimensions of Two Different Digital Games ........................ 274

POSTERS

John C. Adams, Joshua Huling, Janet Simons. Video Cellphones, War, and Vidblinks: Exploring the Rhetorical Constraints of Time and Place ......................... 279
Cheryl Campanella Bracken, Matthew Lombard, Kimberly A. Neuendorf, James Denny, Michael Quillan. Do 3-D movies really reach out and grab you? The case of SpyKids 3-D ................................................................. 283
Brigid Costello, Alastair Weakley, Ernest Edmonds. Creative presence: Supporting artistic collaborations ................................................................. 287
Wijnand IJsselsteijn and Miriam Reiner. On the importance of reliable real-time sensorimotor dependencies for establishing telepresence ...................... 298
Rod McCall, Shaleph O’Neill, David Benyon, Michael Smyth, Fiona Carroll. A Method For Designing Virtual Places ..................................................... 305
David Nunez. A Constructionist Cognitive Model of Presence ......................................................... 313
Miriam Reiner, Gad Halevy, David Hecht, Moram Furman, Daniel Vainsencher. Discrimination Among Three Types of Anomalies in a Visual-Haptic Virtual Environment ........................................................................... 317
Rocio Ruiz Rodarte. Augmented Reality for Museum Exhibits .................................................. 321
Ana Sacau, Coutinho Susana. Presence in Educational and Entertaining Virtual Environments .................................................................................. 325
Phil Turner, Iain McGregor. Telling People About Virtual Places: A Qualitative Perspective ................................................................................ 332
Phil Turner, Susan Turner. Insideness And Outsideness: Characterizing The Experiences Of Real And Virtual Places .............................................. 340
Josh Williams, Jacquelyn F. Morie. Mining Emotion in Virtual Environments ......................... 347
Werner Wirth, Susanne Wolf, Ursina Mögerle, & Saskia Böcking. Measuring the Subjective Experience of Presence with Think-Aloud Method: Theory, Instruments, Implications ........................................................................... 351
Movement, Action, and Situation: Presence in Virtual Environments

Antonella Carassa, Francesca Morganti, Maurizio Tirassa
Università della Svizzera Italiana, Istituto Auxologico Italiano, Università di Torino
{antonella.carassa@lu.unisi.ch, francesca.morganti@auxologico.it, tirassa@psych.unito.it}

Abstract

Presence is commonly defined as the subjective feeling of "being there". It has been mainly conceived of as deriving from immersion, interaction, and social and narrative involvement with suitable technology. We argue that presence depends on a suitable integration of aspects relevant to an agent's movement and perception, to her actions, and to her conception of the overall situation in which she finds herself, as well as on how these aspects mesh with the possibilities for action afforded in the interaction with the virtual environment.

1. Introduction

What about presence? Many a prominent view in current research and literature focus on what presence is and how it develops. Presence is commonly defined as the subjective feeling of "being there" [1] [2] [3]. Several authors considered this feeling of presence as mainly deriving from the immersion in a virtual environment [4] [5] [6]. They defined presence as the result of subjective involvement in this kind of highly interactive virtual environment; presence would be strong inasmuch as the virtual system enables an inclusive, extensive, surrounding and vivid illusion: the immersive quality of a virtual reality system would be enhanced by the perceptive features and the proprioceptive feedback provided by the system. Within this perspective, different authors have developed apparently different conceptions of presence.

Sheridan [7] and Zeltzer [8], for example, described the sense of presence as the sense of being placed in a place different from the physical one. Sheridan, in particular, defined virtual presence as the subjective feeling or mental state in which a subject has the belief of being "physically present with visual, auditory, or force displays generated by a computer". Heeter [9] defined an environmental presence which is yielded by the perception that an environment exists that modifies depending on what you do and seems to consider you as present. Witmer & Singer [2] also took presence to be due to immersion, but related it to the tendency to direct attention toward selected information that is meaningful to the individual. Presence would then be comparable to selective attention, and the sense of presence would be yielded by the allocation of attentional resources. According to these authors, both involvement and immersion are needed to experience presence. This approach, while focusing on immersive properties, also emphasized the role that activity plays in directing attention within complex interactive situations.

The importance of activity in the support and the enhancement of presence in virtual reality was investigated by Flach & Holden [10], who emphasized the necessity that interaction with objects be introduced in virtual environments. On a similar vein, Zahorik & Jenison [11] focused on the role of plausibility in perception/action behaviors; the latter are dealt with in terms of affordances. Mantovani & Riva [12] highlighted the importance of freedom in the actor's action within a virtual environment, as well as the need of a thorough consideration of the social and cultural dimension of actions in both the simulated and the physical world.

In an attempt to combine immersion-based theories with activity-based ones, Sheridan [13] proposed Estimation Theory. It claims that we can never have true knowledge of objective reality; instead, we are continuously making and refining a mental model which estimates reality. This process is made possible by sensing reality and interacting with it. Immersion in virtual reality is a source of stimuli, starting from which a user would create a mental model of the virtual environment and of how she relates to it. It would be the structure of this mental model that determines whether or not the user experiences a sense of presence. Thus, even when she is uncertain about the reality of her perceptions in the virtual environment, such perceptions would be anyway close relatives of those she has in the physical world.

The specific role of interaction with technology in creating presence was firstly considered by Lombard & Ditton [14], who defined presence as the "perceptual illusion of non-mediation". In particular, according to Lombard [15] presence should be divided into those aspects which involve the perception of a physical environment (where the sensory features correspond to those of the physical world), those which involve the perception of social interaction (where the social features correspond to those of the physical world), and those which involve both. In this perspective, presence occurs when a person misperceives an experience mediated by technology as if it were a direct (that is, non-mediated) one. Presence, thus, would not be a property of technology; rather, it could vary depending on how much the user acknowledges the role of technology and could therefore be yielded by different kind of technologies.
We agree with Biocca [16] that all these aspects ought to be integrated within a more general perspective on the nature of mind and agency. It is our aim in this paper to outline such an integrated perspective. We will argue that presence depends on a suitable integration of aspects relevant to an agent's movement and perception, to her actions, and to her conception of the overall situation in which she finds herself, as well as on how these aspects mesh with the possibilities for action afforded in the interaction with the virtual environment.

2. Perception and movement

Imagine you are observing a soccer player. His kicking the ball toward the goal is realized by the increased activation of certain muscles and the decreased activation of others. It, however, involves much more than just a sequence of motor commands. True, any action ultimately consists in the realization of body movements, but how such movements are programmed and executed is much more complex than it may seem.

As a first thing, the player has to take into account, and keep track of, a whole set of physical parameters dealing with his physical features — or, better yet, with the interaction between such features and the world in which he finds himself. Thus, a player who is 160 cm. tall will have to program his movements very differently from one who is 190 cm. tall; in both cases, of course, the movements will have to be programmed for an environment that provides a certain gravity pull, a certain density of the medium (think of the muscular effort needed to realize the same body movement in the air and under water), a certain adherence of the floor surface, and so on. While executing the kick, the player receives feedback information from his own body (proprioceptive feedback from the muscles, the joints, the organs of balance, and so on) as well as from the "external" world (variation in the patterns of brightness, for example, provided by the sun and other lights; variations in the visual landscape in front of him; variations in the relative direction of sounds which he knows are motionless, like the spectators, or moving, like the other players moving beside him; and so on).

Strictly speaking, this information is neither exclusively located within the player's body nor exclusively located in the outside world. Instead, it is, in each case, relational information. The player's feeling of the friction of his foot against the grass, for example, is neither in the foot nor in the grass: it is in the physical features of the ground (its roughness, softness, and so on), in the features of the movement of the foot (its force, its direction, and so on), and even in the player's expectations (a soccer player knows, for example, that to play on a rain-soaked ground will yield different information to those he will receive from a sunburnt one).

The management of such information depends on the creation, the maintenance and the moment-by-moment reactivation of sensorimotor schemes that "tell" the player how to appropriately program his movements in the specific situation in which he finds himself, what sorts of feedback to expect from the world, and so on. That way, his body will "know" what muscular power to exert in order to achieve a certain movement; analogously, while he turns his head, his body will "take it for granted" that the world will be turning in the opposite direction, and so on. Actually, it is also on the grounds of such feedback that he will be able to know how he is executing his kick, or which point in the sequence of movements he has reached.

If the relation between his body and the world is not the right one, that is, if the execution of the programmed movement is not accompanied by the right feedback from the body and the perceived environment, then the player has a problem. Sometimes this may be a surprise, maybe even an interesting one, as it happens when we expect to touch an object, only to find that it is just a hologram. Other times, the surprise may be far less pleasurable, as it would happen if we were to find out that it's not the glass that we wanted to grasp that is a hologram, but the floor upon which we wanted to thread.

If this is true of the real world, it has to be true also of a virtual environment that aims at looking like it or simulating it, and therefore has to take into account the structural coupling [17] [18] [19] of the organism and the world.

There are several aspects to this coupling. Some will be of a comparatively high level: these will concern, for example, the degrees of freedom allowed by the system, or its ergonomics and its cognitive ergonomics: in the very same way that a car had better have the steering wheel in front of the driver, rather than behind her back, so in a virtual environment it is better to move by using the joystick than by pressing a complicated sequence of keys.

An aspect which is instead relevant at the level we are discussing in this section is that, wherever the steering wheel is located, when the driver steers her visual system ought to perceive a world which turns coherently in the opposite direction; her organ of balance ought to receive the appropriate proprioceptive feedback from the head's rotation, and so on. What Gibson [20] used to call the invariants of the physical world must remain invariant in the simulated world, if the minimal level of user's presence is not to be lost, because she cannot but take it for granted that the world will react to her actions in agreement with the basic relational laws of the body/environment interaction.

3. Action

Let us consider the soccer player again. A kick is, manifestly, the realization in the world of a certain sequence of sensorimotor programmes. It also is, however, much more than just that.

If described at a different level, the kick is an action, that is, an event which is consciously and deliberately brought about by the player as a way to physically realize an intention. That intention, in its turn, does not exist in isolation, but is born within a much more complex network of knowledge and plans.

The player inserts his kick within the set of schemes and strategies that he is forming and following instant by
instant. Not only does he know that he is participating in a soccer match, with a relevant set of rules, prohibitions, conventions, and so on; he also knows that he is participating in a specific collective action [21] within a specific strategy within a specific match. He therefore programs, knows, monitors, and controls his kick within this hierarchy of plans; he knows that, after the kick, he will have to find himself in the right position and with the right inertial push for a subsequent run or stop; he has an idea of what could happen after the kick (he may foresee, for example, that he will be able to enjoy a brief moment of rest, and therefore decide that he can afford to spend a supplementary amount of physical effort), and so on. This knowledge is not separated from the kick, and the kick is not independent of it; on the contrary, it contributes to determine, beside the motor programming of the kick and the expectation of a certain feedback, many features that are not intrinsic to the technique of the movement but to its use in a certain context.

It is not only the knowledge of the overall scheme behind the match, the specific strategy of that phase of it and of the specific moment, and the context in which it is realized, that have him choose to kick the ball toward a certain point. Moment by moment, other aspects come into the scene which concern the specific position of that player on the overall field, his ability to evaluate what promising opportunities he has available for his kick, and so on. Thus, if he thinks that the ball can reach a teammate who is both free and in a good position for kicking a goal, he will prefer, all other things being equal, to pass the ball to that mate, rather than to another one who is too far or too heavily hampered by the opponents.

The local goals of the player's actions depend on his general goals and on how he interprets them, and guide his perception of the possible opportunities for actions — in Gibson's terms [22], the affordances he has available. The opportunities that he perceives in the world, in their turn, guide his choice of local goals, as well as his revision or reinterpretation of the more general ones [23] [24].

Again, this all holds for the virtual world as well as for the "real" one. There are at least two factors to be considered here.

One is the sensibleness, and the meaningfulness, of the relations between the actions that the user can do and the effects that they have in the virtual environment. A forward movement of the joystick is better followed by a forward movement of the user in the virtual environment than by a movement to the right and slightly back.

More subtly, the definition of the possible spaces for action in the virtual environment ought to correspond somewhat reasonably to the user's expectations. We are not concerned here with the possibility of actions in the virtual world that are impossible in real life, such as flying, but with the need that such actions take place coherently and in agreement with the user's expectations. If, for example, a passage between two rooms of the virtual environment is too narrow when compared to the user's physical size, she will find it very surprising, and possibly somewhat discomforting, to be able to pass through it, because she will have the feeling that the world does not correspond to her expectations about possible and impossible actions.

What is at issue here is not the practical impossibility of certain actions, but their conceptual impossibility. A virtual environment in which the user can fly may be more sensible than one in which she can pass through a needle's eye, because our everyday experience with our body is more easily projected in the former kind of impossible experience than in the latter: we are accustomed to jumping, to seeing things below us, to viewing landscapes from high vantage points, to imagining what it would be like to be a bird, but not to perceiving sudden and dramatic changes in the size and the proportions of our body.

Another aspect that has to be taken into account is the possibility to choose between alternate courses of action, that is, the degrees of freedom granted to the user. All the rest being equal, an environment which affords several possible actions is more interesting than one which affords few. This happens because human beings need to feel that they are engaged, participating in interesting sequences of events, in interesting choreographies. If the world with which (or, better, within which) they are interacting, be it virtual or "real", is not interesting enough, humans will just get bored, and tend to move their presence toward different worlds, as it happens in daydreaming. The monotony of an environment, therefore, tends to decrease the feeling of presence within it, because the user will have the time, the space, and the cognitive necessity to imagine that she is elsewhere, so as to keep herself engaged in a sequence of actions and events capable of stimulating and maintaining her interest.

While the first aspect we have discussed here brought us back to the previous section, the latter brings us forward, to a further level of analysis of the interactions between the human beings and the world, which we will discuss in the next section.

4. Situation

Let us go back to our soccer player once again. We saw how he decides his actions according to the affordances he perceives in the world, and programs his movements according to the sensorimotor schemes that are part of his normal, and mostly unaware, abilities to move in the world. But this complexity still does not provide a full description of the player's presence and experience.

In the player's subjective perspective, each action that he performs plays a role within a narrative that he tells himself concerning what is going on, what he is doing there and why, with what further and future perspectives, and so on. More precisely, each action that he considers or performs plays a role within a complex weave of such narratives, each contributing to the overall meaning of his being there, on that field, in that very moment, choosing to perform a certain kick, as well as to the specific body movement which ultimately shapes the material counterpart of his mental state. Each narrative may be viewed as a choreography [25] in which the player features as the protagonist; each has an intrinsically autobiographical and social nature, and the overall weave thus results from the
whole previous history of that individual (which includes, of course, his current and past hopes, dreams, and expectations for the future).

Think of a player who is young and full of hopes, one who is so aware of his own talent as to just take it for granted that he deserves to play in a much better team, but who has always been kept on the bench by the coach. The first time he enters the playing field, that match will become hugely important. He might tell himself a story like "I'm here in this lousy arena, with these good-for-nothing mates, but in a few months I'll be playing in the Premier League — then they'll see". This story will contribute in letting this player see certain spaces for action rather than others. On the one hand, his choices will be affected by his eagerness to show his talent and worth; he might thus have a tendency to not pass the ball, keeping it for himself in the hope to draw everybody's attention, and to have the opportunity to goal. On the other hand, even when he passes the ball, the excitement and the anxiety given by the awareness of the importance of a good performance might worsen the performance itself, by hampering his ability to play in the smooth and precise fashion he has learned during his training. An older, more experienced player will probably behave very differently, because many crucial factors are different in him: his drives and motivations, his self-awareness, his aspirations, his knowledge of his own weaknesses, and so on. In a word, the stories he tells himself will be very different.

These considerations may be brought back to our discussion of virtual environments. A first remark concerns the different ways of interacting with technology that different users bring with themselves according to their narratives concerning the environment itself. A user with a sharp, and maybe a little anxious, awareness that she has to deal with a technological artifact will interact with it differently to one who is capable of letting such awareness go to the background and of focusing on what the environment affords. At least in part, thus, the "transparency" of technology depends on the user rather than on the artifact. While these differences may probably be made less sharp with suitable training, they can never disappear, if only because it is not always possible or worth giving a user such training.

There is, however, a second consideration, which has nothing to do with training or with anxiety caused by technology. The interaction that a user has with the virtual environment is driven by the narrative that she tells herself about her being there; such narration depends, in its turn, on her general and local reasons for interacting with the environment, as well as on her individual history and personality.

Think of a flight simulator and some of its possible users. The engineer who designed it will enter the environment in search of possible bugs and mistakes, so to be able to correct them before putting the simulator on the market. An officer, in charge of selecting which of several flight simulators available better fits the needs of the Air Force, will try to pick features like the smoothness and the believability of the interaction with the environment, or to assess the cost/quality ratio of the product. A pilot who uses the simulator to learn to fly a new fighter without the risks and expenses of a real test will focus on the limits of the airplane's maneuverability. When the simulator, now an obsolete model from the military viewpoint, will finally be launched in the electronic games market, a thirteen-years-old will use it with still a different set of purposes, paying no attention, for example, to how many flight accidents she may have, at least until her parents let her keep on playing.

Each of these users will experience a variable sense of presence, according to how much the environment will suit her needs, her interests, and the stories that she brings with herself in the interaction. Searching for bugs is something very different from trying to impress on one's friends.

Conclusions: Presence in virtual reality

We distinguished three levels in the interaction of an agent with her world, be it real or virtual: that of the situation, that of the action, and that of body movement and perception. These levels are not reducible to one another; instead, each of them contains the subsequent one, like the nesting Russian matrioska dolls, and returns as a feedback on the previous one. Thus, a circular relationship of co-determination exists between them.

Normally, an agent will not think of her movement in terms of a motor sequence (unless, of course, she has any reason to do so, in which case the motor sequence may become the action or the situation). Instead, she will choose and perform actions whose goals are part of a broader situation, which she represents as the activity, or the weave of activities, in which she is participating at each moment. This activities are, in their turn, supported by goals, values, knowledge, and roles that give them meaning, boundaries, a history, and possible directions of development.

Therefore, an individual will represent herself not as a monad with no history who "behaves" in an objectively given world, but as an agent who carries on a narrative about herself in the world. What is of interest to her is to follow complex flows of meaning relevant to the different choreographies in which she finds herself. Her representations and actions create her participation to such choreographies from moment to moment.

How does this conception of mind and agency, a constructivist and interaction-based one [19], affect our conception of experience and presence in virtual reality? The kernel of our position is that what is designed is not an intrinsic property of the entities alone, but a property of the interaction between the agent and the entities [26].
The availability of the affordances depend on the activities in which the agent is participating at each moment. Such activities result from the agent's previous history, which goes to constitute both her memory and the processes of recognition and reconceptualization that make such history immediately useful in the current interaction [27][28].

Thus, what happens on entering a virtual environment is not that the user leaves behind the real world, whose role is, at most, that of an external disturbance which decreases or damages presence in the virtual environment. Instead, we bring our experience inside the virtual world, and, in turn, we integrate the virtual world in our experience, which will go to sediment in our overall future history and projects.

Something similar always happens in fiction. A book, a movie, or the tales that are told around a fire are familiar to us because we recognize their meanings in the light of our previous history, and integrate them in the weave of narratives in which we will live from that moment on.

Of course, a virtual environment differs from a book or a movie, in that, while the latter ask and afford us to just put ourselves in the characters' shoes, finding there a meaning of interest to us, in the virtual environment we can actually perform action and receive the corresponding feedback. The possibility of first-person action in the world, that is, the possibility of contributing to the generation and maintenance of world dynamics, and of receiving in turn the possibility (and the need) to generate and maintain our cognitive dynamics, is another crucial factor of presence, that is, of our capability to feel that we are participating in the world in which we find ourselves.

Beside this difference between fiction and virtual environments, our feeling of presence depends, in both cases, on the possibility for us to bring in some interesting meanings, and to integrate them in interesting ways with the meanings that the book, the movie, or the virtual environment proposes to us. In this respect, what counts is not necessarily the writer's or the designer's virtuosity: virtual, or fictional, worlds are not interesting because they provide a perfect duplicate of the array of stimuli that the real world provides, but because they grant us the possibility of recognizing stories that we feel as familiar, that is, stories in which we can bring our meanings, and as interesting, that is, stories which are worth integrating in our future experiences.

When the interaction is such that a good feeling of presence is generated and maintained, several other things will become possible. The first is that the mind supplies with its own capacities, at least to a certain extent, to the "low fidelity" of the simulated world. As we said above, what makes the difference is not technological perfection, but the type of interaction that technology permits.

Secondly, just as actions support presence, so does presence support actions. The feeling of presence is satisfactory when the user manages to make an overall sense of her interaction with the environment. When this happens, she will also manage to make it useful and interesting for her future narratives: in simple terms, she manages to learn something.

Thus, in experiencing a virtual reality environment, the user will bring with herself everything that she has been up to that moment, and her experience with the media will add to her "cognitive history". This may mean that she will have acquired knowledge (concerning the Qumran scrolls, or how to fly an airplane), or that she will have spent a few hours shooting nasty green aliens that want to invade the Earth, or, in the worst case, that she will have suffered from cybersickness — even this is an experience, however unpleasing, that will affect her possible futures.

What the designer does is thus to create an envelope within which interaction with the virtual environment may acquire a weave of narrative meanings. The goal of such enterprise is not intrinsic to the virtual environment, but is born out of the structural coupling between the user and the environment — and, sometimes, between the user, the environment, and a supervisor or a tutor who guides the interaction, as it may happen, for example, in an environment designed for neuropsychological or motor rehabilitation [29].

Acknowledgements

The authors are listed in alphabetical order. Antonella Carassa and Francesca Morganti were funded by the Italian Ministry of Education, University and Research (MIUR), Cofin 2003 prot. 200311 9035 ("Fattori percettivi e cognitivi in compiti di navigazione spaziale in ambienti di realtà virtuale"). Francesca Morganti was also funded by the European Union, project on I-Learning (FP5 IST/FET 2001 - 38861). Maurizio Tirassa was funded by the University of Torino, Fondi individuali di ricerca (Quota ex-60%) for the year 2003. Antonella Carassa wishes to express her gratitude to Giuseppe Mantovani for encouraging her to enter this research area.

References

Exploring the Book Problem: 
Text Design, Mental Representations of Space, and Spatial Presence in Readers

Andre Gysbers¹, Christoph Klimmt¹, Tilo Hartmann¹, Andreas Nosper¹ & Peter Vorderer²

¹ Department of Journalism and Communication Research, Hannover University of Music and Drama
² Annenberg School for Communication, University of Southern California, Los Angeles
(Contact: andre.gysbers@ijk.hmt-hannover.de)

Abstract
Based on the MEC model of Spatial Presence, an experiment (N = 34) was conducted that explores the development of Spatial Presence in readers of text. Two techniques for text writing that may facilitate Spatial Presence were derived from the model and implemented in the stimulus production: the number of verbal spatial descriptions (cues) was varied (low versus high), and one text version included repeated instructions to imagine the portrayed space as vividly as possible. Findings indicate that the mental representation of described spaces is more vivid if much space-related information is presented by the text, but Spatial Presence is higher if less spatial cues are provided. These partially surprising results are discussed with respect to the MEC model and future directions of cross-media theorizing about Spatial Presence.

Keywords--- Spatial Presence, Book Problem, Experiment, Measurement, Questionnaire.

1. Introduction: The Book Problem

Historically, research on Spatial Presence has focused on powerful new media technologies that are very obviously capable to create an illusion of “being there”, such as virtual environments or IMAX theaters. More recently, the emergence of Presence has also been hypothesized in the context of less immersive media, such as television [1]. Even books have been assumed to elicit the experience of Presence in their readers [2]. One argument for the capacity of books to induce a sense of Presence is the amazing aesthetic experience of literary texts that often include detailed and vivid portrayals of spatial configurations. If Spatial Presence can occur in users of non-immersive media such as books, however, direct sensory experience cannot be the only mechanism of Presence. For readers, perceptual processes are apparently less important for the facilitation of Spatial Presence than higher-order mental activities such as cognitive involvement and imagination. By stimulating those higher-order processes, text and small-screen media may be capable to compensate for their lack of power in terms of creating an illusion of “being there”. The goal of this study was to investigate readers in order to explore those mechanisms of Spatial Presence that do not rely on sensory experience, but on imaginary processes. As a theoretical framework, we employed the MEC Model of Spatial Presence, as it predicts the emergence of Presence under conditions of low immersion and the absence of direct sensory input (see section 2.). Hypotheses derived from this model were experimentally tested (3.). The obtained results (4.) allow for some interesting progress in the explanation of the book problem (5.).

2. Spatial Presence in Readers: The MEC Model

The conceptual model of Spatial Presence advanced by Vorderer et al. [3] explains the occurrence of Presence as a two-step process (see figure 1). It explicitly covers Presence phenomena across different media and is applicable to information processing and Presence experiences during reading. According to the model, readers generate a mental representation of the spatial environment portrayed in the text. They do so by processing the space-related information included in the text (bottom-up component) and by adding spatial images and space-related knowledge that was already available to their mind before exposure to the text (top-down component). Readers combine text-based and knowledge-based information to create a so-called spatial situation model (SSM) of the described environment. For example, readers of a travelogue form a
mental representation of how the beautiful valley the author has written about would look like. This mental imagination is the first step in the formation of Spatial Presence.

However, the mere existence of an SSM is not the same as the experience of Presence. According to Vorderer et al. [3], Spatial Presence occurs only if the individual considers him/herself to actually be located within the space that is represented in the SSM (and does no longer believe to be part of the non-mediated, real environment). Spatial Presence means to accept the mediated environment as personal reality or “primary ego-reference frame” (PERF; [4]), to which one’s thinking and (imagined or factual) actions are directed. Just imagining how the nice valley would look like is not enough to feel Spatial Presence, then; rather, people have to regard themselves to be located in that valley.

![Figure 1 MEC Model of Spatial Presence](image)

**Figure 1 MEC Model of Spatial Presence**

The transition from the SSM to the actual experience of Presence is explained through the theory of perceptual hypotheses [5]. Once the processing of mediated information has allowed for the creation of a rich, consistent and enduring SSM, people test the perceptual hypothesis that the space represented in the SSM is their actual surrounding (the so-called medium-as-PERF-hypothesis). If they accept this hypothesis, Spatial Presence emerges; if they reject it, they remain at the stage of having a vivid spatial impression in the mind’s eye but still perceiving themselves as part of their real surrounding. Both media and personological factors determine whether in a given situation the medium-as-PERF-hypothesis is accepted (=formation of Spatial Presence) or rejected (= failure to reach the stage of Spatial Presence).

One of the most prominent stimulus characteristics related to spatial perception and Presence is spatial cues [6]. They allow users to identify boundaries and other spatial structures in the stimulus field. Most spatial cues address the visual (e.g. [7]) or the auditory [8] modality. Spatial cues may be inserted in written language as well. Verbal descriptions of spatial structures do not trigger perceptual processes such as the visual identification of an edge, but rather stimulate space-related cognitions based on existing knowledge structures. Textual spatial cues do not provide direct information on the spatial attributes of the portrayed environment, but inform the readers about which (class of) spatial cognitions they should retrieve from their memory to complete their SSM [3]. The quantity and quality (e.g., comprehensibility) of verbal spatial cues should therefore influence the richness and vividness of readers’ SSM, which consequently would also affect the outcome of the readers’ test of the medium-as-PERF-hypothesis, because a rich and vivid SSM increases the probability that the individual will perceive him-/herself to be located in the environment represented in the SSM. The way in which a text describes spatial structures would then have an indirect impact on the emergence of Spatial Presence.

As higher-order processes such as imaginations are presumed to be most important in the development of Presence during the consumption of text-based media [3], the inclusion of verbal spatial cues is not the only technique of message design relevant to the ‘book problem’. The individual’s motivation to be absorbed in a book’s world and to have intense experiences during reading may affect those higher-order processes [9]. Interested, open-minded readers may engage more actively in imagining the book’s world and thus insert more and very vivid spatio-visual information from their memory to the SSM [3]. From the perspective of text design, this assumption would suggest to insert phrases that animate readers to engage in (spatial) imagination as actively as possible. Such appeals may increase readers’ attention to the text’s spatial cues (which is a pre-condition of the development of an SSM, see [3]), would motivate them to make more active-cognitive contributions to the top-down processes involved in the formation of the SSM, and, most importantly, could make readers actively directing the test of the medium-as-PERF-hypothesis towards the acceptance of the mediated space as PERF (i.e., Spatial Presence) by actively searching for information confirming the hypothesis and suppressing contradicting information. Because of these multi-level effects, imagination instructions in a text are hypothesized to cause a higher probability of Spatial Presence in readers.

In sum, the MEC model allows to elaborate two cognitive processes which specify the “reader imagination” that leads to Spatial Presence. One is the construction of a mental representation of space (SSM) that partly builds on spatial descriptions in the text, and the other is the motivated active imagination of the portrayed space that benefits from vivid spatial memories and a less critical evaluation of the text’s spatial description (support of the medium-as-PERF-hypothesis). According strategies of text design can create Spatial Presence in readers: One such strategy is the inclusion of a sufficient amount and quality of spatial-verbal descriptions (cues), and the other is the
repeated invitation to engage in spatial imagination as actively as possible.

3. Method

To test the assumptions that more spatial descriptions and imagination instructions increase Spatial Presence in readers, an experiment with three conditions was conducted. A text portraying a museum was produced and experimentally varied with respect to the number of spatial cues (very few cues versus many cues) and to the number of encouragements to imagine the museum as vividly as possible (no encouragement versus repeated encouragements). Presence was measured after reading the text with the MEC Spatial Presence Questionnaire (MEC-SPQ; [10]; see 2.2.).

3.1. Stimulus materials

A text describing a visitation of a Mozart Museum was produced in three different versions. All versions were in German language and based on the same master document. The described spatial structure contained one large entry hall and two additional stories with one hall and three exhibition rooms each. The different levels were connected by stairways. All halls and rooms were portrayed in the text. The manipulation of the independent variables verbal spatial cues and imaginary encouragement was realized by adding information to the basic text (version number 1). Version number 1 contained only a very small number of spatial cues, whereas version number 2 contained a large number of spatial cues. The following examples illustrate the differences between these two versions:

Example of version 1: “The entrance hall with a dark, wooden floor, decorated with a dark red carpet, holds a warm atmosphere, although there are only a few objects in it.”

Corresponding example of version 2: “The entrance hall with a dark, wooden floor, decorated with a dark red carpet, holds a warm atmosphere, although there are only a few objects in it. It is 30 meters long, 15 meters wide and five meters high.”

Version number 3 contained the same (large number of) spatial cues as version number 2, but included repeated encouragements to imagine the museum as vividly as possible.

Example of version 2: “The second floor starts with a long and narrow hall, covered by a wooden floor, a red carpet and several paintings of young Mozart on the grey walls.”

Corresponding example of version 3: “Try to imagine this floor as precisely as possible: The second floor starts with a long and narrow hall, covered by a wooden floor, a

3.2. Dependent Measures

After reading the text, the participants completed the MEC Spatial Presence Questionnaire (MEC-SPQ; [10]). This survey tool is based on the model of Spatial Presence experiences proposed by Vorderer et al. [3] and has been developed according to standard social scientific scale development and validation processes [10]. The MEC-SPQ measures each of the following constructs which are supposed to be involved in the experience of Spatial Presence [3] by five-point Likert scales. On the whole, all scales revealed very satisfactory internal consistencies. Two items of the involvement scale and one item assessing SoD were excluded because of low item remainder coefficients. The numbers of items and Cronbach’s alpha values are reported in brackets:

- Attention Allocation (8 items; \(\alpha=0.86\))
- Spatial Situation Model (SSM) (8 items; \(\alpha=0.87\))
- Spatial Presence: Self Location (SPSL) (8 items; \(\alpha=0.92\))
- Spatial Presence: Possible Actions (SPPA) (8 items; \(\alpha=0.94\))
- Higher Cognitive Involvement (6 items; \(\alpha=0.71\))
- Suspension Of Disbelief (SoD) (7 items; \(\alpha=0.92\))
- Domain Specific Interest (DSI) (8 items; \(\alpha=0.89\))
- Visual Spatial Imagery (VSI) (8 items; \(\alpha=0.85\))

The complete questionnaire can be obtained from www.presence-research.org.

3.3. Sample and Procedure

34 students participated in the study. Each person was randomly assigned to one of the three experimental groups (between-subject design). Versions 1 and 3 of the text were read by 12 participants each, Version 2 was read by 10 subjects. The mean age of the participants in the groups was between 24 and 25 years. The groups were nearly equalized by gender.

The participants were invited individually to a quiet room with controlled lighting conditions and were asked to read the stimulus text for seven minutes at normal pace. They were told that the research was conducted to find out more about general experiences of media users during the reception of texts. They were also informed that there was no need to read the text completely in seven minutes; rather, subjects were suggested to read “just like they would do at home”. After seven minutes were over, participants completed the MEC-SPQ, were informed in more detail about the research interest (the measurement of Spatial Presence), received 10 EUR as compensation, and were thanked and dismissed.
4. Results

Most participants (n=26) read four or five pages of the text, four subjects stopped their reading on the third page, and another four subjects managed to read six pages (M=4.47, SD=.86). When the amount of text increased due to additional spatial cues or instructions, the average number of pages that were read slightly decreased (few spatial cues, M=4.67; many cues, M=4.50; many cues + instructions, M=4.25). However, these differences were not significant.

Overall, domain specific interest (DSI) for the presented topic was rather low (M=2.35, SD=.80), but did not significantly differ between the experimental groups. Self-reported scores for visual spatial imagery (VSI) were nearly distributed normally (M=3.41, SD=.73) and very similar across all three groups.

Mean score for attention was M=3.69 (SD=.68), with none of the participants scoring lower than 2.0 on the attention scale. These results indicated that most participants focused their senses on the text (as a basic requirement for initiating spatial presence experiences). The text version with many spatial cues and additional instructions yielded the highest attention score (M=3.90), followed by the basic text version with few spatial cues (M=3.75) and the “medium” text version (M=3.35). These differences were not significant, and no substantial influence of domain specific interest on attention was observed (r=.01, n.s.)

With regard to building a spatial situation model (SSM), participants reported a mean score of M=2.86 (SD=.69). Obviously, the experimental manipulation of the text was successful at this pre-level of spatial presence experiences (see figure 2). Readers of the basic text version with few spatial cues scored lowest on SSM (M=2.51). A large number of spatial cues evoked a more intense SSM (M=2.88), and additional instructions to precisely imagine the described setting resulted in an even higher SSM score (M=3.19). To test for the main effect of text version, an analysis of variance was performed on the SSM scale. According to the model proposed by Vorderer et al. [3], besides media factors, attention and VSI are potential factors that could influence the creation of a SSM and therefore were entered as covariates in the analysis. Both covariates were positively correlated with SSM (attention, B=.28, p<.10; VSI, B=.38, p<.01), and the main effect of the text version was still significant after controlling for these variables (F(2,33)=5.13, p<.05). Post-hoc comparisons showed that only the difference between the “few spatial cues” and “many spatial cues + instructions” groups was significant (Bonferroni correction, p<.05). The difference between “few spatial cues” and “many spatial cues” groups was close to significance (p<.10).

The MEC-SPQ included two subscales to assess spatial presence experiences, self location (SPSL) and possible actions (SPPA). Both scales were highly intercorrelated (r=.72, p<.01). However, factor analysis with oblique rotation showed that all items had higher factor loadings on their respective component and thus could be separated correctly. Despite the descriptive, non-immersive nature of the text, a broad variety of different levels of spatial presence was reported by the readers. As expected, SPSL scores (M=2.88, SD=.88) were significantly higher than SPPA scores (M=2.10, SD=.88; t(33) = 7.02, p < .01).
Comparison of SPSL and SPPA scores across the experimental groups revealed a non-expected pattern, which was opposite to the SSM results. For both spatial presence subscales, highest average scores were yielded by the basic text version with few spatial cues (SPSL, M=3.10, SD=1.00; SPPA, M=2.50, SD=1.08). Additional spatial cues and instructions did not result in more, but contrarily lower values for both SPSL and SPPA (many spatial cues: SPSL, M=2.99, SD=.59; SPPA, M=1.88, SD=.52, many cues + instructions: SPSL, M=2.57, SD=.92; SPPA, M=1.88, SD=.81) (see figure 3).

Analyses of variance were performed on both Spatial Presence scales, individually. As both involvement and SoD are supposed to mediate the transition from SSM to Spatial Presence experiences [3], these variables were entered as covariates. However, the analysis exposed no significant between-group differences (SPSL, F(2,33)=.93, ns; SPPA, F(2,33)=.69, ns). Interestingly, SoD was a significant covariate for SPSL (B=.32, p<.05) and positively correlated with SPPA, as well (B=.21, ns), whereas involvement was a significant covariate for SPPA (B=.40, p<.05), not for SPSL (B=.004, ns).

Exploring the data for involvement and suspension of disbelief (SoD) revealed similar patterns. Overall, participants showed a great variety in both constructs (involvement, M=2.87, SD=.87; SoD, M=3.17, SD=.14). Reading the text versions with more spatial cues or instructions did not lead to higher levels of involvement or SoD. On the contrary, the basic version of the text with few spatial cues yielded highest mean scores for both involvement and SoD (see figure 4). The differences were not significant, however, and in view of the ANOVA results, especially the involvement scores have to be interpreted very carefully (involvement, F(2,33)=.74, ns; SoD, F(2,33)=.20, ns).

5. Discussion

Based on the MEC model of Spatial Presence, an experiment was conducted to explore if and how Presence occurs in readers of texts. Two elements of message design which were derived from the model were supposed to facilitate Spatial Presence during reading: the number of spatial descriptions (cues) in the text and the integration of explicit instructions to imagine the described spatial environment as vividly as possible. It was argued that both design techniques would affect the quality of the mental representation of space (SSM), and that the imagination instructions would in addition cause motivational support for the medium-as-PERF-hypothesis. According to the theoretical model, both mechanisms would enhance the feeling of Spatial Presence.

Interestingly, the findings support our assumptions only partially. With respect to the quality or strength of the SSM, the hypothesized effect both of number of spatial cues and imagination instructions was empirically confirmed. However, the average values in Spatial Presence (both scales) were highest in the condition that was expected to display the lowest scores (few spatial cues, no instructions), whereas the condition that was supposed to hold the highest Spatial Presence scores (many spatial cues plus instructions) turned out to have the lowest values. In
other words, the strength of the SSM was no important factor in determining the confirmation of the medium-as-PerF-hypothesis (and consequently, no determinant of Spatial Presence). Rather, Spatial Presence values were higher if the vividness of the SSM was lower. At first sight, this empirical relation is contradicting the assumptions of the MEC model.

However, the findings on the internal processes (user variables) shed some light on how this relationship might be explained. As books are especially low-immersive media, the importance of user factors (involvement and suspension of disbelief) for the emergence of Spatial Presence had been hypothesized beforehand. The results indicate that SoD was most intense in the experimental condition with a few spatial cues and no imagination instructions, whereas involvement displayed only marginal differences across conditions. One plausible interpretation of these data is that detailed spatial descriptions and imagination instructions in the text facilitated the construction of strong, vivid SSMs in readers, but at the same time, those text elements limited readers’ imagination and fantasy regarding their ‘active illusion’ to be located in the museum. Readers have a better impression of how a described space looks like if they receive more information, but their transportation [11] into the text world is inhibited, because this process obviously requires more degrees of freedom for imagination. The high SoD value in the “low spatial cues + no instruction” condition supports this interpretation, because it points at more active imagination processes in the readers of this experimental group.

Of course, methodological limitations have to be mentioned when this interpretation is considered. First, the number of subjects is small, and variance in most investigated variables was quite large. Technically, these circumstances increase the importance of chance and suggest being rather conservative in interpreting the findings. Second, the text stimulus employed in this study had not been written by a professional author, but was our own creation that might have led to unexpected (and undetected) effects of non-typicality in (some) readers. The differences found between the groups may therefore in part be caused by other variables than those discussed here. Third, reading time was limited to only seven minutes, which may have been not sufficient to allow readers to accommodate to the laboratory context and to enter a typical reading condition in which imaginations of Spatial Presence normally would occur.

These limitations and the partly surprising results suggest an extension or replication of the study, which should employ other genres of text (from professional writers, such as a travelogue or a novel) and expand the number of subjects to allow for a more reliable data analysis. But the findings reported here already foreshadow that the formation of Spatial Presence may have to be modeled differently than predicted by the applied theory. A solid SSM is perhaps not the key determinant of Spatial Presence, as the precise description of space forces the readers to adjust their mental representation to many details, which would hinder them from generating the illusion to be located within the described space. According to this preliminary study, readers will more likely succeed in imagining self-location in the text’s world if they are given more freedom to picture the spatial arrangements in their mind’s eye. Whereas mental representations of space depend on external information, the feeling of Spatial Presence would rather be facilitated if the text leaves the specific spatial configuration of the portrayed environment open to the individual reader’s spatial imagination. The present study has shed some light on the book problem, then, but has raised some questions about necessary distinctions in the modeling of Spatial Presence between the consumption of texts and other (immersive and/or audiovisual) media. With respect to existing theories of Spatial Presence [3], specifications for different media or modalities addressed by a medium may be required if our findings can be confirmed by further research.

Acknowledgements

This research has been funded by the European Community, FET Initiative “Presence Research Activities”, Project “Presence: MEC” (IST-2001-37661). We thankfully acknowledge the EC’s support. We also thank our colleagues from the MEC project for their support, especially Holger Schramm for the writing of the experimental stimuli.

References


Formation of Spatial Presence: By Form or Content?

Sungkil Lee¹, Gerard J. Kim¹, Albert Rizzo², Hyungjin Park¹

¹Dept. of Computer Science and Engineering
Pohang University of Science and Technology (POSTECH)
Pohang, Korea

²Institute of Creative Technologies
University of Southern California
Los Angeles, CA, USA

{gkim@postech.ac.kr, arizzo@usc.edu, yskill@postech.ac.kr}

Abstract
Spatial presence, among the many aspects of presence, is the sense of physical and concrete space, often dubbed as the sense of “being there.” This paper theorizes on how “spatial” presence is formed by various types of artificial cues in a virtual environment, form or content. We believe that spatial presence is a product of an unconscious effort to correctly register oneself into the virtual environment in a consistent manner. We hypothesize that this process is perceptual, and bottom-up in nature, and rooted in the reflexive and adaptive behavior to react and resolve the mismatch in the spatial cues between the physical space where the user is and the virtual space where the user looks at, hears from and interacts with. Hinted from the fact that our brain has two major paths for processing sensory input, the “where” path for determining object locations, and “what” path for identifying objects, we categorize the sensory stimulation cues in the virtual environment accordingly and investigate in their relationships as how they affect the user in adaptively registering oneself into the virtual environment, thus creating spatial presence. Based on the results of series of our experiments and other bodies of research, we postulate that while low level and perceptual spatial cues are sufficient for creating spatial presence, they can be affected and modulated by the spatial (whether form or content) factors. These results provide important insights into constructing a model of spatial presence, its measurement, and guidelines for configuring location-based virtual reality applications.

Keywords--- Spatial Presence, Model, Spatial Perception, Where, What, Form, Content, Brain, fMRI, Sensory Mismatch, Adaptation, Disorientation, Questionnaire, VR System Design, Dichotomy, Immersion.

1. Introduction

Starting from the simple notion of “feeling of being there,” presence has been developed into a multi-dimensional concept over the years. Scholars now generally agree that there are different types of presence, such as spatial presence, social presence, and psychological (or conceptual) presence [7]. Among them, “spatial” presence (also known as physical presence) refers to the sense of physical and concrete space, often dubbed as the sense of being there (e.g. virtual environment). Spatial presence bears particular importance to the virtual reality (VR) “technologists”, interested in providing location-based experiences, because it is seemingly (although not proven) more dependent on the “form (or system)” factors of the VR content.

In fact, there has been a lot of debate over the so called form vs. content issue. In a recent article, Slater argued that presence was about “form,” rather than “content” [13]. He stated that, “… presence is the response to a given level of immersion … presence is about form, the extent to which the unification of simulated sensory data and perceptual processing produces a coherent ‘place’ that you are ‘in’, and in which there may be the potential for you to act ….”. Our interpretation is that what Slater refers to as presence is actually “spatial” presence (rather than general presence) and naturally, he believes that it has more to do with low level spatial cues as perceived through a particular system configuration (“form”), and less with high level “content” factors like story, pictorial realism, attention, game elements, etc.

This paper theorizes on how “spatial” presence is formed by various types of artificial cues in a virtual environment, form and content included. We believe that spatial presence is a product of an unconscious effort...
to correctly register oneself into the virtual environment in a consistent manner. We hypothesize that this process is perceptual, and bottom-up in nature, rooted in the reflexive behavior to react and resolve the mismatch in the spatial cues between the physical space where the user is and the virtual space where the user looks at, hears from and interacts with. This view is in agreement with that of Slater in terms of the importance of the form factors.

In the rest of the paper, we go over series of experiments to test and confirm our theory on the formation of spatial presence in virtual environments. Starting with the premise, the relative importance of form factors to spatial presence and self registration, we first investigated in the relationships among various types of form factors as how they affect spatial presence. We make a note that although many studies have identified important factors to promoting user felt presence, their relationships and interactions have not been fully investigated [7]. Our initial hypothesis was that, among the form factors, spatial cues such as stereoscopy, shadow, and relative motion would be more contributing to spatial presence than detail cues such as geometric and texture resolutions. However, this hypothesis was disproved in our first experiments. This led to two other experiments, one that tested the effects (toward the spatial presence) of sustained attention (a content factor), and the length of exposure along with various form factors.

In the end, based on the experimental results (and other related work), we postulate that while low level and perceptual spatial cues are sufficient for creating spatial presence, they can be affected and modulated by the spatial (whether form or content) factors. Furthermore, we discuss the implication of this conclusion to constructing a model of spatial presence, measuring it and appropriately designing the questionnaire, and guidelines for configuring location-based virtual reality applications. We make a note that some procedural details of the experiments were omitted in the paper for lack of space. We only report results that are statistically significant with the p-value of below 0.005.

### 2. Related Work

#### 2.1 Presence: The Dichotomy

One of the important and defining goals of virtual reality systems is to create “presence” and to fool the user into believing that one is, or is doing something “in” the synthetic environment. Many researchers have defined and explained presence in different ways [7]. Historically, in the context of virtual reality, the concept of presence has been associated much with spatial perception as its informal definition of “feeling of being there” suggests [5][7]. Pausch et al. associated immersion and presence to one’s establishment of 3D reference in space [9]. Similarly, many studies have identified system elements that contribute to enhanced user felt presence, and many of them are spatial or perceptual cues such as providing wide field of view (FOV) display, head tracking, stereoscopy, 3D sound, proprioception, maps/landmarks, and spatial interaction [7].

Other studies in presence have challenged this view and attempted to widen the concept to include psychological immersion, thus linking higher level and “non technological” elements (processed in a top down fashion) to presence such as story and plots, flow, attention and focus, identification/empathy with the characters, social interaction, emotion, pre-knowledge, etc. [7][10][11]. One can argue that there is an (evolving) dichotomy within the concept of presence as illustrated in Table 1 (the table should be taken as an illustration, that is, in reality, the separation is not as clear cut).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2.2 Form Cues: “Where” and “What”</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1 Although these are important contributors to, for instance, conceptual presence, they may contribute to spatial presence depending on the target of the cognitive activity.
The visual system in our brain can be divided into two or more separate pathways. This separation starts to become evident at the level of the retina where two major types of cells, the M and P cells, can be found from which subsequent pathways (into Magnocellular and Parvocellular layers) are formed already within the primary visual cortex (also known as V1) [4]. Each of the pathways splits further into different regions in the visual cortex that have different functionalities. Regions along the P pathway seem to deal primarily with color, object shape and ultimately lead to the inferior temporal cortex which is known to mediate pattern and object recognition (“what” path). Regions along the M pathway are sensitive to orientation, movement and retinal disparity, and lead to the posterior parietal cortex that processes spatial and motion information (“where” path).

Figure 1: The what-where pathways in the visual cortex. The M carries information regarding space and motion for determining object location and the P primarily carries information regarding object local properties for object identification.

In our first experiment, we consider six visual cues: stereoscopy, object (gross) motion, user motion, motion detail, texture quality, and shape detail. The first three are considered “where” cues and the rest “what” cues. Even though we use the words “where” and “what” for these cues from their basic characteristics, it should not be confused that “what” cues can still affect overall spatial perception and vice versa. For instance, in object recognition, researches have found that humans focus on those parts of the scene that are most informative in disambiguating its identity [6]. A similar model has been established for attention as well [6].

2.3 Cross Modal Integration / Resolution

It is generally accepted that multi-sensory feedback is beneficial to both presence and task performance in the context of virtual reality systems [7]. This is only provided that the feedback from each modality is consistent with one another, and the multi-sensory feedback (or input) is configured appropriately for the task at hand [8]. Multimodal sensory mismatch often results in the form of sicknesses, discomfort and other after effects [15]. In relation to the sicknesses, humans are also known to adapt given sufficient amount of exposure. In the process of adaptation, humans resolve the mismatch by constructing one’s own interpretation of the situation, whether by suppressing one modality or fusing them together in some way [12].

3. Experiment I: Where vs. What

Our investigation started with an experiment to weigh the relative contributions toward spatial presence among different types of form factors, “what” and “where,” first with the visual input only, then with both visual and aural.

3.1 Testbed Environment and Independent Variables

In this first experiment, spatial presence levels were measured (with a subjective questionnaire) after having the subjects experience test virtual worlds configured with different combinations of six visual presence elements. We built a simple virtual undersea world as the testbed for the experiment. Table 2 shows the summary of the independent variables and their level design. Figure 2 shows an example of the virtual undersea world presented to the subject during the experiment.

Table 2: Five independent variables and their levels in Experiment I-1 (Uni-modal Case).

<table>
<thead>
<tr>
<th>Type</th>
<th>Variable</th>
<th>Levels</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Where</td>
<td>Stereoscopy</td>
<td>High</td>
<td>With stereo</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>No stereo</td>
<td></td>
</tr>
<tr>
<td></td>
<td>User Motion</td>
<td>High</td>
<td>Fixed user navigation</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>View at fixed location</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Object Motion</td>
<td>High</td>
<td>Fish moves around</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>Fish stays in place</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Motion Detail</td>
<td>High</td>
<td>With deformation (tail wagging)</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>No deformation (No tail wagging)</td>
<td></td>
</tr>
<tr>
<td>What</td>
<td>Geometry</td>
<td>High</td>
<td>High polygon model</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>Low polygon model</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Texture</td>
<td>High</td>
<td>With texture</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>No texture</td>
<td></td>
</tr>
</tbody>
</table>

3.2 Experimental Procedure

Subjects, in a random order, looked at each of the 32 virtual undersea worlds projected on a 50 inch screen from a fixed location for 90 seconds. After looking at each configuration, the subject was asked to fill out a presence questionnaire. The questionnaire comprised of four questions asking to rate the (1) visual realism of the objects, (2) one’s ability to perceive locations of oneself and other objects, (2) the visual realism of the overall

---

2 Our choice of visual elements comprises of those that can be varied by software control. For instance, the effect of field of view was not considered.

3 Although there are 64 combinations of the independent variables, the Fractional Factorial experiment design allows analysis by testing only 32 subject groups.
environment, and (4) the feeling of being in the environment (spatial presence), in the scale from 0 to 100.

Figure 2: An example configuration of the virtual undersea world (Geometry=high, Texture=high, Stereoscopy or Motion can not be illustrated here).

3.3 Main Results

The ANOVA, simple effect tests, and regression analysis showed that the manner in which the visual elements played a role was significantly different for user perception of visual realism and spatial presence. Results showed that “where” cues played an increasingly more important role (with statistical significance) for spatial presence than for visual realism only, but with a marginal difference (See Tables 3). This was rather contrary to our expectation; we expected a landslide victory for the “where” cues, because intuitively spatial cues should be more important for spatial presence. However, we observed the interactions between several “what” and “where” cues played significant roles in creating spatial presence.

Table 3: Regression analysis. Relative weights toward the overall visual realism ($R^2 = 0.97$) and spatial presence of the environment ($R^2 = 0.95$). All results with p value less than 0.006.

<table>
<thead>
<tr>
<th>Factor Types</th>
<th>Variable</th>
<th>Visual Realism</th>
<th>Spatial Presence</th>
</tr>
</thead>
<tbody>
<tr>
<td>What</td>
<td>Geometry</td>
<td>10.18</td>
<td>37.4</td>
</tr>
<tr>
<td></td>
<td>Texture</td>
<td>27.32</td>
<td></td>
</tr>
<tr>
<td>Where (29.87)</td>
<td>Stereoscopy</td>
<td>6.72</td>
<td>29.9</td>
</tr>
<tr>
<td></td>
<td>Object Motion</td>
<td>16.30</td>
<td></td>
</tr>
<tr>
<td></td>
<td>User Motion</td>
<td>6.92</td>
<td></td>
</tr>
<tr>
<td>Interaction</td>
<td>Geometry x</td>
<td>6.70</td>
<td>23.0</td>
</tr>
<tr>
<td>(23.07)</td>
<td>Object Mot.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Texture x</td>
<td>10.70</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Object Mot.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Texture x User Motion</td>
<td>5.70</td>
<td>7.23</td>
</tr>
</tbody>
</table>

With such a marginal difference in the degrees of contribution between the “where” and “what”, the results would not be so useful (e.g. guidelines for system configuration for spatial presence), because the sufficiency or saturation point for the contribution of the geometric detail and texture resolution (“what”) is neither known nor easily quantifiable, and perhaps depends on the user’s individual background. This in turn also make any analysis of the contributing weights among different cues, regardless of whether they belong to the “what” or “where” group, admittedly without merit for the same reason.

3.4 Bi-modal Case: Visual and Aural

Thus, a similar, but separate experiment was conducted to assess the effect of multimodality. A different testbed, an office navigation, was used and aural “where” (3D sound) and “what” (sound quality) cues were used in addition to the visual “where” (stereo, landmarks) and “what” cues (geometric detail and texture resolution) as the control factors. Our expectation was that with multimodality itself as a presence enhancing cue, the dependence of spatial presence on the “what” cue (as manifested in the first experiment) would be reduced. If we were to observe such a phenomenon, we would be able to expect other similar manipulations, such as adding interaction or other channels of sensory input to bring similar effects. However, contrary results were obtained, that is, the relative weights for the “what” cues dominated those for the “where” by about the ratio of 7 to 3. The “where” cues increasingly more important for spatial presence over visual realism was observed as was in the first experiment. As suspected with the results of the first experiment, the individual difference in detail perception (“what”) seems to make the comparison between “where” and “what” inconclusive.

4. Experiment II: Where and What vs. Time

In order to neutralize the individual difference in the contribution of “what” (and even for “where”) cues, the same experiment was run with another independent variable, the length of exposure time. Our hypothesis was that with a sufficient time of exposure, both the effects of the “what” and “where” cues would be saturated, and after that point, the true comparison between the “what” and “where” cues could be made.

4.1 Experimental Design and Procedure

The same testbed (i.e. undersea world) used in the first experiment was used again, but only four configurations were compared. The four test configurations were selected to form the “high what / high where,” “high what / low where,” “low what / high where,” and “low what / low where” groups. The “where” factors constituted the use of stereoscopy, gross object motion and shadow. The “what” factors constituted the levels of...
polygon counts and texture resolution for the object models.

The level of user felt presence was measured in a different way. All the subjects (32 in total) experienced the various configurations (in a random order) for 10 minutes. The level of presence was measured basically the same way by asking the subject to rate “the degree to which feeling being in the undersea world,” but at every one minute in the scale of 0 to 100 during the course of the 10 min. experiment. No visible symptoms of simulation sicknesses were observed from the subjects.

4.2 Main Results

The main result is shown in Figure 3. The level of presence generally increased for all four test combinations with longer time of exposure. The graph in Figure 4 also clearly shows that the level of presence saturating, and near and at the saturation regions, it is the “where” cues that expedites and becomes more important in promoting spatial presence. The relative weights in the early part of the exposure are irrelevant because it is presumed that their effects have not kicked in sufficiently, and individual differences create a large variance.

Figure 4 shows the change in relative weights toward spatial presence among the variables: time (labeled “intercept”, diamond), where (triangle), what (square) and the interaction (“x”). With longer exposure time, the contribution of the “where” cues becomes the most important. It is also interesting to observe that at the end of 10 minutes, there is no interaction between the “where” and “what” cues. This is contrary to the results of the Experiment I where much interaction was observed. The interaction in Experiment I could have been caused by the multi-sensory conflict due to the lack of sufficient exposure (only 90 seconds).

5. Experiment III: Form vs. Content

While Experiment I and II concerned the relationship between the form factors, “where” or “what”, the purpose of Experiment III was to study the relationship between two elements, each representing the two axis of the presence dichotomy, perceptual cues for spatial perception and sustained attention for (psychological) immersion. Our belief was that spatial perception and a top down processed concept such as voluntary attention have only a very weak relationship. In our experiment, subjects navigated through a virtual office with three differing levels (low, medium, high) of visual perceptual cues. The subjects were asked to either to count certain objects or not in the midst of the navigation. Our hypothesis was that sustained attention would have increasingly positive effects toward spatial presence for low fidelity virtual environments (impoverished spatial/perceptual cues), and have no effect in the high fidelity environment (rich in perceptual cues). Thus, we expected the effect of sustained attention would saturate as the environment became richer with spatial cues and its perceptual realism. In order to confirm the sustained attention actually occurred while carrying out the counting task, fMRI of the subjects were taken and analyzed.

![Figure 3: The level of presence along time of exposure among four tested configurations of Experiment II.](image)

![Figure 4: The relative weights toward spatial presence among the variables: time (labeled “intercept”, diamond), where (triangle), what (square) and the interaction (“x”). With longer exposure time, the contribution of the “where” cues becomes the most important.](image)

5.1 Experimental Design

Our experiment was designed as a 3×2 between-subjects experiment. There were two independent variables. One was the visual detail of the virtual environments (bottom up cues) and the other was the (sustained) attention factor (top down cue). The dependent variables were the total score of the presence questionnaire that subjectively rated the degrees of feeling of being in the virtual environments (i.e. spatial presence). The virtual environments consisted of the three different levels of visual detail: synthetic and low in detail (L), synthetic and high in detail (H), and real video (V). The attention...
factor had two levels: with the attentive task (TO), and without it (TX).

Several kinds of visual detail cues were manipulated to create the low and high fidelity versions of the synthetic environment. Those were the geometric detail (polygon counts for objects), inclusion of shadow, object motion, and texture resolution. Due to the counting task, user motion was set to passive navigation in a fixed path. The display was provided in monoscopy as the special-purpose IMRI compatible HMD (Head Mounted Display) did not support stereoscopy.

The subjects were instructed to count the number of pencils with special colors in the synthetic environments or in video environment while navigating. The colors of the pencil body or the cap could be one of four: red, green, blue or white. The colors of the pencil and the cap were mixed in a random order. The subjects were asked to count the one with red body and blue cap.

5.2 Experimental Procedure

Group I (12 subjects) experienced virtual environments in the IMRI system, and group II (other 24 subjects) experienced virtual environments without it. The boxcar design was used. Given a test environment with a visual detail level (L, H, or V), the subjects went through a series of tasks, FIX, TX, and TO, three times. FIX means a fixation task representing the resting baseline for comparison with activated state. At first, the scanning triggered the presentation of a crosshair (fixation baseline) for 12 seconds prior to the first task block. This fixation was followed by a block of 30 seconds blocks of TX and TO. This process was repeated 3 times for each L, H and V. The sequence of L, H and V was pseudo-randomly chosen. Thus for example, the first step might be FIX-HTX-HTX-HTX-HTO, the second, FIX-LTX-LTO-FIX-LTX-LTO, and the third, FIX-VTX-VTO-FIX-VTX-VTO.

After finishing each step (e.g. FIX-HTX-HTX-HTX-HTO), subjects filled out the presence questionnaire. Subjects from the group I were instructed not to move their heads to insure head fixation. For this reason, they answered to the questionnaire with voice with minimal exchange of words. Subjects from group II plainly wrote their answers to the printed questionnaire.

We used ten questions to rate the degree of feeling of being in the virtual office and other related qualities of the virtual experience. Our questionnaire largely considered spatial presence. Each question was answered in the scale of 0 to 10.

5.3 Main Results

The results of ANOVA are shown in Figure 5. The first figure of Figure 5 represents the results for the total presence score. It shows that the means of the presence scores for each level in the visual detail factor (L, H and V) were significantly different (α= 0.05, Pr < 0.0001). According to the SNK (Student-Neuman-Keuls) Test, the score for V was the highest, H the middle, and L, the lowest (as expected). On the other hand, the difference in presence scores between TO and TX were not statistically significant (α= 0.05, Pr = 0.1225). The analysis also showed no significant interaction between visual detail factor and attention factor. (α= 0.05, Pr = 0.4319). This result partially supports our hypothesis that spatial presence and attention have a weak relationship. In fact, the result shows they are independent and unrelated.

The brain activations analysis using SPM showed significant differences in the brain pattern only between TO and TX. Figure 6 shows the brain images rendered into the standard single subject image. It shows that the cingulate, inferior parietal, inferior frontal, middle frontal and sub-gyralf regions were particularly activated. These activated regions are evidences of the sustained attention.

Our original hypothesis was that sustained attention would positively affect spatial presence in a virtual environment with impoverished perceptual cues, but would have little or no effect in an environment rich in them. The experimental results showed they were not related at all.

Waterworth et al. [17] suggested in their FLS (focus, locus and sensus) model that sense of (spatial or physical) presence is the strongest when attention is most occupied by perception of the environment (physical or electronic), and the weakest when attention is most occupied with mental reflection. They explained that changes in the balance between conceptual (abstract) reasoning and perceptual (concrete) processing affect the nature of our experience of the world around us. Their FLS model suggested that the subjective duration depends on the amount of conceptual processing performed during an interval, relative to the level at which an individual habitually performs. For example, if conceptual processing has a heavy load, people’s experience of duration is short and people pay little attention to the world around them. In those situations, they are “absent minded” and do not present in the world. And when the conceptual processing load is light, they have longer experience of duration and can frequently sample what is going on around them, whether natural or synthetic. In this sense, presence arises when people mostly attend to the currently present environment within and around the body.

Our result is consistent with that of Waterworth’s model and we claim that introducing high level elements like attention, emotion, scripts do not really help user build a spatial model of the place and leave the user with feeling visiting a concrete place.

Our results may also be explained by the fact that spatial presence or spatial perception is largely a low level perceptual phenomenon that goes on involuntarily, while conceptual presence is high level top down, and...
voluntary reasoning. Thus the only way they can be coupled is when the target of the conceptual reasoning is the physical (or virtual) world itself (e.g. thinking about where the desk is). Even though people cannot afford to pay attention to the surrounding environments during the attentive task, but they still know that they are already in the synthetic environments or real environments and continue to receive perceptual cues processed automatically. However, the high level cognitive activity may be inhibiting the spatial memory construction process of the perceptual system.

Interestingly, the debriefing session revealed a difference in spatial perception depending on the perceived difficulty of the task. Those who thought the counting task was easy showed tendency to feel increased presence by the inclusion of the task. This is another evidence of the reduced mental load on the conceptual processing leaving room for formation of higher spatial presence.

Figure 5: The effect of visual detail and attention on presence. There are significant differences among L, H and V (above), but none between TX and TO (below).

6. Summary and Discussion

6.1 Model of Spatial Presence

In some sense, it is obvious that spatial or “where” cues are important for spatial presence. Despite the common sense, it seems there is a big confusion due to many different definitions of presence and conflicting results from various studies that considered different types of cues as contributing factors to presence. However, very little of these studies considered the effects of saturation and time of exposure. Our position on how spatial presence is formed is that it is a product of basically a bottom up perceptual process that gathers spatial cues to actively place and register the user in the seemingly surrounding environment and that it takes some amount of time. Our results show that this process can be affected by provision of spatial perceptual cues “to set the stage” [3]. We further speculate that, the spatial cues must be consistent with the external stimuli to be effective. As for the content factors, we believe that they must be spatial in nature to create synergistic effect with the form factors. Thus, a spatial attentive task (e.g. search) with rich form factors would have created the highest possible presence. Our model is depicted in Figure 6. Slater has similarly recently speculated on the existence “minimal” perceptual cues that are sufficient to invoke high presence when coupled with top down reasoning that creates a personalized experience [14]. Our model further extends this idea.

Figure 6: Contributing factors to spatial presence.

6.2 Measurement of Spatial Presence

As spatial perception lies in the core of spatial presence, measurement of spatial presence can be carried out by testing various spatial memory and behavior. Spatial presence can be viewed as a type of spatial representation with particular characteristics, for instance, a sense of inclusion, appropriate size, and perspective (e.g. ego-center). In the Experiment II, in addition to the spatial presence, the size characteristics were asked of the subjects, and spatial presence (and its change according to time) correlated highly with an appropriate size perception (e.g. large enough to include the user), another evidence of a gradual registration of oneself into the environment. In addition, the level of presence must be linked to the concreteness of this representation strengthened by the amount of cues and time of exposure. Slater et al. has already used the level of concreteness of spatial cognition as part of his presence questionnaire [16].
6.3 VR System Design

The implication of the study is important for interactive multimedia or virtual reality system design. Employing expensive VR devices will be superfluous if the purpose of the system was non-spatial. On the other hand, VR as a technology will have a unique value in providing strong spatial context for those applications that require it such as many training and educational systems. For instance, if indeed it is possible to induce psychological immersion by manipulation of story, plots and abstract interaction, then, the digital contents such as the interactive story or games can be conveyed sufficiently using the conventional desktop interfaces rather than employing expensive and often difficult to use and engineer VR setups to create spatial contexts.

7. Conclusion

Establishing a model of presence is important because it serves as one of the basis for designing and evaluating virtual reality applications. A model of presence refers to a detailed analysis of the contributing elements and their mutual interactions. In this paper, we have argued for a model for spatial presence based on results from a series of experiments, manipulating various types of artificial cues. We believe that spatial presence is a product of an effort to correctly register oneself into the virtual environment and this process is perceptual, and bottom-up in nature, rooted in the reflexive behavior to react and resolve the mismatch in the spatial cues between the physical space where the user is and the virtual space where the user looks at, hears from and interacts with. In particular, we postulate that while low level and perceptual spatial cues are sufficient for creating spatial presence, they require sufficient amount of time to take effect, and can be affected and modulated by the spatial (whether form or content) factors. More studies are needed and planned to further verify our proposal.

Acknowledgements

This work was supported in part by the BK 21 program of the Korean Ministry of Education, the ITRC program of the Korean Ministry of Information, and the POSTECH POSRIP.

References


Presence and Sexuality
Matthew Lombard, Matthew T. Jones
Temple University
{lombard@temple.edu, matjones@temple.edu}

Abstract
This paper considers the relationship between sexually arousing media content and specific media technologies from the perspective of presence theory.

"I'm new to this area, but why isn't anybody in this group studying presence and sex?"

1. Introduction

Presence is a naturally interdisciplinary topic and presence scholars have explored its application in a wide range of settings and contexts including art, space and undersea exploration, business, health and medicine, and many more. Given its popularity and prevalence, it is somewhat surprising, however, that presence researchers and theorists have not explored presence in the context of sexual media content and the technologies that deliver it.

After defining the relevant key terms, this paper outlines reasons presence scholars should examine presence concepts and theories in the context of sexuality, considers the nature of presence responses in this context, and proposes a series of conclusions regarding the implications of presence developments in this area.

2. Defining terms

To examine presence and sexuality, it is first necessary to define the terms presence, pornography, erotica and sexual media content. Presence is defined following Lombard and Ditton as “the perceptual illusion of nonmediation” [1]. This refers to a phenomenon in which an individual perceives a mediated experience as an authentic first-hand experience to which he/she may respond physiologically, cognitively and emotionally as he/she would in the nonmediated setting. A variety of dimensions of presence have been proposed, some of which are discussed below.

Following Mosher [2], pornography is defined here as media content for which the exclusive goal is to elicit or enhance the sexual arousal of its users. This content is typically although not always sexually explicit.

A related term used to describe sexual media content is erotica. Linz and Malamuth write that erotica “is often used to refer to literary or artistic works that have a sexual quality or theme” [3]. Because some of the actual content of erotica is similar to pornography, distinguishing between the two is a difficult and value-laden endeavor. Erotica can be defined as sexual media content for which a primary (not necessarily the exclusive) goal is to elicit or enhance the sexual arousal of its users.

Sexual media content describes any content presented or experienced via a technology and which is designed in part to elicit or enhance sexual arousal of users. This includes not only pornography and erotica but content presented in swimsuit issues of sports magazines, Victoria Secret television specials, beer commercials featuring attractive and scantily-clad women and men, etc.

Finally, the technology (media format) that carries sexual media content can logically include not only media such as videotape, the Internet and DVDs but adult novelty items intended to be used sexually.

3. Reasons to examine presence and sexuality

Before exploring the relationships between sexual media content and technologies and presence, we turn to some reasons this topic deserves scholarly attention.

First, creators and distributors of sexual media content don't use the term, but they clearly seek to produce an experience of presence, a sense of being physically, and in some cases emotionally, close to the people represented in the content. Their efforts have been closely tied to the evolution of media technology and in many cases have driven the development, innovative use, and profitability of such technology. Tierney notes that "virtually always, from Stone Age sculpture to computer bulletin boards, [the erotic] has been one of the first uses for a new medium” [4]. Steinberg [5] comments that, "[E]very new technological achievement quickly finds its way, like water flowing downhill, to a sexual application. When the photographic process was first discovered, one of its first uses was to create enticing images of naked women. When motion pictures were born, underground sex films immediately followed. One of the prime economic foundations of the home video revolution has been the sex video market." More recently, sexual content appeared early on the worldwide web [6], where "[o]nline pornographers have been among the first to exploit new technology for more than a decade, from video-streaming and fee-based subscriptions to pop-up ads and electronic billing” [7]. It's played a key role in high definition [8] and interactive television [9; 10], video-on-demand [8], and now mobile communication [10] technologies. "Though some analysts don't like to talk about it, many new ways of delivering content first became profitable via pornography. As far back as the printing press ... industries have all grown technically on this type of content. And new technologies themselves are often successful because of it" [11]. Today
companies like the aptly named Vivid are investing in research and development to produce new technologies that will deliver sexual media content [12; 13]. Interactive, real-time, remote sex toys, holograms, and life size, anatomically correct (and soon, animated) sex dolls are among the many technologies being developed. Presence scholars need to better understand how the pursuit of presence has guided, and especially how it is likely to guide, the evolution of all of these media technologies and the consequences of this evolution.

Another reason that scholars should examine presence concepts and theories in the context of sexuality is that sexual media content is consumed on a daily basis by many millions of people around the world. Estimates of the size of the 'adult entertainment' industry vary from 1 to 14 billion USD annually [14; 15; 16]. Major media, hotel, and other, often multinational, corporations may not publicize it but make millions in profits [17; 18; 19]. Nearly a third (32.3%) of men and a fifth (17.9%) of women report having seen an x-rated movie in the last year [20]; over half (54%) of Americans report having had sex via phone, e-mail or text message [21]; the most frequently searched-for word on the Internet is "sex" [22]; and some estimate that "over 60 percent of all visits on the Internet involve a sexual text message" [23]. Pornography has gone mainstream all over America. From movies to television shows to music videos and magazines, porn stars and porn iconography are everywhere, pointing to a national comfort level that few of the many contexts in which presence concepts have been studied, a large, diverse and global public already is regularly consuming sexual mediated content and experiencing presence in the process.

The prevalence and popularity of sexual media content and evolving technologies to deliver it are likely related to the important role sexuality plays in our lives. Along with air, food, water, sleep, and warmth, sex is classified as a biological or physiological human need, the base level in Maslow's "hierarchy of needs" [25; 26; 27]. Beyond the biological imperative to reproduce, humans need and highly value emotional intimacy and connection with sexual partners (part of the Belongingness and Love needs in Maslow's hierarchy). Whether evolutionarily or culturally based, it is clear that sex and intimacy are critically important to most human beings. How and with what success they seek fulfillment of these needs via sexual media content and technology should be of great interest to presence scholars.

Pornography has long been the subject of controversy, with some arguing it is harmful in a variety of ways and others arguing for its benefits [28; 29]. Pornography, especially violent pornography, has been charged with causing people to objectify and disrespect women, to become callous and desensitized to their mistreatment, to become sexually promiscuous, and to behave violently toward women [30; 31; 32; 33; 34].

Others argue pornography and other sexual media content contributes to important educational and therapeutic [35], hygienic [36; 5] and safety [36] functions. One leading manufacturer of internet compatible sex toys (SafeSexPlus.com) advertises this final advantage in its name. As an advertisement for "Ultra-Realistics" sex toys proclaims, the user gets "all the pleasure and all the excitement with none of the risk!" [37]. All of these potential negative and positive consequences of mediated sexual content and experiences provide a rich and important context for the application of presence research and theory.

The evolution of presence-enhancing technology regarding sexuality also raises critical ethical issues. Commenting on the $6500, life-size, anatomically correct RealDoll, attorney and feminist M.C. Sungaila writes, "Knowing that it's out there and that somebody thought this was a good idea -- to make money off the complete objectification of women -- is discomforting to say the least" [38].

Rheingold notes that "Given the rate of development of VR technologies, we don't have a great deal of time to tackle questions of morality, privacy, personal identity, and even the prospect of a fundamental change in human nature. When the VR revolution really gets rolling, we are likely to be too busy turning into whatever we are turning into to analyze or debate the consequences" [39].

For these reasons and others, and despite the potentially controversial and politically sensitive nature of the topic, scholars should turn their attention to the context of presence and sexuality.

4. Characteristics of presence in the context of media with sexual content

Presence is a multi-dimensional concept, and as noted above, a variety of dimensional structures of presence have been proposed [40; 41]. Here we use the relatively detailed dimensions identified by Lombard and Ditton [1] based on a review of presence literatures to consider how each dimension applies in the context of the intersection between media form and sexual content.

4.1. Social richness

Informed by social presence theory [42] and media richness theory [43], Lombard and Ditton [1] identify the first dimension of presence as social richness. As a characteristic of a medium social richness generally refers to the use of multiple sensory channels; as a characteristic of the medium user, it refers to the subjective experience of warmth and intimacy in the mediated interaction. These qualities are not only contingent upon one another (increased sensory channels = increased intimacy), but are obviously desirable in the mediated sexual encounter, where a high degree of physical and/or emotional intimacy would be expected. As we will see, media content and technology designed to provoke sexual arousal through the experience of presence seeks to maximize the level of social richness in the encounter through both increasing the number of senses involved and increasing the bandwidth of sensory information.
4.2. Realism

The second dimension Lombard and Ditton [1] identify is realism, and they describe two distinct forms, perceptual realism and social realism.

4.2.1. Perceptual Realism Perceptual realism refers to a presence experience in which the mediated representation accurately simulates or reproduces the sensory experience that would be expected in the nonmediated context. This is an essential aspect of the mediated sexual experience since human sexual arousal has evolved in response to real (unmediated) stimuli and not mediated stimuli. Despite this evolution, however, mediated content designed to provoke sexual arousal (e.g. pornography) can do so by maximizing perceptual realism. Malamuth observes that “since mass media did not exist in our ancestral history, our mechanisms for discriminating fantasy versus reality may not be sufficiently sharp to totally avoid any long-term impact of exposure on our feelings, thoughts, and behavior” [44]. Anderson [45] and Reeves and Nass [46] similarly point out the deceptive effect that media have on human perception. Extending this logic to include pornography, we can assert that the reason mediated images of bodies are arousing is that, on some primitive level, we respond to mediated bodies as if they were real bodies. It would make sense to further suggest that the more perceptually realistic sexually stimulating images seem, the more likely they are to evoke sexual arousal.

Efforts by manufacturers of adult novelty toys have placed a premium on developing products high in perceptual realism. Companies such as TopCo and Doc Johnson have developed and produced artificial genitals out of various materials designed to reproduce the qualities of real skin. Patented compositions with names like Cyberskin™, Futurotic™ skin, Realistic® skin, and Ultra-Realistic 3.0® skin not only demonstrate an attempt to maximize perceptual realism, but also provide evidence that the technology is steadily progressing through stages of development. Toward this same goal of maximizing perceptual realism, Playboy has announced the introduction of “Spice HD,” a high definition television channel dedicated to sexually arousing content.

4.2.2. Social Realism As opposed to perceptual realism, social realism describes a presence experience wherein the behavior and language of depicted social actors are true to life or realistic in nature. The proliferation of “amateur” and “reality” pornography attests to the importance of an unscripted, unprofessional product that permits access to socially real people. Despite the lack of perceptual realism that exists in amateur pornography as a result of low quality recording equipment and lack of professional skills, the content remains exciting as a result of its authenticity.

4.3. Transportation

Beyond realism, or perhaps a precondition of it, is the issue of physical location. In a chapter which discusses the problem of physical location in the context of erotic internet interaction, Waldby notes that “The pretext for any computer mediated communication between participants is separation in space” [47]. This, of course, can be extended to include any technologically mediated communication (computer or otherwise). As a result of this, the experience of presence in the mediated situation is contingent upon the perceived transportation of users. Lombard and Ditton [1] describe three ways that location is perceptually altered by presence as transportation: “you are there,” “it is here” and “we are together.”

4.3.1. "You are There" “You are there” transportation describes the user ‘traveling’ into the mediated environment and feeling as if he/she were a part of the mediated world. Virtually any pornographic film or video aims to bring the user "into the action." Weaver [29] observes that contemporary pornography presents sexual content "in a 'you are there as it happens' documentary style."

4.3.2. "It is Here" A second form of presence as transportation, termed “it is here”, brings the mediated representation into the space of the media user rather than the other way around. There exist some sophisticated and highly realistic sex toys fashioned from casts/molds of the genitalia of porn performers that would seem to have the potential to transport the anatomical likeness of the particular performer into the user’s space [48; 49; 50]. Advertisements for these products emphasize the connection that they have to the flesh and blood porn performer from whom they were molded. To illustrate, an ad featuring the “Juli Ashton Ultra Realistic® Vibrating P*** with Anus” boasts that it is “molded directly from Juli’s vagina” [49]. These replicas function to bring the porn performer (at least in fragments) to the porn consumer. Beyond the reproduction of mere fragments of performers’ bodies is the reproduction of the entire body in the form of a “love doll.” An advertisement for the “Jill Kelly Suction Doll” announces “You’ve seen her on the screen, now see her between your sheets” [50]. And beyond the tangible replicas in these examples, adult entertainment company Digital Playground is currently working to develop holographic imagery intended to sexually arouse. In addition to again demonstrating how pornography is a driving force behind technological innovation of media, this shows that resources are being allocated toward developing technology capable of transporting mediated people to the viewer.

4.3.3. "We are together" A third and final form of presence as transportation specifies a shared space in which a sensation of “we are together” is experienced by mediated communicators.

A number of products and services with adult content ranging from the very basic to the very intricate have been designed to evoke a sense of interacting together in the same space between people who are actually in different physical locations. What is crucial to this sense of shared space is the real-time (or apparently real-time) nature of the interaction.
In a 1995 article for Wired magazine titled “The next best thing to being there,” Robert Rossney attests to the importance of interaction and feedback within a mediated erotic encounter. Through an investigation of an online “peep show” service called “Virtual Connections, Ltd.,” Rossney discovered that he was aroused based on feedback from the woman with whom he was interacting. He writes that “It’s one thing to look at a picture of a scantily clad woman. It’s another thing entirely to ask her to remove an article of clothing and see her respond by whipping off her panties and flinging them aside” [51]. Clearly the interactive nature of this sort of experience goes a long way toward reconstituting some of the lost sense of (“we are together”) transported presence in the mediated erotic encounter.

Presumably, the ability to physically “touch” a performer (even if it is through computer interface) and observe her response is more arousing than only visual, aural or text-based real-time communication. This technology is available and being refined as well [52]. Several sex toy manufacturers have used simple remote control technology to permit genital manipulation at a distance. Products such as the “Remote Controlled Vibrating Panty” [50], the “Remote C*** Blaster” [49] and the “Remote Control Egg” [37] are designed to permit a person holding the remote control to covertly stimulate the woman who is wearing the device that is receiving the signal. This technology illustrates the central role of real-time interaction in the mediated erotic encounter.

What is lacking in remote controlled sexual devices is the sense of reciprocity that is so central to the unmediated sex act. A remote control only transmits signals, it cannot receive signals. In her discussion of online text-based erotic encounters, Waldby makes the observation that the technology used to interact “both substitutes for the face-to-face negotiation of proximate sexuality and simulates certain aspects of that proximate relationship, involving the projection of a limited kind of telepresence through the simultaneous and interactive production of pleasure in the other’s body” [47]. Although Waldby is referring exclusively to text-based sexual encounters (referring to more elaborate forms as “cumbersome and literal minded”), the principal importance of reciprocal communication is well illustrated.

Toward the goal of constituting a sense of reciprocal interaction in the shared space of the mediated erotic encounter, some innovators have devised technology capable of allowing the real-time give and take integral to fostering the intimacy of the “we are together” form of presence as transportation. The remote sex technology offered by “F*** You, F*** Me” (www.fu-fme.com) permits users to interact sexually with Windows compatible “genital drives” that act as surrogates for their partner’s sexual organs. Still more complete and elaborate is Dominic Choy’s computer interfacing sex doll which, “[u]sing signals from the internet as well as sound and touch sensors…would allow a user wearing a virtual reality headset to have virtual sex with someone in another part of the world...” [53].

Likely the most sophisticated and comprehensive solution to the problem of creating shared space in the mediated erotic encounter is theorized by Howard Rhiengold [39] who describes a scenario in which long distance sexual partners stimulate each other through bodysuits that are enmeshed with sensors.

Aside from technologies designed to bring two (or more) people simultaneously into real time, shared space, mediated encounters, interactive DVDs attempt to create a sense of “we are together” transportation between a live user and a recorded performer. “Joone,” president of Digital Playground – a company that manufactures interactive adult DVDs, has said that “When you’re watching a regular porn movie, you’re watching it in the third person...This way [with interactive DVDs] it’s a first-person experience” [54]. Through the use of a DVD remote control, users are able to interact with performers, choosing sexual positions and other aspects of the para-sexual experience. Many current DVD series feature interactivity with popular adult performers. Titles include “My Plaything,” “Virtual Sex,” “My Digi-Girl,” and “Come Play with Me” [48].

4.4. Immersion

Another form of presence – presence as immersion – occurs in two varieties: psychological and perceptual.

4.4.1. Psychological Immersion Drawing from Palmer [55] and Quarrick [56], Lombard and Ditton define psychological immersion as a feeling of being “involved, absorbed, engaged, [and] engrossed” [1]. Cybersex addiction provides a good example of how psychological immersion functions. Millions of Americans are “cybersex” addicts [57; 58; 36], and while much of their media use involves highly iconic (rather than perceptually rich or immersive) communication via chat rooms and e-mail, they become so deeply involved in the experience that they consider it the equivalent of an affair (Schneider, 2000).

4.4.2. Perceptual Immersion Perceptual immersion refers to the involvement of multiple sensory channels in the mediated encounter. If one can only see a stimulus, the experience is less immersive than if one can see, hear, touch, taste and smell a stimulus because the involvement of multiple sensory channels permits cross-validation of experience. As a result of the fact that the natural (unmediated) sex act is sensually immersive in that participants are necessarily confronted with the sight, sound, feel, smell and taste of each other’s bodies, it makes logical sense that the producers of media offering sexual content seek to approximate these sensations. This explains why pornography, or any sexual media content, has been an innovative and driving force behind developing perceptually immersive presence experiences. For example, Digital Playground’s homepage (www.digitalplayground.com) states explicitly that “Digital Playground’s ultimate innovations will combine all the senses, bringing your fantasies one step closer to reality.”

One of the key challenges for innovators attempting to create immersive experiences (sexual or otherwise) is the involvement and coordination of multiple sensory channels.
The Spice TV catalog [48] offers “Virtual Sex Sets” which include both an interactive DVD of a particular porn performer and a sex toy reproduction of the genitals of that performer. In offering visual, aural and haptic dimensions of experience, these “Virtual Sex Sets” take an initial step toward creating a perceptually immersive experience.

Beyond mere involvement of multiple senses, the coordination of those senses is crucial to fostering a perceptually immersive presence experience. If one sensation provides information that contradicts another sensation, the presence experience will be lost. Eric White’s “Virtual Sex Machine” is a device which seeks to coordinate haptic experiences with visual ones. Argento reports that “The device, powered by three separate motors, receives the signal from the computer and then reproduces sensations of what’s happening on the screen. As White says, what happens on the screen, happens to you” [59]. Placing the agency of the immersive experience in the hands of the user, Immersion Corporation has developed a computer mouse capable of permitting the user to “grope” by transmitting texture and feeling to the hand of the user as he or she glides the cursor over the surfaces of the onscreen image. This subjective agency capitalizes upon Heeter’s observation that “A sense of presence in a virtual world derives from feeling like you exist within but as a separate entity from a virtual world that also exists” [60]. In both of these examples, visual and haptic sensations are combined and coordinated to induce a multifaceted feeling of perceptual immersion.

4.5. Social Actor Within Medium

“Presence as social actor within medium” [1] or “parasocial interaction” [61] describes an interaction that a media user seems to have with the mediated performer that is akin to a face-to-face interpersonal interaction. Because of its ability to sustain a level of feigned intimacy through the gestures of private interaction, this parasocial interaction is ripe to be exploited by media producers seeking to create a sexually arousing experience for viewers. Even as early as 1956, when Horton and Wohl first theorized the parasocial encounter, examples of its sexual applications came into play with the discussion of the popular 1950’s radio program The Lonesome Gal, which featured a seductive feminine voice speaking in first person to an audience of single men at the end of the day.

Pornographic photography frequently features models who make eye contact with the viewer (a technique known as direct address). Audio-visual media allow visual and verbal direct address to further encourage parasocial interaction. One advertisement by Video Marketing Concepts informs us that, “This wholesome girl-next-door talks dirty to you as she unveils her naturally curvaceous body.” Additional selling points, such as “close-ups 3 inches from your face!” directly address the interaction between viewer and performer as if the barriers of time, distance and the screen did not even exist.

Other examples of sexualized parasocial interaction do not even involve flesh and blood performers. Highly realistic two and three dimensional computer generated images manufactured with programs such as Adobe Poser©, Kinetix 3D Studio Max©, Alias Wavefront© and Maya© allow artists to render incredibly lifelike images of bodies [62]. A “Miss Digital World” pageant has even been held to exhibit these deceptively attractive graphic creations [63].

Some innovative adult software carries things a step further by bringing virtual images to life and allowing for computer mediated interaction with a digitally created partner. The Girlfriend™ program has the potential to permit the user to cultivate a relationship with a virtual woman. An advertisement notes that “Now you can have your own girlfriend…a sensuous woman living in your computer!” and goes on to explain that “your girlfriend starts with a vocabulary of over 3000 words and will continually learn new words, feelings, and ideas. This program truly grows the more you use it” [64]. This and other sure to be more sophisticated programs in the future may be so interactive and adaptable, that they blur the line between parasocial interaction and true interpersonal interaction to the point that a malfunctioning hard drive may constitute the death of a lover and friend. Due to the additional dimensions of interactivity that sexuality provides, using programs such as Girlfriend™ would likely be a far more intense experience than interaction with nonsexual virtual companions like the “Tamagotchi” [65], the electronic pet that requires interaction from its owner to stay alive.

4.6. Medium as Social Actor

A final variation of the presence experience that should be addressed in terms of its role in mediated sexual content is “medium as social actor.” This form of presence involves social responses of media users not to entities (people or computer characters) within a medium, but to cues provided by the medium itself” [1]. Dildos, vibrators, masturbation sleeves, penis pumps, and various other sexual devices that are designed to give pleasure but do not rely upon interaction with other people (real or virtual) to operate could be said to constitute media that function as social actors. Current products such as Realdoll (mentioned above) and sexually capable androids of science fiction exemplify this concept. Here an issue concerning the role of internal fantasy comes into play. When a person makes use of a sexual device, they probably are not fantasizing about the device itself any more than a person masturbating is fantasizing about his/her hand. This being the case, it is difficult to attribute the arousal of the user solely to the device since internal fantasy may play a significant role in the arousal process (this may apply to some extent to any of the types of presence discussed and how they function with regard to sexual content).

5. Conclusions, implications and future research

This preliminary exploration of the topic of presence and sexuality leads us to a series of conclusions:
1. The topic area of presence and sexuality has received little to no attention from presence scholars. In addition to wanting to avoid political and other controversies, it seems likely that this is related to a common tendency in journalists’ reports about pornography to maintain “an arm’s-length disdain passing for objectivity,” in which sexual media content is “treated as a sociological phenomenon, just not one that is part of any culture that they -- or by implication any [news consumers] -- feel part of” [19]. Academics should not hesitate to acknowledge they are part of a culture and species for which sexuality is important and explore interesting, important and common phenomena related to presence in this context.

2. As with many contexts in which presence is relevant (e.g., art, business, education, politics, etc.), the people who create sexual media content and technologies to deliver it don’t use the term presence, but that’s exactly what they are pursuing. Perhaps more than in any other context, presence is their primary goal – it’s the purpose of every sexual product from still images to realistic, interactive dolls. And as with the other contexts, these creators have contributed to a fascinating evolution of technologies in pursuit of this goal.

3. Perhaps unlike many contexts which presence scholars explore, products in the context of sexuality are designed and sought by consumers to create every type of presence. Regardless of the conceptualization scheme, every dimension of presence – from “realism” to “medium as social actor” [1], “spatial” and “social” [41], and certainly “spatial,” ”engagement” and ”naturalness” [40] – is central to the experiences desired by users and therefore pursued by content/technology creators in this area.

4. Sex is one of the most complex, sensory-involving aspects of human experience, so it represents the ultimate challenge to content and technology producers who are trying to create an actual or perceived mediated encounter that reproduces the experience of physical intimacy. Not only must they reproduce the key physical and emotional aspects of the experience, their products must be affordable and available for use in the home (as opposed to large, very expensive and public IMAX theaters, simulator rides, etc.). It seems likely that the number and consistency of sensory inputs and outputs are critical to their success [1] along with many other factors identified by presence researchers in other contexts. Current products may be lacking, but technologies will likely evolve quickly: “Cyberdildonics is still far more like having sex with an electronic device than with a human ... [but] ”these products are merely version 1.0” [13]; “Given the pace of technological innovation, sexual experiences straight out of science-fiction novels may not be very far away” [66].

5. In addition to the longstanding ethical issues and questions about positive and negative effects related to pornography and other sexual media content, the evolution of technology in the context of presence and sexuality is likely to bring new compelling and important ethical questions to the fore, perhaps in the not-too-distant future. These include the (im)morality of having sex with androids, of replacing nonmediated sex and relationships with idealized virtual versions, and the implications of these for our psychological health (e.g., our ability to distinguish acceptable virtual and unacceptable ‘real’ behavior), the social cohesiveness of society, and eventually even changes in the nature of being human [39].

6. Because of the ethical issues that it raises and for all of the reasons discussed in section 3 above, presence researchers should monitor the industries and technologies that produce and deliver sexual media content and begin to study presence phenomena in the context of sexuality. The specific benefits of this work include new tools to measure presence and its correlates, the ability to consider traditional presence issues and questions in a fresh context, and the potential to contribute to positive societal changes. The new tools include physiological measures of genital responses, primarily penile tumescence [67; 68; 69] and vaginal pulse amplitude [70], that unlike other objective physiological measures such as GSR, can be attributed to very little other than a specific source (here sexual arousal) that is logically highly correlated with presence. Exploring presence and sexuality would provide new data and perhaps understanding of longstanding issues in the presence community such as “the book problem” (whether and how highly symbolic media such as text can evoke presence), the appropriate trade-offs in different form and content characteristics necessary to efficiently evoke presence, and the ethics of accurately reproducing nonmediated experience. And we have the opportunity to contribute to the long debate about the many potential harmful and beneficial consequences of consuming sexual media content, possibly helping to minimize the former and maximize the latter.

References


[48] Spice TV catalog: for men and women who want it all (Summer Preview, 2004)


[64] Sexxy Software (2004). [Catalog] P. O. Box 220, Hatfield, PA 19440


Presence and the Aura of Meaningful Places

Blair MacIntyre1, Jay David Bolter2, Maribeth Gandy3
1College of Computing 2School of Literature, 3Interactive Media
Communication and Culture Technology Center
GVU Center, Georgia Institute of Technology, Atlanta, GA, USA
{blair@cc.gatech.edu, jay.bolter@lcc.gatech.edu, maribeth.gandy@imtc.gatech.edu}

Abstract

We propose the term aura to enrich the current language for designing and analyzing media experiences, especially when using augmented reality, mixed reality and ubiquitous computing technology. Aura describes the cultural and personal significance that a place (or object) holds for an individual. An MR application can exploit aura to make the user’s experience more compelling or educationally rewarding. Aura provides a necessary complement to the concepts of presence, which is commonly used to evaluate VR applications, and of place, which refers to the more generic significance of places, particularly in CSCW applications. We use the Oakland Cemetery in Atlanta, Georgia to illustrate the concept of aura. A number of research questions about the relationship of aura, presence, and place are suggested.

Keywords--- Augmented Reality, Mixed Reality, Presence, Aura, Cultural Heritage

1. Introduction

For more than a decade, since the publication of the first number of the journal Presence, presence has been a key concept for understanding and evaluating the effectiveness of virtual environments. VE researchers have used the term to describe the mental state of the user in response to being immersed in a virtual application [24][31]. They have offered a variety of definitions and proposed a variety of measures for presence, and presence remains a powerful conceptual tool for many applications. During that same period, however, two related developments have occurred in computing that call for a broader application of presence, while testing the limits of the current definitions and concepts used in presence research.

The first development is that the research community and the computing industry have shown increasing interest in devices and applications that integrate virtual information into the user’s physical working or leisure environment. For example, ubiquitous computing researchers have focused on pushing computation into the world where it blends in with other available tools [30], designers of tangible media emphasize the importance of physical manipulation [11], and augmented reality (AR) systems tightly couple synthetic media with the physical world to create perceptually-integrated physical/virtual worlds [7]. For simplicity in this paper we will refer to all such systems that combine physical and virtual worlds as mixed reality (MR). MR techniques have now been explored in a wide variety of domains, including battlefield visualization, maintenance and repair, air traffic control, urban design, office work, games, and medicine.

It is not clear that presence per se is useful for evaluating MR applications, at least not as it is commonly defined by VR researchers. If presence is defined as a sense of being in a physical place, then one might be tempted to say that MR applications get presence “for free,” since the subject can see and interact directly with the physical world. (As we will discuss later in this paper, the mediated nature of MR applications makes the situation more complicated, especially when one considers the more general definitions of presence that focus on mediation rather than sense of place1.) Researchers in MR applications have in general avoided using presence, preferring a variety of other approaches and concepts, including the concept of place [6], which we will discuss below.

The second development is the rise of a new class of applications, which use MR, VR, or even desktop multimedia to present culturally significant experiences. Researchers in Europe and North America are deploying digital technology to enhance the user’s experience of cultural heritage objects, such as paintings and historic artifacts, as well as such sites as museums, castles, historic homes, and battlefields [1][3][10][22][27][29]. There are also entertainment applications in which computer-controlled digital information is displayed in situ in the physical world. (In a sense, much of Disney World and other theme parks are really mixed reality experiences.) Again the concept of presence alone does not seem adequate to describe these applications. Marsh shows that media theory (the work on earlier media forms, such as film and drama, as well as the work of contemporary digital theorists) can be used to expand on the concept of presence for these applications [16]. Marsh argues, and we agree,

1 The International Society for Presence Research defines presence in relation to mediation. See http://www.ispr.info
that new media experience designers can draw on the techniques of filmmakers (and other media) to engage the emotions of the user. This emotional response of a person to a significant place or object is often powerful and can be leveraged by media experiences.

Many designers who create experiences in historic sites are already explicitly trying to leverage this psychological response. However, without a common framework and terminology for these aspects of places and objects, it is difficult to discuss the principles on which the media experiences are based and thus learn from previous designs. Over the past four years, our own work has increasingly focused on experiential MR applications, such as tours of, and dramas set in, historic locations [14]. The appeal of these experiences (for both ourselves and other researchers) derives from the rich historical, cultural, or personal meaning that such places and objects have for the participants. If a place is emotionally engaging to begin with, an in-situ experience should be more engaging than a film, web site, or VR application that is experienced away from the site.

To describe and eventually to measure the user’s sense of emotional engagement, we are proposing a concept that we call “aura.” We believe that aura provides an important complement to presence and can enrich our understanding of users’ responses to a variety of different computer-mediated experiences, especially in cultural heritage sites. In addition, aura can serve as a bridge to relate the VE research in presence with the research by the MR and CSCW communities on the concept of place.

In the rest of this paper, we define and elaborate on the concept of aura. As an example, we discuss MR experiences we are designing to exploit the aura of a historic site in Atlanta, Georgia. We also discuss the complex relationships between aura and the concepts of presence, place, and embodiment. Finally, we suggest ongoing research questions: in particular, how aura might be measured and correlated with presence.

2. Definition of aura

Consider the following question. Since almost anyone can afford to hang a reproduction of a painting like the Mona Lisa on their wall, why do millions bother to travel long distances to Paris to visit the Louvre in person and crowd around Da Vinci’s painting? Clearly it is important to these visitors to be in the physical presence of the painting; there is “something” about being near the real thing. This something is the Mona Lisa’s “aura”, which is the sum of the painting’s historical (e.g., it’s age and who painted it), cultural (e.g., we place a high value on such art objects), personal (e.g., you may have visited the Mona Lisa during a long ago vacation with a cherished friend) contexts for a viewer. In this case, for many in Europe, North America, and elsewhere, both the Mona Lisa and the Louvre museum (as a place) have aura.

Media theorist Walter Benjamin introduced the concept of aura in the early 20th century to describe the effect on the viewer of traditional art objects [2]. Benjamin argued that new media technologies, specifically photography and film, led to new forms of art, in which uniqueness (or aura) no longer mattered. Benjamin’s historical analysis is still influential among film and other humanistic theorists. We wish to adopt and adapt Benjamin’s notion of aura; however, we do not agree that “reproductive technologies” (such as film in Benjamin’s day or the computer now) always destroy aura. Nor do we agree that aura is an exclusively bourgeois value that popular art and entertainment necessarily want to reject2. We wish to apply the concept of aura to a wide variety of culturally significant objects and places and to consider how computer applications can be designed to augment or inform the user’s experience of such objects and places. We offer the following definition of aura:

The aura of an object or place is the combination of its cultural and personal significance for a user or group of users.

The “cultural significance” refers to the shared meaning for a large community; the known history of the object or place plays a major role in the community’s shared interpretation. “Personal significance” refers to the individual associations that the place or object may have for a particular user. An object or place can have elements of both sorts of aura.

Strictly speaking, all aura is personal, because “aura” describes the psychological response of one individual to the object or place. The personal nature of aura is fundamental to the concept: aura can only exist if the individual can connect the object or place to his or her own understanding of the world. Anyone can construct a trivial example to illustrate this point (e.g., an Inuit from northern Canada may have no knowledge of, and thus experience no aura for, the Mayan ruins of Tulum). The implication of this assertion is worth stating explicitly: increasing the connection to a person’s understanding of the world can increase the aura for that person.

Although all aura is personal in this sense, we can also speak of the collective aura of a well-known cultural object, such as the Mona Lisa, because thousands or millions of individuals in our society have a similar response to the painting. Many such one-of-a-kind art objects and historic locations have aura. But even mass-produced objects may have aura: e.g. the store-bought birthday card that a friend gives you. The modernist artist, Marcel Duchamps, picked out common objects, his so-called “ready-mades” and designated them as art by placing them in a museum, giving them aura1.

2.1 Aura as contextualized presence

Researchers in the BENOGO project, whose aim is to create the experience of “being there without going there”, are attempting to understand (and leverage) aspects of

---

2 In a subsequent paper, we intend to discuss in more detail the question of aura in 20th century media, including new digital media.

1 In a complicated way, Duchamps and other modernists such as the dadaists were playing with and calling into question the concept of aura, as Benjamin himself recognized.
unique physical locations to create more compelling VR experiences. They have suggested that research on presence in VR has been hampered by the fact that most VR applications offer their users a visual world that is generic. As they put it, if presence is the feeling of “being there,” then we still have to ask: where is “there”? They argue that presence can best be achieved by placing the user in a meaningful context, and call this approach “contextualized presence” [27].

The BENOGO research in VR suggests to us another way to conceive of aura. Aura can be thought of as the difference between a place or object and a perfect reproduction of that place or object: for example, the difference between the physical world and the holodeck in the television and film series, Star Trek, or the difference between the Mona Lisa and a perfect copy. We understand this difference, however, not in any metaphysical sense, but rather in terms of user response. Assuming that the user knows or is told that she is in a VR replica of the Louvre and not in the physical Louvre, her experience will be influenced by that knowledge even if the sensory input is exactly the same as it would be in the physical location. Thus, even a perfect VR exhibit could lack aura and thus not be as compelling as an experience in the physical location. Similarly, if someone gave you a perfect copy of the Mona Lisa (and told you it was a copy), you would not react to it in the same way that you would to the real painting. This distinction highlights both the strengths and the potential limits of projects like BENOGO. By increasing a user’s sense of place, the BENOGO researchers are showing how to create more compelling, more “auratic” VR experiences. It is not clear, however, that it will ever be possible to create virtual experiences that are “just like being there without going” as long as the user knows they are in a VR environment.

2.2 Other uses of the term “aura”

Some in the CSCW, VR and HCI communities have used the term “aura” in a different sense, typically to refer to some invisible property around an object. For example, in VR, aura has been used to indicate the extent of the physical interaction between two objects [8]. The Audio Aura system used the term “aura” to refer to the invisible information space through which the user moved, which was made manifest through audio [19]. However, the humanities have been using the term in Benjamin’s sense for much longer, and we believe that Benjamin’s sense of the term can also prove useful in explaining a range of digital (and other) media experiences.

2.3 An example of the aura of place: the Oakland Cemetery

The historic Oakland Cemetery in Atlanta, Georgia, USA will serve as our principal example of a location with aura. (See http://oaklandcemetery.com). Oakland was the major burial site for the city from 1850 to the early 1900s, when the grounds were more or less filled. Oakland physically embodies the history of the conflicts and achievements of the city. There is a section for Confederate soldiers, one for African-Americans, one for the Jewish population of the city, and one for paupers, whose wooden grave markers have disappeared. Famous southerners, major political figures (including 6 governors of Georgia and some 25 mayors), as well as economic leaders and the “inventor” of Coca Cola, are buried in Oakland [26]. Oakland is also a “garden cemetery”, designed and used as a park, and is part of the Martin Luther King, Jr., National Historic District. For these reasons, Oakland possesses considerable aura for those who live in Atlanta and are familiar with this history.

Any cemetery has a certain cultural significance for anyone in North American or European society. For this population, a cemetery evokes some combination of apprehension and pensiveness because of associations about the dead, although this may not be true for all cultures. Any physical cemetery would produce some such response, even for people who know nothing about its particular history. This common cultural understanding of the meaning of cemeteries, as well as an understanding of acceptable behavior in a cemetery, is what defines what we might call, following Dorough and Harrison [9], the *placeless* of the cemetery for our culture.

In addition to this general cultural meaning, however, a particular cemetery may have special significance. The circumstances of Oakland Cemetery, along with its art and architecture, give it aura for those who know its history. Similarly, personal relationships with occupants of the cemetery (e.g., a relative or friend) would imbue a particular cemetery with personal aura. Aura can have a positive or negative valence. For example, an African-American visitor to Oakland might find the monuments to the Confederacy repugnant; the cemetery would still have aura for such a person. Oakland’s aura can vary in intensity as well as valence: long-time residents of the city may cherish it, but some residents and many visitors may know little about it. Oakland’s potential aura is great, because most visitors can find something meaningful in the range of historic figures buried there, its relationship to the history of the region, and its art and architecture.

Because of its value and location, Oakland receives about 40,000 visitors a year, ranging from tourists to city residents to children on school trips. There are a variety of traditional, human-guided tours for visitors, designed to focus on the interests of different audiences. For example, school children may receive a tour highlighting the historic figures they are studying in their civics classes, while tourist may receive a tour focusing on “colorful” inhabitants. Each of these tours has been designed assuming certain knowledge of (and thus aura for) the cemetery and have the (implicit) goal of increasing the aura of Oakland to enhance the visitor’s experience and make the visit memorable.

3. Aura and media technologies

The design of the tours of Oakland cemetery suggests a larger issue: how can aura be evoked by various media technologies, especially MR technologies? In our
discussion thus far, we have implied that by creating an experience in a place (or near an object) with aura, the experience itself will share in that aura. In this section, we address this idea explicitly.

Media technologies do not themselves possess aura in the mind of the user; the aura belongs to the object or place. The technology may evoke the aura a person already holds for an object or place and *leverage* that aura to improve the media experience. In addition, we suggest that a media technology can *enhance* the aura of an object or place by increasing the connection to the user’s personal experiences and knowledge (see Figure 1). On the other hand, media technologies can also interfere with or diminish the experience of aura, which was Benjamin’s original point about the technology of film [2]. For example, an inappropriately humorous use of ghosts in the Oakland experience could make it harder for the user to understand the contributions made by the historical figures or may cause negative reactions in visitors with personal relationships to occupants. It also seems that, by reducing aura, designers may allow people with overwhelming associations for a place to experience that place, similar to the way that VR exposure therapy can help people gradually overcome their fears (e.g., Virtual Vietnam [21]).

As Figure 1 indicates, the relationship between aura and the media experience is reciprocal. The designer leverages aura to improve the media experience, and the media experience can also enhance the aura of the object or place for the user. In fact, the aura of the place or object is often so tightly integrated into the media experience that it becomes difficult to distinguish between the processes of leveraging and enhancing.

Most people have experienced aura through a media experience, such as the sense of awe that might be conveyed by a documentary film on the construction of the Egyptian Pyramids. The film itself does not have aura, but evokes the aura of the pyramids. Clearly, both the media technology used and the mode of presentation can contribute to the degree of aura that is evoked. The media form chosen (documentary or action film, *in-situ* MR, desktop hypermedia, etc.) provides affordances and imposes constraints on the aura of the experience. An action film like *The Mummy*, while situated near the pyramids, takes advantage of our knowledge of ancient Egypt, but would not evoke the same sense of aura as a well-done documentary. Likewise, an MR application that allows the user to walk around the remains of the pyramids in Egypt should be able to leverage this physical proximity in a way other media (including our previously mentioned documentary) cannot. The choice of the mode of presentation (descriptive text, third-person narrative, dramatic mode) should also affect the aura evoked by the application.

To sum up, we could envision various media presentations of Oakland using traditional as well as digital technology: an MR experience, a historical photograph, a historical painting, a VR experience, a desktop multimedia presentation, and a traditional film documentary. Each of these presentations might evoke some of the aura of Oakland, but they would do so in very different degrees. The relationship between the content of an experience, the media form, and the user is sufficiently complex that it is not possible to construct any simple scale of technologies and their ability to evoke aura. In fact the relationship between media form and aura raises a number of research questions.

### 3.1 Enhancing aura

If users are not aware of the significance of a place or object, they will not experience aura. This is especially true in educational applications. For example, many visitors to the Oakland cemetery would not know about many of the significant aspects of cemetery (e.g., history, art, architecture, famous residents). Thus, most of the current tours explicitly try to enhance the aura by presenting a variety of interrelated information to the user. This in turn requires adding to the user’s existing network of knowledge by making new links: for example, connecting Oakland Cemetery to the user’s knowledge of the Civil War.

### 3.2 Leveraging aura

We hypothesize that in order to leverage aura, a media experience must bring the object or place “closer” to the user. We suggest that auratic proximity cannot be measured simply by physical distance, but rather by the user’s perception of their relationship to the object or place. While we expect users to feel the aura most strongly if they are in the same physical place, any media experience that engages the user sufficiently can leverage aura.

We cannot assume, therefore, that “better” (e.g., more visually accurate, more interactive, etc.) media technologies necessarily correlate with a greater sense of aura. As noted above, the perfect VR environment, the Star Trek holodeck, would achieve the illusion of the absence of mediation and induce a high degree of presence, but because the user knows the environment is an illusion, the holodeck might not evoke the aura of the place being simulated. Older, less sophisticated technologies may evoke greater aura because of our cultural assumptions about them.

For example, from the later nineteenth century until recently, our culture has usually assumed that (in the absence of “tricks”) photography showed what was “really” there [17]. Because of this widely shared assumption, a photograph may still be regarded as better at evoking aura than a computer graphic image, even if the photograph is a grainy, black-and-white image—indeed precisely because it is grainy and black and white. When we look at a
4. Aura and the mediation of presence

The technology of mediation is an important factor for both aura and presence. As we remarked above, if presence is defined as a sense of being in a physical place, then one might be tempted to say that MR applications get presence “for free.” The subject in MR always has a sense of place and therefore always has a sense of presence.

But in fact, even in an MR experience, the world is always filtered at least partly through media: the computer graphics, text, or audio that augment the user’s view of the physical environment. A VR experience is (almost) entirely mediated, since VR cuts off the user’s ability to see and (often) to hear the physical world. An alternative way to define presence, then, is to focus on the user’s capacity to forget about the mediating technology.

This is the strategy of Lombard and Ditton and some other presence researchers. The website for the International Society of Presence Research defines presence as “a psychological state or subjective perception in which even though part or all of an individual's current experience is generated by and/or filtered through human-made technology, part or all of the individual's perception fails to accurately acknowledge the role of the technology in the experience.” (see http://ispr.info) According to this definition, the key aspect of presence would be the illusion of non-mediation or transparency. Some researchers prefer an operational version of this definition, in which an application has presence if the user acts and reacts as if she is in the corresponding real world setting.

If presence is defined with Lombard and Ditton as the feeling of the absence of mediation, then MR applications can have more or less presence, just as VR applications do. An MR application combines the physical world with a virtual world. If the combination is seamless, then the user could experience a high degree of presence. However, the illusion of presence (the integration of the physical and virtual) can be broken by technical or design based problems and inconsistencies in the experience. Lag in the tracking system might cause virtual content to fail to remain registered with the physical world, while the lower resolution of a virtual image might stand out dramatically against a real object. Even when the technical aspects of the experience are perfect, inconsistency, lack of realism, or poor design of the sensory and behavioral aspects of the virtual elements could cause breaks in presence.

Although presence and aura are clearly related (both are psychological responses of the user to the experience), the relationship is not simple or unidirectional. The relationship is also related to the nature of the media technology used (see Figure 3). It seems to us that aura can enhance presence. It also seems that in some cases presence can increase the aura by making the object or place seem closer. Yet, the user’s knowledge of and personal relationship to the place are probably more important than presence for generating a feeling of aura. Consider the various media technologies that could be used for an experience in Oakland Cemetery. A desktop multimedia experience would be relatively low in the sense of aura as well as in presence. A VR version of the Oakland cemetery experience would be higher in presence, and for this application, we suspect that the greater the presence, the greater the sense of aura. Yet neither of these experiences may evoke as great a sense of aura as an historical photograph of Oakland Cemetery, such as the one in Figure 2. Put another way, a generic VR application could conceivably achieve a high degree of presence through perfect immersion without evoking much or any sense of aura in the user.

5. Aura, space, and place in mixed reality

The concept of aura also bears an important relationship to the concepts of space and place in MR systems (following the terminology of Harrison and Dourish [9]), which are both situated in and defined by their relationship to the physical world. Much has been written about the practical benefits of MR related to leveraging the physical, perceptual, and social affordances of being in a physical space. At the most basic level, MR systems try to directly leverage our physical and perceptual abilities and
the corresponding affordances of the space in which they operate. For example, the URP system [28] allowed the designer to physically manipulate the various inputs to a simulation of an urban space. The KARMA system [7] leverages the user’s perceptual ability to merge maintenance instructions with the physical equipment on which they are superimposed.

Physical spaces and objects also offer social affordances, such as eye-to-eye contact and other physical cues, that we unconsciously leverage during our social interactions to signal interest, negotiate conversational flow, and so on. These affordances have been studied extensively in the context of distributed collaboration [20] and media spaces [9]. They have also been demonstrated in tangible interfaces like the Designer’s Outpost [12] and the Augmented Flight Strips [15]. In these systems, the users work directly with physical props that are transparently linked to their virtual counterparts and leverage the existing social dynamics of the groups to enable them to manage a larger number of objects than they could with a non-tangible system.

Over the past several years, researchers have developed a sophisticated understanding of the user’s relationship with their work environment in MR systems and have developed the concept of “place” to describe that relationship [9][18]. Harrison and Dourish refer to the often-implicit social and cultural meaning of a space as its placeness [6]. They argue convincingly that the shared meaning, and therefore mutual understanding, which people attach to real places is what designers are often seeking when creating MR systems. Research has focused on the implications of place with an eye toward deep integration of information technology into work settings. While Harrison and Dourish do consider places that have personal significance for the user, such as homes, they focus on design applications that do not exploit the unique characteristics of such places.

As we would expect with an influential concept, research has led to suggestions to nuance the distinction between place and space. In addition to the BENOGO project cited above, Perry and Brown have discussed the concepts of place and space in the context of personal geographical-based information systems (that is, maps and guidebooks) and argued for a more subtle understanding of the relationship of between place and space [5]. Both maps and guidebooks and have elements of both place and space. What interests us here is movement toward a concept of aura. One important purpose of a guidebook is to inform the user about the special historical and cultural meaning of a particular place. Guidebooks enhance aura as they augment the user’s sense of place.

Such work points to a lack in the current conceptual framework. Just as place was developed to supplement the concept of space, we believe that place now needs to be supplemented by the concept of aura. In previous research, the concept of place has been applied most often to the social context and meaning of the workplace environment. Harrison and Dourish’s key example is media spaces, such as the PARC environment. The PARC system functioned in researchers’ offices, which were fitted with cameras and audio links to allow researchers to interact without being in the same physical location. The concept of place is well suited for describing the user’s relationship to such spaces for CSCW, because the offices had no cultural significance, although they were differentiated by personal use.

For a new class of applications, where the digital technologies are used for entertainment and informal education in well-known cultural settings, we need a new concept like aura. Just as place provides an analytical tool for understanding and improving CSCW, aura provides a tool for understanding and improving experience design.

5.1 Aura and embodiment

In addition to the concept of place, Dourish has stressed the importance of embodiment in application design. He argues for a new paradigm, which he calls “embodied interaction,” an approach that combines elements of tangible computing and social computing [4][6]. Embodiment is also a key aspect of aura. The importance of embodiment was made clear in Benjamin’s original formulation, when he claimed that aura was generated by the “here and now” of the place or object. “Here and now” describes the immediate, embodied presence of the object or place.

Tangible computing explores ways to incorporate the virtual information provided by the computer into the user’s physical environment: in short, how bits can become embodied. The design strategies of tangible computing have been applied to generic work situations (KARMA [7] or Flight Strips [15]) in which the computer can display information on the user’s physical environment or gather information from the user’s physical interactions. In most cases, therefore, tangible applications explore the physical affordances provided by generic objects (printer parts, paper flight strips), not by culturally unique objects (such as particular paintings or the artifacts of a particular battle). AR experience design can be seen as a new form of tangible computing in which the success of the interaction depends on the uniqueness of the object or place.

A similar qualification can be made in regard to the other aspect of embodied interaction. Social computing usually refers to the use of digital tools to foster the negotiation of social meanings — again, the typical case would be CSCW. Dourish explains how communicative meaning is achieved through negotiation among the users of such systems. AR experience design, however, can be seen as a form of social computing in which the applications are representations rather than tools and have a symbolic meaning within the culture. For that reason it might be better to call the design strategy “cultural” rather than “social” computing.

Such cultural applications do not necessarily offer their users an equal role in the negotiation of meaning. A negotiation of meaning occurs between user and designer similar to that between a playwright and an audience in dramatic presentation. That negotiation takes place at the interface between the virtual and the physical aspects of the experience, because the user must interpret what the virtual elements mean in this particular physical context. For
example, in a dramatic experience presented in the Oakland Cemetery, the designer might create virtual characters that represent ghosts of people who lived in Atlanta in the 19th century. The user would interpret the status and meaning of these characters, according to many of the same conventions that she would use if she were watching a drama in a theater. The principal difference is that she is not seated in a theater, which is in fact a space that the drama asks her to imagine as a particular place. Instead, she knows that her body is located in the actual space of Oakland, which she shares with the virtual characters. The user’s embodied presence at the site adds a vividness that distinguishes AR experiences both from conventional stage drama and from social computing applications in the workplace.

6. Future work

Like presence and place, the value of the concept of aura will become apparent if it proves useful for analyzing a class of media experiences with the eventual goal of improving them. For example, one of the appeals of computer-augmented tours has been the potential for customizing experiences for each visitor, based on simple demographic information. We believe that the framework of aura can provide guidance on how to approach the problem of what data to collect and how to customize the experience. One version of tour of a famous battlefield could be designed assuming recent familiarity with war. Depending on the region in the world a visitor is from (e.g., Eastern Europe), the content could be related to the current conflicts (e.g., the conflicts following the fall of the Soviet Union).

At this initial stage, however, there are many research questions regarding aura that must be addressed if the concept is in fact to become useful as an analytical design tool. First, the interrelationship of aura and presence suggests that both of these concepts could be thought of as aspects a more general concept that we might call “engagement.” Although presence has generally been applied to VR, we could also step back and refocus on the more general definitions of presence as a measure of engagement in any computer-mediated experience, not just perceptually immersive media such as VR (e.g., people are frequently “completely immersed” in books, movies, and computer games).

Second, there is the question of how to best measure aura. It seems that, at best, aura can be measured qualitatively. Because aura, like presence, is a psychological condition of the user, it should be possible to use similar techniques to those used for presence. In particular, questionnaires and think-aloud protocols have been developed for presence and are widely used [32][25]. The BENOGO project has put forth evaluation techniques for their concept of contextualized presence, and these might be modified to assess aura. While we hope that a general set of questions can be devised to measure aura, it may be that the site-specific nature of aura requires evaluation instruments specific to each place or object.

Finally, the development of instruments to measure aura would allow us to ask a set of questions about the correlation between aura and other properties of the experience. For example,

1. Is there a direct, positive correlation between aura and presence in an MR experience?
2. What effect does aura have on the user experience (i.e. increased learning or engagement, as might be implied by studies of in situ education [23])?
3. Does user participation enhance aura or not? Among HCI researchers and experience designers, interaction is generally held to be a good or even necessary feature of digital applications. But does participation enhance aura in a media experience? The question is worth asking, because, for example, it is not clear that allowing a visitor to touch the Mona Lisa in the Louvre would enhance her sense of wonder at the painting. It might have the opposite effect.

6. Conclusion

Aura is a defining feature of a class of culturally and personally specific applications for entertainment and education that can be used formally and explicitly in various forms of experience design. All media technologies should be able to leverage and extend the aura of a place or object for a person, although MR technologies that create situated experiences would seem to do so more readily.

There are two key points to remember regarding aura. First, aura arises from the significant historical, cultural and personal aspects of a place or object, in contrast to the more generic social and cultural constructions associated with all instances of the place or object. Second, aura is a relationship between a person and the place or object—media experiences do not in general have aura, but can leverage or enhance the aura a person feels.

The concept of aura provides an important complement to presence and can enrich our understanding of users’ responses to a variety of different computer-mediated experiences. In addition, aura can serve as a bridge to relate the VE research in presence with the research by the MR and CSCW communities on the concept of place.

7. Acknowledgements

This work has been supported by a grant from GVU Center at Georgia Tech. Much of our thinking on this topic has occurred in the context of our Augmented Reality Design class, which has been supported by an equipment donation from Hewlett Packard. We would like to thank the students who have taken our AR Design classes over the past four years for many challenging discussions that have helped us refine the concepts presented here. Finally, we would like to acknowledge the gracious support of John Avery and Kevin Kuharic at the Historic Oakland Foundation, who have enthusiastically supported our work (and our student’s research) over the past two years.
References


Working Memory and Presence: Reconsidering the Role of Attention in Presence

David Nunez
Collaborative Visual Computing Laboratory, University of Cape Town
dave@cs.uct.ac.za

Abstract

This paper presents a working memory based model of attention in presence, in which presence depends on the allocation of cognitive resources to process an environment. In the model, media decoders (cognitive modules specialized in decoding particular media) extract data about the environment in an abstracted form. In order for presence to occur, a certain amount of working memory is required for processing this abstracted data. However, as the extraction process itself requires working memory, the media can have a large impact on the presence experience. To demonstrate its usefulness, the model is used to explain Biocca’s book and reality problems. To conclude, a load-based methodology for testing the model is proposed.

1. Introduction

Attention has been recognized as an important variable in presence (for instance in [1]). The importance of attention has been made pre-eminent by Biocca [2] in his posing of the ‘reality problem’ in presence theory. Briefly put, the ‘reality problem’ asks how it is possible for a person to be physically in an environment, and yet not experience presence there. This is a problem which affects most current presence theories – partly because they begin with an assumption that virtual environments are approximations to reality (for instance, [3]), and partly because they do not explicitly consider the role that attention plays in the cognitive processes which give rise to presence. This paper proposes that Baddeley’s concept of working memory [4], which explicitly deals with the allocation of cognitive resources, can be fruitfully applied in presence research to explain a variety of phenomena such as the reality problem, the book problem, and why text-based virtual environments tend to produce inferior presence experiences to immersive displays.

2. Attention, presence and working memory

Attention as a variable in presence exists in numerous places in the literature. For example, Barfield & Weghorst [5], in proposing a conceptual framework, specifically formulate attention as a resource allocation process which acts as a mediator in presence. Similarly, Wirth et al [1] propose that before presence can occur, the media must attract and hold the subject’s attention. Empirically, Schubert et al [6] in a factor analysis of eight questionnaires found attention as a first order factor.

2.1 Cognitive perspectives on attention

Attention and its relation to various tasks and cognitive processes is a well studied phenomenon in applied psychology. For example, [7] presents evidence of the importance of attention in relation to reading comprehension, and others have studied its role on the task of aviation (cf. [8] for a review). These tasks, like presence, involve inferring an environment from a set of perceptions. A large number of these explanations apply Baddeley’s model of working memory [4]. The working memory system is similar to the better known short term memory system in that it has a limited capacity, but unlike short term memory, which is conceived mainly as a system involved in encoding memories and remembering, working memory is seen as a general purpose working space where cognitive processes are applied to data [4]. Working memory is thus not simply related to remembering and encoding memories, but rather as the site where most cognitive processes take place, and where the contents of awareness exist [4].

Working memory is often studied in relation to attention, and it is generally accepted that all stimuli which are attended to equate to the contents of working memory [9]. The relationship between selective attention and working memory is not as clear, although it is thought that working memory space is allocated to new stimuli based partly on the processes currently running in working memory [4]. What is clear, however, is that the amount of working memory resources assigned to a particular task is analogous to the amount of attention which is being focused on that task, and working memory is thus a useful indication of how much attention has been allocated to a task [9].

2.2 The allocation of resources to a task

How working memory is allocated to a task is a complex process. Simply having more information to process does not imply that more working memory will be allocated. Information which is meaningfully related is chunked into more abstract complexes, effectively
freeing up working memory space [4]. Thus, a highly immersive display system which renders a large number of correlated variables may require little working memory to process, as the information can be easily chunked; however, if a latency were to develop in the display of one of the variables or channels (for instance, in sound rendering), then the temporal discrepancy would prevent the chunking of sound together with the other variables, and thus more working memory would be required to process the scene. This same mechanism can explain why stimuli which come from ‘outside’ the virtual environment (as discussed in [10]) can reduce presence – these stimuli will not chunk with the stimuli associated with the virtual environment, and will thus require more working memory to process. If enough of these distractions abound, then they will begin to impinge on the working memory which is necessary for successfully processing the virtual environment, and the subject will begin to become unaware of elements of the environment. This reduction in the amount of attention focused on the virtual environment will lead to a reduction in the sense of presence.

2.3. The link between working memory and presence

An important question that remains is how presence is linked to working memory. The working memory model has been used to explain phenomena which, from a micro-cognitive perspective, are similar to presence. These include investigations into working memory in reading comprehension [7], as well as studies on working memory and navigation [11]. These tasks involve the processing of a subset of external stimuli so as to decode some meaning (be it spatial or otherwise) to allow further inferences about the space or behaviour in the space. If one accepts that such a processing task is necessary for presence, and given that all processing requires some working memory, then it follows that for presence to occur, some amount of working memory will be required to process the environment.

Using working memory in this way changes the relationship between presence and attention slightly. Rather than considering attention primarily as a filter or selector for external stimuli, this view considers attention as the allocation of cognitive resources (in the form of working memory) for the processing of particular aspects of the external stimuli or environment. Thus, instead of thinking about presence occurring because the virtual environment is being attended to rather than the real environment, or because the real environment is not providing any distractions [10], one would speak about the amount of working memory allocated to processing the virtual environment. From the working memory perspective, distracters from outside the virtual environment negatively affect presence because they require working memory to process; this in turn must be taken from the limited pool of resources which are also being used to process the virtual environment.

3. Working memory and the processing of media: A two-process model

Central to applying working memory theory to presence research is an idea of the role that working memory plays in the processing of media, given that all virtual environments are encoded in some medium [12]. This is a particularly interesting problem, because presence can arise from a wide variety of sources, from text-based media to highly immersive media such as a cave display. From a cognitive perspective, one might pose the problem thus: How can virtual environments be encoded in a variety of representations while still having the same basic cognitive effects?

One possible solution to this problem comes from separating the act of processing a virtual environment into two sub-processes. In the first process, the medium which encodes the environment is decoded so as to transform the perceptual stimuli into more abstract representations (such as scripts or schemata [13] or spatial situation models [1]). Then, in the second process, these representations are used to draw inferences and make decisions about the virtual environment. It is during this second process, when higher level cognitive processing is occurring about the virtual environment, that presence is most likely to occur [1]. Shifting the ‘presence cognitive event’ (if one may be so vague) to the second process in this way allows this simple model to predict presence occurring regardless of the medium that the environment was encoded in (as suggested in [14]).

3.1 Media decoders and working memory allocation

The mechanism by which perceptual stimuli are decoded (the first process discussed in section 3 above) is a complex one, because the specific processes applied to the stimuli will vary depending on the way the stimuli are encoded. For example, decoding a photograph of a room requires different cognitive processes to decoding a verbal description of that room, even though the final product of that decoding will be similar in both cases. Also, it is important to consider that media require varying degrees of effort to decode – for instance, decoding a photograph occurs much faster and with less conscious effort than decoding a written text. Finally, one needs to take into account individual differences with regard to the ease with which people are able to decode some media (for instance, reading skill increases with age and with practice [7]).

One can model this process by considering the decoding of each medium as being done by a separate cognitive module whose input is a set of perceptual stimuli, and whose output is a set of abstract representations of the content of the medium (schemata, spatial situation models, etc.). Each of these media decoding modules is a collection of strategies and processes for decoding a particular medium; thus one can think of there being a writing decoder (after evidence
from [7]), a film decoder (after evidence from [15]), a diagrammatic decoder (after evidence from [16]), and so on. When a new set of perceptual stimuli are considered for processing, the appropriate media decoder would be selected on the basis of a small set of key features in the stimulus set (for instance, the basic shape of letters in the spatial memory buffer might trigger the writing decoder). Once the media decoder has been activated, it would proceed to decode the stimuli. If a media decoder attempts to decode the wrong type of medium (for instance, in the case of a picture being embedded in text as occurs in a magazine), then the error would signal for the selection of a new decoder.

One can link the notion of media decoders to working memory without much difficulty. Because decoding a stimulus requires cognitive processing, it will therefore require some degree of working memory [4]. How much working memory a particular media decoder will use is, however, a complex question which will require empirical investigation. It seems reasonable to suggest that decoding some media will require more working memory than decoding others. For instance, images and video are relatively easy to decode (partly because there are brain structures which are specialized for this task), whereas writing is harder to decode (as it requires first a visual pass to decode individual letters and words, and a parallel second semantic pass to decode the meaning of the sentences as a whole [7]). It also seems reasonable to suggest that some media decoders become more efficient with use and practice; for instance, reading requires time to learn and generally improves with practice, eventually becoming effortless. Similarly, some film genres make use of conventions which must be learned at first, but are later decoded with little effort [15]. From a working memory perspective, this increase in efficiency of the decoding process and associated subjective sense of effortlessness comes from a decrease in the amount of working memory used by the decoder [4].

4. Addressing the book and reality problems with working memory

An important question that one might ask about the use of working memory theory in presence is ‘does it work?’ Clearly, empirical evidence would be the most satisfactory way of answering such a question (see section 5 below for a discussion of how this evidence could be collected). Given that this idea is still in its preliminary stages, an alternative method of checking its validity can still be applied – by checking if the model is capable of responding to the book and reality problems [2]. These problems were posed as general limitations of current presence theories (particularly those which emphasize the role of immersion and underemphasize the role of attention). Therefore, it seems that any new theory of presence, particularly one dealing explicitly with attention and non-immersive environments, should be able to respond to these two problems.

4.1 The reality problem

The reality problem deals with a common experience – someone who, while physically in the real world, experiences very little presence there (due to the subject being lost in their thoughts or something similar). In this situation, the fact that the subject is completely immersed with ‘high-fidelity’ sensory stimulation from the environment seems to have little effect on presence. Working memory can be used to explain this by examining the cognitive state of the subject. What are the contents of working memory at the time? If the subject is thinking deeply about something other than the environment they are in, then very little working memory capacity is likely to have been allocated to processing the environment. Thus, very little information exists for the drawing of inferences about the environment or about even forming memories of their experience; indeed, if the environment is not allocated much working memory, the subject will not even be aware of it [4], and so very little presence will result. Applying working memory theory to the reality problem has the further benefit that it allows one to explain this phenomenon in degrees (one could be said to be more or less present in one’s environment even if lost in thought). This is explained by stating that the degree of presence in any environment (real or otherwise) is a function of the amount of working memory allocated to processing that environment. Thus, if one is paying a little attention to the environment, only a small amount of working memory will be allocated, and a maximum of only a little presence can be expected. However, if a lot of attention is given to the environment, then a lot of working memory will be allocated, and the maximum amount of presence expected will be large. Note that under this model, the amount of working memory allocated to processing the task does not guarantee an increase in the degree of presence experienced. The amount of working memory allocated only predicts the maximum degree of presence which a subject can experience. If the display system represents low fidelity or low-immersion stimuli, then the presence experience will not be particularly intense, regardless of the amount of working memory allocated to processing the environment [5].

4.2 The book problem

The book problem asks how books, arguably the lowest form of immersive display, can lead to presence experiences. The two-process model described in 3.1 above, together with media decoders and working memory can be used to respond to the book problem. In the two-process model, presence is seen as occurring in the second process and therefore the medium representing the environment is relevant only in terms of how it is decoded by the relevant media decoder. Thus, if a decoder is available for the given medium, and the decoder does not use too much working memory for processing, enough working memory should be available for the second process, and thus presence is a possibility.
If one takes into account that media decoders can account for individual differences in media competency (as discussed in 3.2 above), then it is also possible to explain individual differences in presence experiences (as discussed in [14]).

5. A working memory methodology for presence

As working memory has been well researched by psychologists, it would be sensible to consider their methods and adapt them to presence research. In general, working memory research makes use of a memory load paradigm, which is used to investigate to what extent a particular main task (such as navigation or reading) relies on working memory. Subjects are divided into groups, each of which is given a loading task of varying intensity. This loading task (for instance, remembering a list of unrelated items [4]) is used to place a demand on the working memory systems and is performed simultaneously with the main task under investigation. The intensity or load of the loading task is then the independent variable of the study, and performance on the main task is the dependent variable. The degree to which the main task makes use of working memory will be shown by the rate at which its performance declines with increasing working memory load (cf. [4] for numerous examples of this paradigm in use). In a presence study, memory load tasks can be easily integrated into a virtual environment. Integration can prevent confounds related to having a task which is unrelated to the environment and may become a distracter. One might, for instance, create a task where the subject is required to gain and remember passwords (for loading semantic working memory) or button press sequences (for loading spatial working memory) to progress through the virtual environment. Such a task would allow the experimenter to vary the memory load (length of the password, for instance), but because the task is integrated into the virtual environment, it would not act as a reminder that the subject is in an experiment and thus break presence. This ensures that the task serves only to load working memory rather than moving the subject’s focus away from the experience of being in a virtual environment to the experience of being in an experiment.

Acknowledgements

I would like to acknowledge the support of the CAVES (Collaborative African Virtual Environments System) project in providing the means for conducting this research.

References

The Sense of Being ‘There’:
a Model for the Space of Presence

Anna Spagnolli, Luciano Gamberini
Dept. of General Psychology, University of Padova, Italy
{luciano.gamberini@unipd.it, anna.spagnolli@unipd.it}

Abstract

With respect to the common definition of presence as the ‘sense of being there’, most studies concentrate on the ‘sense of being’ part of the concept, addressing users’ interior state during the technologically mediated experience. In this work, we would like to emphasize the role of the ‘there’ part, introducing a model for the “space of presence” not as a component but as a defining dimension of presence itself. The paper starts by emphasizing the central role of place in the constitution of presence, and then elaborates on this concept in the light of recent developments in human geography, ethnography and cognitive science. Such space consists of physical, cognitive and cultural resources dynamically structured by the unfolding action and is then very different from physical or even mental spaces. The conclusive section illustrates some of the main implications of the proposed model, that do justice to the richness of the concept of presence and show its potential insightfulness in the study of mediated experiences.

1. Locating presence

As the architect William J. Mitchell provocatively puts it, electronic environments ‘are not just interfaces; we are beginning to live our lives in them’ ([1] p. 31). For some authors, the setting up of a space represents the crucial characteristic of new digital media. Technologies are no longer tools through which we act upon the environment, but environments themselves. From scrolling a document on a two-dimensional sheet, to driving a car on a simulated three-dimensional road, we intervene on a reality that is electronically generated and possesses its own spatial rationale [2]. It is not clear, however, whether such space, reflected also in the terminology (‘cyberspace’, ‘virtual environment’, ‘navigation’) is to be taken literally or metaphorically. On the one side, the tangible consequences, observable events and shared practices emerging from digital environments seem to legitimate them as spaces, worth ethnographic approaches such as those adopted by Christine Hine in her ‘Virtual Ethnography’ of on-line communities [3]. On the other, there is no physical counterpart to these territories and the user is ultimately more likely to be described as located at her desk using the computer than within a digital space. On the one side, digital media are singled out because of their different spatial-temporal characteristics [4]. On the other, it cannot avoid similarities with older technologies that blur any sheer differentiation with older supports [5]. In the area of virtual environments, the issue of digital spaces has a great relevance, where a crucial theme is the user’s sense of being within the simulation. The way in which such feeling is interpreted depends strictly on the way in which virtual spaces are considered: for instance, when they are considered as by-products of energy and information, then ‘virtual presence’ will be taken as a mental illusion [as in 6].

The concept of ‘Presence’ has the advantage of introducing in the study of technologies a reflection on the change in the main coordinates of human experience, spatial and temporal. Being present is tantamount to being somewhere: whenever a person is qualified as ‘present’ above any other attributions, her location is the salient, characterizing feature. In Martin Heidegger’s words, “‘Being-in’ is thus the formal existential expression for the Being of Dasein, which has Being-in-the-world as its essential state’” [7, p.80]

The nature of human presence is then illuminated by a reflection on the constitution of human place and can be extended to see if it holds to include digital spaces as well. Which are the places of human experience? How do technologies contribute to draw them? In the conventional definition of virtual presence as ‘the sense of being there’ [8, 9, 10] the environment is invested with a crucial role: the deictic ‘there’ points directly to the surrounding space and leaves to it any further specification of the users’ psychological experience. While a spatial dimension has been frequently added to theories on presence in virtual reality [11 for a review], it has not be given an actually constitutive force and is merely limited to a geographical surrounding. In the rest of the paper, we will start from the assumption that there is no presence without a place and reflect on the way in which technologies affect both of them.

2. Human places
At the beginning of the 20th century, philosophers and thinkers were struck by the extent to which the human sense of space was affected by cultural changes. The historian Stephen Kern illustrates how writers such as Proust, poets such as Marinetti and directors such as Münsterberg began rendering the psychological space in their work [12]. About in the same period, the concept of ‘place’ made its re-appearance after centuries of disregard, emphasizing an experience of space not reducible to the pure, abstract, homogeneous dimension portrayed by Newtonian physics. ‘Place’ identifies a space conventionally and recognizably associated to meaning on a psychological and cultural basis: it is transfigured by emotional meaning, scanned by different professional wisdoms, rich with valuable resources, modeled by aesthetics canons [13]. It is this kind of ‘space’ that pervades our experience and that, far from limiting our appreciation of space, orients human action[14].

As Edward Casey [15] explains in his essays on Place, place has reappeared when the Cartesian dichotomy between body (res extensa) and mind (res cogitans) has been abandoned. In ontological discourse, the mind was loosing centrality in explaining the psychological experience. Philosophers have started to re-found the mental processes as emerging from the involvement with the environment instead of being pure ideas or universal structures. According to the phenomenological school started by Merleau-Ponty and continued by psychologists such as Mark Johnson [16] basic orientation categories such as up and down, front or back, symmetry or asymmetry would depend on the spatial orientation of the human body. More radically, thinkers such as Heidegger argued that the involvement with the world is the necessary condition of Being. Finally, other scholars such as Foucault studied directly the way in which the spatial structure influences the individual experience and incarnate the social organization.

The Copernican revolution stating the primacy of the relationship with the environment in the human experience rejects the entrapment within the confines of the individual mind, and privileges the public, negotiable experience. Place is neither an objective space (homogeneous and measurable) nor a mental space (pure and abstract). The field of human geography develops a relational concept of space in which space is ‘folded into’ social relations through practical activities’ ([14] p. 769), and neuropsychology support the adoption of action as the reference point for coding spatial information such as objects’ distance or position [17, 18]. Action becomes the criterion according to which place is organized [19,20] coherent with a perspective emphasizing the human involvement with the world. It also introduces a temporal dimension on place, for each activity is a developing process exploiting cues from the past and projecting possibilities on the future.

The main lesson to be derived from this brief interdisciplinary foray is that places are not physical containers of human presence, but the main expression of human presence itself. In the following paragraph, we elaborate on these aspects with respect to technologies, benefiting from a link with the theories on ‘situated action’ and distributed cognition.

3. The space of situated action (Presence in mediated places)

The body has its own modalities to reach out in the space and structure the environment. Pioneer psychologists have detailed the different sensory-motor parameters that organize the animals’ space (e.g. [21]), anthropologists have outlined the characteristics of the human sensory spaces [22] and Environmental Psychologists have investigated the relationship of the human behaviour with the socio-physical space [23] sometimes together with Architects and Engineers [24]. In the model proposed here, body -including the brain, the motor system, the sensory apparatus and other biological mechanisms jointly involved in action - is acknowledged as a crucial parameter for the organization of human experience in the world. While observing the body of a person immersed in a virtual task we will see that her movements accompany action as an essential part of it even though they are not possibly tracked by the virtual system or even necessary to the action itself. In a group of exploratory observations with 4 people engaged in immersive virtual environments of various emotional tone, participants changed their posture independently of the emotional nature of the park but in concomitance with the beginning of certain kinds of actions: they needed to take a more active posture during navigation and preferred to take a more relaxed one while watching a picture [25]. Mediated experiences are often considered free from physical impediments, but the body appears to be an inevitable component in the organization of a complex course of action. Downplaying its role while designing a technology, besides leading to unnatural and unhealthy work conditions, would neglect its capacity to interface our high level cognitive abilities in structuring our presence.

In organizing the space of human action, the body benefits from tools working in intimate relationship with it, which postmodernists would call ‘prostheses’ [26]. Since time began, prostheses have supported human activity in near space (the ‘A’ area in Figure 1), overcoming bodily limitations while manipulating, cleaning, constructing
objects and inevitably has been influencing back the human beings and the spatial-temporal structure of their action (the bidirectional ‘i’ arrow in Figure 1). Today, human possibilities have been modified, amplified and differentiated by technical innovations, further extending the repertoire of human spaces (the area ‘B’ in Figure 1). From this point of view, technologically mediated environments do not represent a discontinuity in the human landscape, but a further instance of a pervasive phenomenon, its peculiarity residing in the specific kind of mediation afforded. In figure 1, the human actor is at the center of two concentric circles; the inner one represents the space reached by the body, possibly incremented with the help of local technical artifacts (area ‘A’); the outer one represents the space reached through Information and Communication Technologies (area ‘B’), extending human presence on a remote physical space.

Activity theory [27, 28] and cultural psychology [29, 30] maintain that any action relies on some kind of tools, not only material such as arms, but also symbolic such as language or plans. To inhabit a space through action, cognitive-cultural artefacts are needed (language, mathematics, norms, preferences) to integrate the physical dimension with meaning, expectations, implications. They serve as tools, schematically represented by the objects C1,2,3 in Fig. 1. The digital space is inhabited according to what it entails in terms of memory and culture, myths, fears, joy. In a study [31, 32], participants visited a virtual museum presenting a selection of Lang pictures from the International Affective Picture System (IAPS, [33]) divided into positive, negative and neutral ones; the analysis of the human-interface events (observation time, distance from the picture during the observation) automatically recorded during the interaction revealed that users’ spatial behaviour changed according to the emotional nature of the picture observed. In treating virtual stimuli as threatening or pleasing, participants were influenced by streams of sensory data (real and virtual) as suggested by Slater and Steed [34]. It is worth stressing again that human place is created by action, which works as a catalyst attracting the physical, cultural and cognitive resources orienting the relationship with the environment. Such resources are selected for their relevance to the ongoing action out of the complete cognitive and cultural endowment of an actor, represented by the outmost semi-sphere ‘S’.

Fig 1: The space of mediated action, consisting of physical (A, B) and cognitive-cultural resources relevant (G) and possibly relevant (L) among those available (S)

The actor’s presence is distributed on a place represented by the cone G in Fig. 1. The cultural-cognitive resources are rooted into the physical realm (see for example [35]); conversely, the physical environment needs to be culturally and cognitively shaped in order to take on its familiar appearance (see for instance [36, 37]). The fact that the place is structured according to the action performed is shown in a study where an outbreak of fire turns an the task from exploration to escape [38]. As shown by a deep human-interface events analysis and by a qualitative video analysis supported by split-screen technique [39], subjects transformed their movement style to cope with the new situation in a manner similar to the one that many people would have used in a real fire
emergency. Participants transformed their joystick and HMD movements in quality and quantity, changing the trade-off between speed and precision in the light of a new activity. Place is then not a mere portion of space, but the heterogeneous ensemble of resources that converge on action and have spatio-temporal features.

The resources excluded from place are not equivalent, because some of them are possibly connected with the ongoing action (D_{1,2,3} in the figure, semisphere L). They create a grey area of discarded alternatives and relevant objects that human cognition cannot consider because of its own limitations. Notice also that the place in the picture is simplified: if the place of a person at any single moment were to be depicted, then the illustration would show a complex chaining of partially overlapping cones. The cones overlap in the space created by the relationship; where they do not, there is an asymmetry emerging from that local relationship, due to different resources, belongings, participations. In any case, the cone is not an individual exoskeleton with a fixed extension carried around by individuals across all circumstances; it is defined by the interaction.

4. Implications

New technologies, especially digital interactive media, accentuate the flexibility of the cone. The modes of presence multiply because human action expands on unprecedented spatio-temporal configurations and because of an extensive use of the cognitive-cultural dimension. The definition of presence as distributed on the heterogeneous ensemble of resources that converge on action has important consequences on the way in which the human interaction with technologies can be conceptualized. Some of the most important are described below.

   a) Technologies do create virtual places: such places do not need to occupy a material slice of territory, they need actions with peculiar spatial-temporal configuration. To outline the characteristics of this virtual places, one has to describe the actions that are possible there and the cultural and physical resources available. It also won’t be necessary for such resources to be located in close continuity in the same stretch of land, but to be part of the same course of (inter)action. So people talking on a mobile phone are inhabiting the same communicative place supported by mobile phone technology and characterized by physical and cognitive-cultural resources.

   b) When the gamut of potentially accessible places and people is spreading, it is not only the spatial and temporal distance that is manipulated, but the kind of participation, access and actions afforded. Mediated places will have own mode of presence, including roles, competence, benefits. They are ways of locating ourselves in human places as such they are never neutral, but positioned vis-a-vis other alternatives.

   c) The concept of presence acquires a great force representing an active process, not a state. If place is related to action and possible actions, then human beings can be engaged in several places at the same time and change them dynamically.

   d) Presence is always relational: being in a place always means being involved in peculiar ways with specific objects. As a consequence, presence is not confined within the realms of the individual mind, but is shared, espoused, communicative. It can be recognized by an observer, both in real life and in research.

   e) For the acting subject, digital spaces are simultaneously physical, cognitive and cultural [40], since none of such dimensions can exclude the other. The position of the body continues to influence the interaction for its possible exposure to other events besides the navigation, for the structure it imposes on the interaction, for the change of state (fatigue, ...), for the practices available.

   f) Finally, when the concept of presence is used, one must specify the place to which it refers. Being present does not mean anything without specifying where one is present. And again, such specification may not be vaguely based on rough categories, but should be defined with respect to specific places for action.

5. Conclusions

The concept of ‘presence’ provides a chance of reflecting on the effect that technologies have on the basic coordinates of our experience. The model proposed in this paper has two main advantages for what entails virtual environments.

First, it can be applied to presence in ‘mixed’ real and virtual situations that are not accountable for in terms of dichotomised models such as the ‘Break In Presence’ (BIP) model by Slater & Steed [34], as shown by Spagnolli and Gamberini [41, 42] and highlighted by Slater, Brogni and Steed [43] and Brogni, Slater, Steed [44], where a binary response is deliberately adopted in order to provide a viable solution to the Presence measurement problem. Among other things, mixed responses are at the basis of the most significant virtual environment applications in clinical psychology. Considering, as suggested by Riva [45], the applications for which controlled trials with at least 10 patients were run, it appears that all situations include a complex relational setting with therapist, patients and virtual words, located in a mixed (real and virtual) environment. Presence departs from the virtual environment not when it departs from the digital perceptual stimuli per se, since they do not define a place, but when the ongoing
activity is no longer primarily mediated by the virtual environment and oriented to it.

Second, the role of perception in presence is balanced by the role of other cognitive, physical and cultural resources within the well known framework of situated action [20] and distributed cognition [36]. The consideration of the complex context in which a technology is used borrows from central concerns in design and ergonomics studies, where a greater consideration of the users’ environment and activity scenario is pursued [46, 47, 48]. In the model depicted here, which needs to be investigated and tested further, presence becomes a central, inclusive concept in the design, development and evaluation cycle of technologies and will give new artefacts a richer human dimension.

Acknowledgements

The authors would like to thank the two anonymous referees for their extremely valuable suggestions.

References.


Social Presence and Interpersonal Trust in Avatar-Based, Collaborative Net-Communications

Gary Bente, Sabine Rüggenberg, Nicole, C. Krämer
University of Cologne
{bente@uni-koeln.de, sabine.rueggenberg@uni-koeln.de, nicole.kraemer@uni-koeln.de}

Abstract
The paper deals with social presence effects of embodiment in collaborative net-communications. Social presence was measured by an extended questionnaire including an item set on emotional closeness of the partners. Principal component analysis revealed a consistent structure with the five dimensions: co-presence, closeness, comprehension, contagion and coordination. In addition a measure for interpersonal trust was applied which differentiates between cognitively and affectively based trust. Correlation analysis showed significant relations between closeness and affectively based trust. Results of a media comparison including face-to-face interactions, phone, chat and avatar-based net-communications point to distinct response patterns for the different social presence factors and interpersonal trust. Only with respect to co-presence the avatar-based interactions led to higher scores than phone or chat interactions. Data are discussed with respect to multi-dimensional conceptualisation and multi-level measurement of social presence.

Keywords--- embodiment, avatars, virtual environments, social presence, collaborative work

1. Avatars in net-based collaboration: research objectives

As early as in the 1980ies Jaron Larnier, founder of the legendary VPL Inc. introduced a virtual environment, the so called RB2 system (Reality Built for Two), which should serve as a telecommunication medium, allowing two people to meet in a shared virtual world and to interact with the virtual objects and with virtual representations of each other. Later he commented on the use of avatars within virtual environments: „One intriguing implication of virtual reality is that participants must be able to see representations of one another, often known as avatars. Although the computer power of the day limited our early avatars to extremely simple, cartoonish computer graphics that only roughly approximated the faces of users, they nonetheless transmitted the motions of their hosts faithfully and thereby conveyed a sense of presence, emotion and locus of interest” [1]. Based on the technological developments of the last decade, recently various avatar communication systems have been introduced [2, 3, 4, 5, 6] which overcome many restrictions of the early days. As each of these developments focused on different technical issues as well as on distinct applications, the resulting systems differ remarkably in some of their basic characteristics, as there are the bandwidth of conveyed behavioral dimensions, the immersiveness of the display technology, the immediacy of transmission of behavioral cues, etc. While early stage research was more concerned with the technical components which create immersion and presence recent studies focus on the applicability of avatar-platforms in different contexts. Evidently, the possibility to detect, store, transmit and thus influence nonverbal behaviour like gaze, movement, gestures and postures in real time made a strong argument for the usage of avatar platforms in communication and perception research [4, 7].

A more recent impetus for the application and systematic evaluation of avatars came from the areas of computer supported cooperative work (CSCW) and computer supported collaborative learning (CSCL). Based on theoretical concepts like common ground [8, 9, 10, 11] and group awareness [12, 13, 14] researchers in these fields expect significant improvement of collaborative learning and problem solving by including nonverbal channels into the net-based communication...
process. Embodiment by means of avatars appeals to these scholars as a possibility to facilitate the net-based communication process, to improve social relationships, to positively influence group processes and cohesiveness and to create higher levels of interpersonal trust. Interestingly, a closer look at the concepts of “common ground” and “group awareness” reveals a wide conceptual overlap with the concept of social presence. So besides the challenge to integrate user friendly and efficient avatar systems into collaborative environments there is a theoretical challenge to model and empirically verify the specific communicative functions and socio-emotional effects of embodiment. As previous work has shown [15] questionnaire data are not sufficient to catch the subtle and transient low level effects inherent in nonverbal communication, like e.g. the use of specific gestures, the timing of mutual gaze, the coordination of verbal and nonverbal activities etc. With respect to the conceptual integration mentioned above it will thus be necessary to provide behavioral measures of social presence, common ground, group awareness and interpersonal trust in addition to subjective verbal reports to figure out to which degree the variables depend on each other or whether they have a common behavioral base. Last but not least it has to be shown that the use of avatars causes desirable effects beyond the experience of social presence and group awareness in the sense of task performance, outcome and efficiency of collaborative work. These various aspects describe the objectives of our current research project which is funded by the DFG (German Research Foundation) within the special interest program “Net-based Knowledge Communication in Groups”. In the following, we will present a pilot study conducted within this project aiming at the validation of questionnaire instruments on social presence and interpersonal trust in a media comparison paradigm.

2. Social presence and interpersonal trust in avatar environments: a pilot study

Major objectives of the pilot study were (1) the development and evaluation of a questionnaire for measuring social presence (SP) and interpersonal trust (IT) in avatar-based collaborative net-communications and (2) a first tentative media comparison focusing on the different-tial effects of the ABC desk as compared to face-to-face-communication (ftf) as well as audio-based and text-based net-communication. Following the suggestions of Biocca et al. [16] the development of the SP questionnaire was based on an integrated approach. They define 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” (p. 2). The degree varies from the peripheral sense of spatial co-presence of the other to progressively higher levels of social presence. Those are characterized by a deeper sense of psychological involvement, access, and connection to the intentional, cognitive, and affective states of the other. Higher levels include a sense of behavioral engagement leading to actions that are perceived as linked, reactive, and interdependent. As mentioned above, we also aim at identifying criteria for social presence in terms of correlates that are said to be crucial with regard to the outcome of mediated collaborations. One of the aspects mentioned to be important in net-based interactions is trust [17, 18, 19]. The well-known assumption that „trust needs touch” [20, 21] offers the question by which aspects the trust-building process an be facilitated even in net-based communication settings where getting in “touch” is at least in a literal sense impossible. In this context the possible relation between the perceived social presence of the communication partner offered by the characteristics of a communication medium and the perceived interpersonal trust [22] between participants has been discussed. However, no integrative analyses of these two aspects can be found in the relevant literature. Thus, we aim at investigating the relation of social presence and interpersonal trust – both on the level of subjective experience.

2.1. Method

2.1.1. The ABC-desk: An avatar-based collaborative desktop environment

ABC desk (avatar-based collaborative desktop environment) developed at the University of Cologne is conceptualized as a low immersive open desktop system (no HMD, or shutter-display) focusing on the real time transmission of a broad range of nonverbal behavior during net-based communication and collaboration [2, 3]. The setting very much resembles a video conference situation, where the interaction partner is just visible on a screen or in a screen window. In contrast to the video conference situation the avatar-system allows for masking the identity of participants, for the recording and the experimental control of the behavioral data. Moreover, interlocutors in an avatar-based virtual environment can refer to a shared virtual object world, which means that they can contingently handle the same objects e.g. in a collaborative task. In particular, the ABC desk was constructed to allow for (1) the real time interaction of up to three interlocutors including nonverbal signals like head movements, body movements, gestures and eye movement, (2) the experimental variation of the visual appearance of the interlocutors, (3) the online filtering of behavioral cues, (4) the recording of verbal and nonverbal behavior, (5) the interactive and/or algorithmic modification of
behavior protocols, and (6) the offline rendering and replay of stored movement data. Non-verbal behaviour is detected by means of Cybergloves, Polhemus trackers and a high resolution eye-tracking system, which we developed for this purpose. Data are transmitted via Intranet (TCP-IP). Animation is performed on the target computer by means of an AVI-CODEC developed for this purpose. The CODEC transforms angular data into 3D-animations and renders the movement of a low-times a second there is no speed problem at all. For this study the ABC desk was used as a Virtual Video Conference (VVC): The interaction partner’s avatar is presented on a separate monitor in full-screen mode (see figure 1). The virtual representatives were high resolution avatars (see figure 2) which are rendered by a commercial 3D animation tool (Kaydara Filmbox©) accessing a continuous data stream which is also provided by AVI codec. Besides the VVC two further mediated communication settings were used in this study: a) text-based communication supported by a chat tool and b) audio communication.

**Figure 1:** Base version of the avatar-based collaborative desktop environment (screen shot during calibration phase showing to the inter-locutors their own avatar)

**2.1.2. Measurement of social presence and interpersonal trust** For the measurement of social presence an initial set of 58 five-point Likert scale items was created by translating those used by Biocca et al. [17] into German. These were extended by further newly created items, pertaining to the perceived closeness of the participants during the interaction as discussed for example by Kumar and Benbasat [23] or Tu [24]. Examples of the items are given in table 1.

Referring to the literature a 20 item questionnaire was created for the measurement of interpersonal trust, which should particularly be able to differentiate between the two components (1) **cognitively-based trust** (competence of the other participant) and (2) **affectively-based trust** (trustworthiness of the other participant) as described by Kanawattanachi und Yoo [18].

**Figure 2:** High resolution avatars

**2.1.3. Participants and Design** 48 (24 male and 24 female) student participants took part in the pilot study. Matching for sex the participants were randomly assigned to four different communication settings: chat (text), phone (audio), avatar (audio-video), and face-to-face (full bandwidth). The interaction partner was always a confederate of the experimenter. The participants were involved in a collaborative problem solving task, where they had to select a potential employee for a certain job. The interactions were time limited and lasted between 5 to 10 minutes. After ending the communication the participants had to answer the questionnaires on social presence and interpersonal trust.

**2.2. Results**

**2.2.1. Dimensions of social presence** Principal component analysis of the social presence items (VARIMAX rotation) resulted in a five component solution explaining 63% of the variance (see table 1 for a list of factors and item examples). Four of these factors are in accordance with the social presence dimensions conceptualised by Biocca et al. [16]. The factors were named **co-presence**, **comprehension**, **connectedness** and **contingency**. One additional factor emerged which exclusively contained the added items on perceived acquaintance and intimacy of the interaction partners. The component was named **closeness**. The labels of the principal components are listed in table 1 together with the labels suggested by Biocca et al. [16] and the psychological dimensions they refer to, as there are: spatial, social, cognitive, emotional and behavioural relatedness. Internal consistency values were good to excellent for all scales (see table 1 for Cronbach’s alpha). Also, the scales showed consistent reciprocity, i.e. the item pairs asking for one’s own and the partner’s experience always loaded in the same direction on the same factor.
2.2.2. Dimensions of interpersonal trust

Principal component analysis of the 20 item questionnaire on interpersonal trust (VARIMAX rotation) issued a two factor solution, explaining 65% of the variance. As expected, the factors clearly reflect the conceptual differentiation between cognitively and affectively based trust. Table 2 shows item examples and the results of internal consistency tests which were excellent for both factors.

Table 2: Factor structure and internal consistencies of interpersonal trust scales

<table>
<thead>
<tr>
<th>Scale</th>
<th>Item amount and example</th>
<th>Cronbach’s Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>cognitively based trust</td>
<td>10 items</td>
<td>.918</td>
</tr>
<tr>
<td>(I think, that I could also rely on his/her competences when facing challenging tasks.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>affectively based trust</td>
<td>8 items</td>
<td>.901</td>
</tr>
<tr>
<td>(I would confide even private information to my partner.)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2.2.4. Differential media effects on social presence and interpersonal trust

The assumption that the consideration of social presence as a multidimensional construct could be useful for further investigations was confirmed by cross media comparisons. Table 3 shows the results of ANOVAs and pair-wise comparisons (ex-post Scheffé tests). The different communication groups (text, audio, avatar, and ftf) varied with regard to the level of perceived social presence showing a different pattern of presence aspects for each setting. The findings did not confirm a general superiority of the avatar-based communication setting in facilitating the social presence experience of the participants. Certainly, there is no linear relation between bandwidth and social presence as could be expected e.g. on the base of media richness theories or cues filtered out models of computer mediated communication [25, 26]. Figure 4 visualizes the direction of the significant results. Only for the aspect of co-presence the results met the assumption that the usage of an avatar will stimulate the presence experience in a better way than text- or audio-based technologies (see figure 4a). However, this result has to be interpreted very carefully as the avatar-effect is close to zero. What can be said is that in contrast to ftf, audio and text have negative effects and the avatar is neutral. While co-presence can be considered an individual non-evaluative judgement closeness is constituted by an interpersonal and emotional experience. Here, embodiment might have positive as well as negative effects depending on the nonverbal behavior of the vis-à-vis (one might see somebody as a stranger because he/she is unknown or we may judge the behavior we observe as strange). It is thus quite plausible
Table 3: Media differences in social presence and interpersonal trust (ANOVA of average factor score)

<table>
<thead>
<tr>
<th></th>
<th>1 FtF</th>
<th>2 Avatar</th>
<th>3 Phone</th>
<th>4 Chat</th>
<th>F</th>
<th>p</th>
<th>Pairwise comparison (Scheffé Test)</th>
</tr>
</thead>
<tbody>
<tr>
<td>co-presence</td>
<td>1.37</td>
<td>-.014</td>
<td>-.601</td>
<td>-.749</td>
<td>36.07</td>
<td>.000</td>
<td>1&gt;2, 1&gt;3, 1&gt;4 (p = .000)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2&gt;4 (p = .023)</td>
</tr>
<tr>
<td>closeness</td>
<td>.145</td>
<td>-.119</td>
<td>.589</td>
<td>-.615</td>
<td>3.54</td>
<td>.022</td>
<td>3&gt;4 (p = .027)</td>
</tr>
<tr>
<td>comprehension</td>
<td>-.184</td>
<td>1.04</td>
<td>-.667</td>
<td>-.190</td>
<td>10.12</td>
<td>.000</td>
<td>2&gt;1, 2&gt;4 (p = .006)</td>
</tr>
<tr>
<td>(inverted scale)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2&gt;3 (p = .000)</td>
</tr>
<tr>
<td>contagion</td>
<td>.342</td>
<td>-.059</td>
<td>-.190</td>
<td>-.093</td>
<td>.654</td>
<td>.590</td>
<td>-</td>
</tr>
<tr>
<td>contingency</td>
<td>.014</td>
<td>.206</td>
<td>-.209</td>
<td>-.011</td>
<td>.33</td>
<td>.803</td>
<td>-</td>
</tr>
<tr>
<td>cognitively</td>
<td>.509</td>
<td>-.912</td>
<td>.667</td>
<td>-.264</td>
<td>10.19</td>
<td>.000</td>
<td>2&lt;1 (p = .001)</td>
</tr>
<tr>
<td>based trust</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2&lt;3 (p = .000)</td>
</tr>
<tr>
<td>affectively</td>
<td>.206</td>
<td>-.162</td>
<td>.286</td>
<td>-.330</td>
<td>1.04</td>
<td>.384</td>
<td>-</td>
</tr>
<tr>
<td>based trust</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-</td>
</tr>
</tbody>
</table>

Figure 4: Significant media effects in the factors (a) co-presence, (b) closeness, (c) comprehension and (d) cognitively based trust
that both modalities which include the visual channel (ftf and avatar) show mean values close to zero (averaging positive and negative effects). As expected, however, phone did much better than chat in creating the feeling of acquaintance and intimacy (see figure 4b).

Comprehension indicates an evaluative judgement with respect to cognitive rather than emotional relatedness. Again audio does best in creating the experience of attentional engagement (see figure 4c).

Direct comparisons however show that there is only one significantly deviant condition, the avatar condition, which produces negative responses in mutual understanding (positive values). The reasons for this can only be speculated about. Certainly, the avatar representations are something new and uncommon to the participants and might focus attention more on the visual channel than usual, thus creating additional mental workload and distraction from the content of the dialogue. Especially technical deficiencies might be seen as a possible cause for such a misled attention. The results on cognitively based trust confirm this result as they are mostly consistent with the comprehension data (see figure 4d). Again the audio condition shows the best scores while the avatar condition generated the lowest levels of cognitively based trust. Again irritating aspects of the new communication technology might be responsible for these negative effects.

3. Discussion and conclusion

The questionnaire instruments introduced for the measurement of social presence and interpersonal trust instruments could prove their internal factorial validity as well as the external validity in distinguishing effects of different communication media and ftf communications. The extended item set as compared to Biocca et al. [16] constituted a new factor with highest variance explanation which was called closeness. The factor represents feelings of interpersonal immediacy and emotional closeness to the vis-a-vis. Only the traditional co-presence factor however revealed media effects as expected, showing significantly stronger co-presence effects for the avatar condition relative to phone and chat. As the co-presence values for the avatar-based encounters however were close to zero the data can hardly interpreted as a positive effect in itself. Moreover, one could also assume that the nonverbal information did not receive the persistent attention and did not become salient in the expected way. This result strongly indicates the necessity of additional behavioral measures for social presence. For example measures of gaze direction on one hand and measures of nonverbal activity could be analysed to point out specific structural differences or similarities between ftf and avatar-based communication.

Figure 5 exemplarily shows such an integrated analysis for two interlocutors in a three-minute sequence of an avatar-mediated collaborative interaction. As a global measure of nonverbal activity the upper curves in the two graphs show the movement complexity (number of body parts in action) of each partner. The bar-charts underneath show the gaze pattern, where the black areas indicate a gaze towards the avatar window in the desktop workbench (see figure 3a). The first minute (up to dotted line) was spent in chatting and the last two minutes in collaborative problem solving.

The graph clearly indicates a strong correlation between nonverbal activity and visual attention which is similar to ftf situations. This kind of correlations could serve as benchmark values when testing social presence effects of emergent communication technologies. The inclusion of behavioral measures in general thus seems to be of high relevance for a further conceptual clarification of social presence leading to more objective definitions of the various dimensions and also to a differentiation with respect to related concepts like common ground, group awareness or interpersonal trust. In any case it has to be prevented that the concept of social presence is expanded to a degree that all known aspects of interpersonal communication and group dynamics are included. Such a meta-concept would have no discriminative value at all and would immediately lose its potential to relate subjective media experiences to technical parameters.
Acknowledgements

This research was supported by grant BE 1745/4-1 within the Special Priority Program “Net-based knowledge communication in groups” from the Deutsche Forschungsgemeinschaft (DFG, German Research Association).

References


Quantitative evaluation of sensation of presence in viewing the "Super Hi-Vision" 4000-scanning-line wide-field video system

Masaki Emoto, Kenichiro Masaoka, Masayuki Sugawara, Fumio Okano

Abstract
Purpose. To decide the system specifications for a future broadcasting system with a stronger sensation of presence, it is essential to understand the effect of viewing angle on this sensation. This study aims to establish a clear quantitative relation between the viewing angle of the displayed images and the viewer’s sensation while watching them.

Methods. We have developed a 4000-scanning-line video system with a wide field of view. It is called Super Hi-Vision, and it enables images to be presented with enough resolution and brightness on an almost flat screen. We measured viewers' body sway while they were viewing still images presented by the system with six variations of visual angle, under the assumption that the smaller the difference between the real world and the scene presented by video systems becomes, the smaller the difference in the response of human equilibrium would be between viewing the real world and viewing the scene presented by video system. The total distance of body sway (hereafter called total body sway) was calculated as an index of the response.

Results. The total body sway shortened as the field of view increased, and the effect tended to saturate over 76.9 arcdeg.

Keywords— Sensation of presence, Field of view, Posture control, Equilibrium, Body sway, Super Hi-Vision, 4000-scanning-lines

1. Introduction

Technological advances have made it possible for us to construct and test video systems that have a very wide visual field, very high resolution, and the ability to present stereoscopic images. One goal of constructing these systems has been to make viewers feel as if they are in the space displayed by the video system, i.e., to convey to viewers the sensation of presence. We (Japan Broadcasting Corporation, NHK) have been studying audio-visual systems with a wide field of view and three-dimensional audio with the goal of making one the center of a future broadcasting system. A wide field of view and three-dimensional audio tend to increase the sensation of presence, but the quantitative evaluation of the sensation has not been achieved. It is essential to establish methods that can quantitatively evaluate the sensation to design a future broadcasting system. We have developed a 4000-scanning-line video system as an experimental apparatus to evaluate the wide field effect on the sensation. It is called Super Hi-Vision, which consists of a high-resolution camera, a wide field-of-view display, a 22.2 multi channel audio system, and a disc recorder. The display system enables wide field images up to 100 arc-degrees to be presented with enough resolution and brightness on an almost flat screen.

The first prioritized method for the quantitative evaluation of the sensation would be a subjective one. However, the subjective evaluation of presence is very difficult, because it is highly subjective with significant ambiguity between viewers as to what constitutes it. Questionnaires on the sensation have been conducted [1][2][3], but their validity is not clear for our objective. An alternative psychophysical strategy without direct subjective evaluation was tried, whose index was the induced tilt angle of the subjective vertical line after 15 seconds of adaptation to slanted images [4]. The amount of induced angle saturated at more than 80–90 arcdeg. for the displayed images.

Another objective index that has been tried is the response of a subject’s posture control while standing. The simple act of standing on two feet is a dynamic one maintaining equilibrium, and it has three major feedback inputs: somatosensory, vestibular, and visual. The change in only the visual input as one’s eyes open and close affects the human posture control system [5][6]. Therefore, it is possible that the sensation of presence felt from presented images also affects posture control through visual input. Many studies have quantified the effect of visual input on the posture control by measuring the body sway in response to a variety of visual input stimuli [7-10]. In almost all of these studies, the magnitude of the vection, which is the sensation of self-motion induced by viewing motion images, was measured as an index of the sensation of presence, and was related to the viewing angle or a portion of the retina. Brandt et al. pointed out that peripheral vision plays an important role in vection [7]. This suggests that
wide field images including the peripheral visual field affect the viewers' sensation of presence. The result of another study on the effect of reciprocating rotated stereoscopic and monoscopic images on the viewers' body sway showed that the effect saturates over 90 arcdeg. of field of view, and was more effective when the subjects viewed stereoscopic images [8]. These studies showed that the motion of picture captured by peripheral vision significantly contributed to vection. This is consistent with the characteristics of the peripheral vision, whose resolution is lower than central vision but whose sensitivity to the motion is higher than central vision [9]. The large area of peripheral vision might be responsible for the contribution.

Moreover, a comparison of central and peripheral vision areas of equal size showed that central vision played a more important role in stabilizing posture control than peripheral vision [10]. The contribution of peripheral vision to posture control can be minimized by using still images. In TV programs, motion pictures showing enough shaking to induce vection are rare. Therefore, an evaluation of presence by using only vection while viewing television is not sufficient; presence must also be evaluated using still pictures.

In this study, we used still images to investigate the relationship between viewing angle and the response of the human equilibrium system in order to eliminate the vection-like effects of motion-sensitive peripheral vision as much as possible. Our assumption to be tested was that the smaller the difference between the real world and a scene presented by a video system becomes, the smaller the difference in the responses of human equilibrium would be to viewing the real world and the presented scene. Most of the previous studies on the relationship between viewing angle and human equilibrium have projected images on the inner surface of a dome screen that could present images up to 180 arcdeg. It is useful to evaluate the relationship using a flat screen, because TVs have flat or almost flat screens. In fact, the video system we used in the experiments has 4000 scanning lines on a flat screen [11], which provides enough resolution and brightness over a 100 arcdeg. field of view. Thus, we evaluated the relationship between the viewing angles of up to 100 arcdeg. and the response of the human equilibrium system. The field of view was varied over six angles, from approximately 30 arcdeg. to approximately 100 arcdeg., which is the normal viewing angle of the 4000 scanning line system.

Strictly speaking, we measured the excursion of the intersection point of the action line of the vertical supportive force with the supporting surface, which is different from the projection point of the mass center of body to the surface of the force platform [12][13]. They are almost equivalent only when the subject is standing still on two legs. Therefore, we use the word "body sway" in this paper to indicate the excursion of the intersection point of the action line of the vertical supportive force with the supporting surface.

2. Methods

2.1. Viewing conditions and test images

The field of view was varied over six angles by shrinking oversampled images to sizes corresponding to the angles. Table 1 shows field of view, number of scan lines, and number of pixels in a scan line. The aspect ratio, which is the ratio of the horizontal and vertical size of the images, was the same (16:9) for all images. The smallest images (approximately 30-arcdeg. horizontal field of view) corresponded to HDTV (High-definition Television, or Hi-Vision in Japan) and the biggest ones corresponded to the 100-arcdeg. 4000 scanning line system. The viewing

<table>
<thead>
<tr>
<th>Field of view (arc degree)</th>
<th>100</th>
<th>93.3</th>
<th>87.3</th>
<th>76.9</th>
<th>61.6</th>
<th>33.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scanning lines</td>
<td>4000</td>
<td>3555</td>
<td>3200</td>
<td>2666</td>
<td>2000</td>
<td>1000</td>
</tr>
<tr>
<td>Pixels per line</td>
<td>7110</td>
<td>6320</td>
<td>5688</td>
<td>4740</td>
<td>3555</td>
<td>1777</td>
</tr>
</tbody>
</table>

Figure 1. Still images
distance was 3.1 meters, at which viewers with a visual acuity of 0 in LogMAR (minimum angle of resolution) cannot resolve or perceive scan lines on the display surface. The alternative to varying the visual angle is varying the viewing distance. This was not used because viewing distance affects body sway [14]. Three landscape pictures (Figure 1) were taken on reversal film (Fujifilm Fujichrome Velvia100F ISO100) with a 4 by 5 inch still camera (Cambo wide ds) mounting a wide-field lens (Schneider Super-Anguron 5.7/47LX). The camera was set at a height of 1.5 meters with its axis horizontal. The developed films were oversampled to 14220 pel to 8000-line digital data with a drum scanner (Heidelberg DC3900) and trimmed to images with a 16:9 aspect ratio. Shrunken images (see Table 1) were made by image processing while keeping the ratios of upper and lower and left and right hemispheres identical. The shrink ratios were determined to integer to minimize the degeneration in image quality.

2.2. Apparatus

Images were presented on an approximately flat screen with sufficient resolution and peak brightness (approx. 50 cd/m²). Table 2 shows the specifications of the video system part of Super Hi-Vision. First, the horizontal visual angle of the system was decided to be 100arcdeg. from the results of a past study that evaluated the sensation of presence by using the tilt angle of the subjective vertical line induced by slanted images [4]. Secondly, the resolution of the system was decided to be 7680pel*4096line for viewers with a visual acuity of 0 in LogMAR to be unable to resolve scan lines on the display surface at 3.1m viewing distance. However, any display or video monitor with 7680*4096-pixels had not been developed, but 1.7-inch LCoS (Liquid Crystal on Silicon) panel with 3840*2048-pixels was available. Therefore, we use two 1.7-inch LCoS panels with 3840*2048-pixels in order to obtain the resolution for green channel in the first projection unit, and one LCoS panel for red and blue channel respectively in the second projection unit [15]. Figure 2 shows the projection units. The improvement of resolution of green channel is more effective than other channels, because the contribution of green channel to brightness is larger than other channels. The relative positioning of the two panels for green must be accurately offset by 0.5 pixel [16]. The optical output of the display is approximately 5000 lumen, resulting in peak brightness on a 320-inch screen of about 50 cd/m² (screen gain 0.85, or about 40 cd/m² with a 450-inch screen whose screen gain is 1.5). Because there are two projection units for dual-G and R/B, the images from these units are not projected the same position on the screen without any correction. Therefore a convergence error correction scheme was developed to convert the red and blue images so that the convergence error on the screen is corrected.

Viewers' body sways were measured with a force platform (Nihon denki sanei 1G06). The platform had two outputs of X and Y, which corresponded to lateral and sagittal body sway respectively. When the amplitude of the platform output was 0.1 volt, the eccentricity of the body sway was 1 cm. These outputs were filtered with a pre-filter (NF Corp. Multifunction filter 3611), and then captured with a PC and an analog/digital converter (Interface CBI-3133A, 12-bit, sampling frequency: 120 Hz)

<table>
<thead>
<tr>
<th>Table 2 Specifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pixel number</td>
</tr>
<tr>
<td>Scanning</td>
</tr>
<tr>
<td>Projector</td>
</tr>
<tr>
<td>Screen size</td>
</tr>
<tr>
<td>Screen gain</td>
</tr>
</tbody>
</table>

Figure 2. Dual green projection unit (lower) and red/blue projection unit (upper).

2.3. Procedure

Viewers stood for 120 seconds on the force platform with the inner sides of their feet in contact, which called the Romberg foot position. Body sway data for 120 seconds were captured, but only the 30 sec-90 sec data were processed to avoid any instability at the start of the period. The total body sway for 60 sec was calculated. The difference in total body sway subtracts the total sway for each viewing angle from the sway for 33.2 arcdeg. were evaluated.
2.4. Test subjects

Twenty healthy adults, 5 males and 15 females (age, mean 32.6 years; range, 24 to 50) who had normal posture control participated as test subjects, after having provided informed consent. Their heights ranged from 151 cm to 180 cm and had a mean of 163.2 cm. Their visual acuity ranged from 0.2 to -0.3 in LogMAR. They were instructed to view the center part of the images at the same height as their eyes while relaxing. This was to avoid the effects of eye position on body sway.

3. Results

Figure 3 shows the excursion of body sway. The x-axis is the lateral sway component, and the y-axis is the sagittal sway component. Figure 4 shows the mean difference of the 20 subjects in total body sway subtracts the total sway for each viewing angle from the sway for 33.2 arcdeg. plotted against field of view. The wider the field of view is, the shorter the total body sway becomes. A repeated-measurement ANOVA with two factors (picture and fields of view) showed that Mauchly's assumption of sphericity was assumed in the factor "picture" (p=0.708) but not in the factor "field of view" (p=0.009) or their interaction (p=0.012), and the main factor of picture (p=0.615) and their interaction (p=0.939, Greenhouse-Geisser epsilon=0.484) were not significant, but the main factor of field of view (p=0.046, Greenhouse-Geisser epsilon=0.595) was significant. This significant decrease in the total body sway with field of view suggested that viewing wide-field images stabilized human equilibrium. To test whether the stabilizing effect saturates or not with increasing field of view, Helmert contrasts within subjects were performed. The results showed a significant difference between 33.2 arcdeg. and more than 61.6 arcdeg. (p=0.014) and showed a tendency, than 76.9 arcdeg. (p=0.071). There was no difference above 76.9 arcdeg. (p=0.476, 0.705, 0.773). This suggested that the human equilibrium system stabilizes as the field of view of still images increases and that the images effect on equilibrium tends to saturate over 76.9 arcdeg. of field of view.

4. Discussion

The total body sway results suggested that still-image field of views of more than 76.9 arcdeg. have little effect on human equilibrium. Hatada et al. reported that the induced tilt angle of the subjective vertical line after 15 seconds' adaptation to slanted images saturated at 80~90 arcdeg. [4]. They used two scenery images and one geometric image (grating pattern). Their results using an open landscape picture similar to our landscape pictures are consistent with our results showing that the field of view's effect on the induced tilt angle of subjective vertical line saturates over approx. 80 arcdeg. The major differences in the experimental conditions between their study and ours are screen shape (dome vs. flat), picture shape (circular vs. rectangular), evaluation index (induced tilt angle of subjective vertical line vs. total body sway), viewing distance (0.85 m vs. 3.15 m), and number of subjects (4 vs. 20). While there are some differences in the experimental conditions between the two studies, their consistency suggests that the total body sway can be an index of the
sensation of presence. These results suggest that we can perceive the maximum sensation from viewing scenery images if the field of view is 80-100 arcdeg or more.

4. Conclusions

To evaluate the sensation of presence while viewing a wide-field video system, we studied the effect of viewing angle on the response of the human equilibrium system. The study was conducted using a 4000-scanning-line video system with a flat screen. The results showed that the effect saturates over 80-100 arcdeg field of view, which is consistent with the results of a past study that evaluated the sensation by using the tilt angle of the subjective vertical line induced by slanted images, and they verify our assumption that the smaller the difference between the real world and the scene presented by video systems becomes, the smaller the difference in the response of human equilibrium system becomes. These results suggest that it is desirable for a wide-field video system to have a 80-100 arcdeg viewing angle to convey the full sensation of presence.

Further studies on the sensation of presence in viewing motion pictures and on the assessment of the sensation using both subjective and objective indexes will be needed before we can arrive at desirable specifications for a wide-field video system.

References

Relaxation Island: virtual, and really relaxing

Freeman, J.¹, Lessiter, J.¹, Keogh, E.², Bond, F.W.¹, & Chapman, K.¹
¹Department of Psychology, Goldsmiths College, University of London; ²Department of Psychology, University of Bath
{J.Freeman@gold.ac.uk, J.Lessiter@gold.ac.uk, pssemk@bath.ac.uk, F.Bond@gold.ac.uk, K.Chapman@gold.ac.uk}

Abstract

To investigate the relationship between presence and mood change, subjective measures of presence and mood were taken in relation to participants’ experiences of a therapeutic narrative presented either alone (i.e. audio of narrative only) or embedded within an audio-visual VE of a tropical island. The study explored the influence on mood and presence of the narrative and the VE, interrelationships between mood and presence, and any effects of mental imagery ability. The narrative was associated with decreases in negative and increases in positive discrete emotion ratings. Presence and increases in relaxation were rated more highly when the narrative was presented within the full VE than on its own. Participants’ engagement with the media presentation was positively related to reductions in reported negative affect. Mood change also correlated with mental imagery ability; however, separate analyses show that mental imagery ability only correlated with the target mood change for the narrative only condition (i.e., where no visual display was provided - where mental imagery skills are most useful).

Keywords--- presence, ITC-SOPI, 3 dimensions, emotion, sense, physical, space, engagement, ecological validity, mood change, visual analogue scales, positive, negative, affect, mental imagery

1. Introduction

We define presence as a participant’s sense of being there in a mediated environment, arising from a perceptual illusion of non-mediation; a hybrid of Barfield, Zeltzer, Sheridan and Slater [1] and Lombard and Ditton’s [2] definitions. 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. [3]; Freeman et al. [4]; Welch et al. [5] Slater and Usoh [6]; Slater Usoh, and Chrysanthou, [7] – 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.

The study of viewers’ experiences of audio-visual media such as film and TV has also at times focused on evaluating the impact of technical developments. However, viewers’ of linear media (and participants in interactive media) do not usually spend time using the media just to experience the technology. Rather, they use media to fulfil specific functions – be they related to communication, entertainment, education, relaxation, or as a means of passing the time.

As reported by Freeman [8] [9] 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 three dimensional structure of presence.

The three dimensions have been labelled 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 content; and

(3) Engagement: a participant’s sense of engagement and interest in the content of the mediated environment; determined primarily by content.

To understand people’s experiences of mediated environments, an understanding of the effects of content is required. To date, relatively little research effort has been devoted to obtaining an understanding of the effects of content variations on presence. However, in psychological research on mood induction and emotion a prime focus has been on the effects of content variation.

According to the three dimensional model of presence outlined above, variations in both media form and media content can influence presence. That content can affect both presence and emotion suggested a need for research to understand the relation between the two. Previous research has provided some evidence for this relation – through experimental research on presence and arousal [10] [11] and in a clinical context using virtual reality environments as therapeutic tools [12] [13].

Emotions are transient states of feeling, of relatively short duration, having a rapid onset and 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 is. Psychologists have found that a wide range of emotive stimuli can induce short term mood changes. These include films/stories [14], music [15] and emotive sentences [16].

The study reported here aims to determine whether a novel virtual environment (VE), conceptualized, designed
and specified by i2 media research ltd and Goldsmiths College and coded by project partners at the Interactive Institute (Sweden) as part of the EC funded project, Engaging Media for Mental Health Applications (EMMA), is effective in ameliorating (negative) mood states of stress and anxiety and promoting positive mood states (of happiness and relaxation). The VE, called ‘Relaxation Island’, is described in detail elsewhere [17]. In brief, it comprises several zones (‘waterfall’, ‘beach 1’ ‘beach 2’, and ‘cloud’). Each zone has been developed to facilitate the delivery of instructions 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.

In this study, the effects of a narrative focusing on a breathing technique scripted from an Acceptance and Commitment Therapy (ACT) perspective was explored under two conditions: eyes-closed, narrative-only vs. narrative plus the visual and background audio stimuli of Relaxation Island. This study was novel in that visual representations of imagery that might facilitate relaxation and acceptance techniques were tested here (e.g., calming sea waves, sounds of a tropical island).

This study was conducted to answer the following research questions:
(a) Does presentation of the ACT narrative result in any mood change?
(b) Is any such mood change enhanced by the presentation of the narrative within an audio-visual VE? (Specifically, is Relaxation Island really relaxing?)
(c) Is presence higher in the full VE presentation than in the narrative only version?
(d) Are any differences in mood change elicited by the two presentation conditions correlated with differences in their respective presence ratings?
(e) Are individual differences in mental imagery ability correlated with any observed effects?

2. Method

2.1. Participants

Twenty participants (10 male, 10 female) aged between 20 and 56 years (mean age: 30.2 years, SD: 8.5) took part in this study. All were students or staff of Goldsmiths College, University of London. Each participant received an incentive of £10 for taking part in the study.

2.2. Design

The study used an independent (between) groups design with two levels of one factor: presentation or not of an audio-visual virtual environment to accompany a therapeutic narrative. For clarity, note that both groups experienced the therapeutic narrative. One group was merely exposed to the therapeutic narrative. The other was presented the therapeutic narrative along with the visuals of relaxation island environment and corresponding background environmental audio (e.g., sounds of the sea, tropical birds).

2.3. Measures

Screening Measures
A range of screening measures were administered to participants to enable control of possible between group differences:
- Acceptance and Action Questionnaire (AAQ: Hayes et al., [18]). A relatively new measure comprising a 49-item scale to assess people’s willingness to accept their undesirable thoughts and feelings, whilst acting in a way that is congruent with their values and goals. Rated on a 7-point Likert scale from ‘Never True’ (1) to ‘Always True’ (7). Higher scores indicate greater psychological acceptance);
- General Health Questionnaire-12 item (GHQ-12: Goldberg & Hillier [19]);
- Depression Anxiety Stress Scale-short form (SF-DASS: Lovibond & Lovibond, [20]: 21 items);

Results from the above measures are not reported in this paper.

Mental Imagery Measure
The Short Betts’ Questionnaire on Mental Imagery (QMI: Sheehan [21] based on Betts [22] was administered prior to any media experience to obtain a measure of participants’ mental imagery abilities. The 35-items measure imagery skills in a variety of sensory modalities.

Mood Measures
- Positive and Negative Affect Schedule (PANAS: Watson, Clark & Tellegen [23]: 20 items, 10 for each general affect scale);

Presence Measures
- ITC-Sense of Presence Inventory (ITC-SOPI: Lessiter, Freeman, Keogh & Davidoff [24]: 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 two media experience conditions (instructional narrative only vs. instructional narrative presented within an audio-visual VE of a tropical island).

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. These were that they: (a) should not be taking any form of prescribed medication (except oral contraceptives), (b) should not be suffering any diagnosed emotional/psychological disorder, (c) were not currently receiving any form of psychological therapy/counseling, (d) had normal (or corrected to normal) vision, and (e) had a good grasp of the English language.

After consent was obtained, participants were asked to complete a battery of pre-test questionnaires that were fixed in the following order: Randomised Short Betts Questionnaire on Mental Imagery, Acceptance and Action Questionnaire, Depression Anxiety Stress Scale – short form, General Health Questionnaire-12, (discrete emotion) Visual Analogue Scales, and the Positive and Negative Affect Schedule. The latter two (emotion) scales were intentionally presented immediately pre-test to account for any effects on mood of questionnaire completion and therefore 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 1.9 meters 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 whether or not to have their eyes open or closed.

Participants in the eyes open condition 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 which was a projected size of 96cm by 129cm (28.5 * 37.5 degrees visual angle). Participants navigated their way to the beach zone (see Figure 1).

Participants in the eyes closed condition were instructed to close their eyes once the lights had been dimmed. 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), for participants in the eyes closed condition, the experimenter navigated through the environment from the same starting point (the signpost) and followed the same path (to ‘beach 2’) as for participants in the eyes open condition. In this condition, the VE was not projected (instead it was displayed only on the experimenter’s PC) and thus was not (even potentially) visible to the participant. This controlled for the possibility that participants might feel frustrated at being restricted from viewing something they knew was being presented.

Figure 1: Beach Zone 2 (Relaxation Island)

On arrival at the beach zone the navigator 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 ‘eyes open’ participants’ navigation was restricted to panning left and right. The pre-recorded narrative then began. Those in the eyes open condition received the full audio-visual VE with the narrative, while those in the eyes closed condition simply listened to the narrative. The experimenter remained quiet (but in the room) for the duration of the experience.

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 experimenter then led the participant back to the office where the post-test battery of measures was completed. These were 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.
3. Results

Mood state change scores (post-test minus pre-test) were calculated for each of the VAS (7 scales) and PANAS (2 scales) mood rating scores\(^1\). One-tailed significance is presented if the result obtained was predicted. T-tests and correlations were run to explore the aims of the study. Descriptives for the presence and mood state dependent variables in each condition are provided in Table 1.

3.1 The instructional narrative media presentation affects mood

Presentation of the instructional narrative significantly raised VAS rated positive discrete mood ratings of relaxation ($t_{19}=3.38$, $p<0.01$) and significantly decreased VAS rated negative discrete mood (sadness: $t_{19}=-3.32$, $p<0.01$; disgust: $t_{19}=-2.24$, $p<0.05$; fear: $t_{19}=-2.15$, $p<0.05$; anger: $t_{19}=-3.58$, $p<0.01$).

3.2 Mood change is higher for the full VE than for the narrative only media presentation

As shown in Figure 2, there was a significant difference between the two conditions on relaxed mood change. The increase in VAS-rated relaxation was significantly higher when the narrative was presented within the VE than when it was presented on its own ($t_{18}=2.18$, $p<0.05$).

There were no other significant differences between the two conditions for any of the other VAS or PANAS (positive affect [PA] or negative affect [NA]) mood change scores.

3.3 Presence ratings are higher for the full VE than for the narrative only media presentations

As shown in Figure 3, higher presence ratings were given to the full VE condition than to the narrative only condition on all ITC-SOPI factors. Significant differences were as follows:

- Sense of Physical Space ($t_{18}=2.66$, $p<0.05$: two-tailed)
- Engagement ($t_{18}=1.90$, $p=0.05$: one-tailed)
- Ecological Validity ($t_{18}=1.78$, $p<0.05$: one-tailed)
- Negative Effects ($t_{18}=2.66$, $p<0.05$: two-tailed)

3.4 Aspects of mood change and presence are related

There were significant correlations between the changes (post-pre) in VAS happiness and presence ratings (Sense of Physical Space: $r=0.63$; Engagement: $r=0.81$; Ecological Validity: $r=0.64$; $p<0.05$ and $n=20$ for these correlations). No other VAS by presence dimension correlations were significant.

There was also a significant correlation for the sample as a whole between the (post- minus pre-) change in PANAS Negative Affect and ITC-SOPI Engagement scores. As (the) engagement (dimension of presence) ratings increased so did the reduction in negative affect ($r=-0.48$, $n=20$).

When the different conditions were analyzed independently, however, an inconsistent pattern of correlations between VAS/PANAS and ITC-SOPI presence scores emerged.

3.5 Mental Imagery ability is related to aspects of mood change

Overall, (post- minus pre-) change in VAS rated relaxation correlated with mental imagery ability ($r=0.50$, $n=20$, $p<0.05$); however, separate analyses showed that mental imagery ability only correlated with increases in relaxation for the narrative only condition (i.e., where no visual display was provided; or in other words where mental imagery skills are most useful: $r=0.58$, $n=10$; $p<0.05$: one tailed).

---

\(^1\)‘During’ test mood state scores (VAS and PANAS) are not reported here.
Table 1. Descriptives for the presence and mood state dependent variables

<table>
<thead>
<tr>
<th></th>
<th>NARRATIVE +VE Mean (SD)</th>
<th>NARRATIVE ONLY Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sense of Space</td>
<td>3.05* (.67)</td>
<td>2.24 (.69)</td>
</tr>
<tr>
<td>Engagement</td>
<td>3.35* (.46)</td>
<td>2.89 (.61)</td>
</tr>
<tr>
<td>Ecological Validity</td>
<td>3.16* (.76)</td>
<td>2.54 (.79)</td>
</tr>
<tr>
<td>Negative Effects</td>
<td>2.33* (.34)</td>
<td>1.73 (.62)</td>
</tr>
<tr>
<td>VAS-Happiness</td>
<td>8.07 (11.23)</td>
<td>-6.61 (12.50)</td>
</tr>
<tr>
<td>VAS-Anger</td>
<td>-9.98 (12.91)</td>
<td>-14.45 (17.70)</td>
</tr>
<tr>
<td>VAS-Disgust</td>
<td>-6.55 (14.11)</td>
<td>-4.90 (8.64)</td>
</tr>
<tr>
<td>VAS-Relaxation</td>
<td>25.36* (11.11)</td>
<td>6.56 (24.90)</td>
</tr>
<tr>
<td>VAS-Fear</td>
<td>-6.5480 (13.08)</td>
<td>-3.4500 (7.24)</td>
</tr>
<tr>
<td>VAS-Sadness</td>
<td>-12.2000 (15.47)</td>
<td>-6.8200 (9.51)</td>
</tr>
<tr>
<td>VAS-Surprise</td>
<td>-.05 (10.88)</td>
<td>8.03 (13.70)</td>
</tr>
<tr>
<td>PANAS-PA</td>
<td>-3.90 (4.33)</td>
<td>-2.70 (3.16)</td>
</tr>
<tr>
<td>PANAS-NA</td>
<td>-2.40 (4.62)</td>
<td>.10 (2.13)</td>
</tr>
</tbody>
</table>

* p < 0.05

4. Discussion and Conclusions

The study described here was designed to explore whether the effects on mood of a therapeutic narrative were enhanced through its presentation within an audio visual virtual environment. This study is part of a larger factorial experiment that is currently in progress, on which a mediation analysis is planned. This particular study used a relatively small sample size (n = 10 in each of the two conditions), and thus the results are considered preliminary. Nevertheless, some interesting findings emerged.

The narrative itself was associated with significant mood change; reducing discrete negative emotion ratings, and increasing relaxation. Most importantly for our research questions, the presentation of the narrative within the audio visual virtual environment resulted in significantly greater increases in relaxation relative to presentation of the narrative alone, but there were no effects on the more global affect ratings (PANAS). It is an interesting and important finding that a VE can be developed to manipulate a specific target emotion. The current results suggest good potential for the use of virtual environments as devices to improve relaxation for the general population. It is envisaged that the effect measured here with relation to relaxation could generalize to other mood states. This is in addition to the already well documented clinical efficacy of VE based therapeutic applications such as the treatment of phobias.

As expected, presence ratings were higher for the condition that provided a visual representation of the media space (narrative presented within an audio-visual VE) than for the media presentation that did not (audio only presentation of the narrative). It is important to note that presence ratings for both conditions were with reference to the “mediated environment”. Clearly, taking presence measures in relation to an audio only stimulus could be described as problematic according to some theories of presence (e.g., Biocca, [26]; Waterworth & Waterworth, [27]) because by definition the experimental context required participants to “fill in the gaps” from the minimal sensory cues provided by the narrative. Indeed, such an experience could be referred to as (at least partly) matching Waterworth and Waterworth’s [27] definition of absence. The current study was conducted as part of a much larger (ongoing) research programme, focusing both on testing theories of presence and on evaluating the applied potential of virtual environments as mood devices for a range of applications.

That presentation of the narrative within a full VE was associated with higher presence provides further evidence for the extensively documented effects of media form on presence. Correlations indicated some relationship between presence and mood change, though these were not consistent. It is possible that this instability is attributable to low power from the small sample size used in the study. Further analyses, and experiments, are planned to better understand the relationship between aspects of presence and emotion and the direction of the relationship(s). Within future experiments it is planned to manipulate presence and emotion independently, though care is required in the design of these studies given the theoretical starting point that both can be affected by content variations.

Finally, because there is a theoretical possibility that individual differences in mental imagery ability could have negated any difference in mood change between the two conditions used in this experiment, we explored whether mood change was related to mental imagery ability. That mental imagery ability only correlated with the increases in relaxation for the narrative only condition suggests that
whilst a relaxing virtual environment (such as Relaxation Island) promises applied potential for the general population, people with high imagery abilities may have less to gain from the presentation of a therapeutic narrative within a full audio-visual VE than people with lower imagery abilities.

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


[9] 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


Why and How Do the Telepresence Dimensions Influence Persuasive Outcome?

Anne-Cécile Jeandrain
E-Management Competence Center Coordinator
UCL School of Management
Catholic University of Louvain, Belgium

Abstract
Research on the relationship between telepresence and persuasion is sparse. Especially absent is research on the specific contribution of each telepresence dimensions (engagement, naturalness and spatial presence) on persuasive outcome. The author proposes a theoretical framework to examine this relationship. She then empirically tests research propositions in a laboratory experiment. The data, collected from a sample of 63 individuals in U.S., indicate that the telepresence factors have a differential influence on global attitude and attitude confidence. The implications of this research for future work are discussed.

Keywords— Telepresence, persuasion.

1. Introduction

Imagine that you are navigating in a desktop virtual environment. This environment is composed of several worlds, oriented to entertainment, education, and so on. One of these worlds lets you the opportunity to enter in a store and make your shopping trip as you do in the “normal” life (walking through the aisles, picking and manipulating some products and so on). Few years ago, this way of online shopping was considered as the future of e-commerce. Indeed, it represented an attempt to integrate in a single tool buyer-seller relationships, experiential transaction, and post-purchase service – in other words, the optimal solution to bring online compelling experience to consumers and to build one-to-one relationships with them. Now some “3D stores” are disappeared and others are transformed in promotional tools for 3D designers and no more in such integrative marketing tools as initially conceive by companies. A reason of such failure was the reluctance of companies to pay a fee for having their online presence via such “sophisticated” desktop interface without knowing if the output will be positive or negative, at which extent, and at which conditions.

Prior research regarding what effects virtuality might have on consumers in a marketing perspective have mainly focused on virtual product experience [1, 2, 3, 4, 5] and on virtual agents [6, 7, 8]. The extant literature lacks empirical research on the impacts of large virtual environment on consumers’ typical responses to product-oriented message (i.e., affective, cognitive and behavioral) [9, 10, 11]. More specifically, does the commercial virtual environment influence the consumers’ product brand attitude, her/his confidence in this judgment and his/her purchase intention? If any, in what way does this influence occur? Are these global configurations more or less effective than those focally to the product? These questions (and many more) still remain unanswered.

To address the aforementioned issues, marketing-oriented literature concerning virtual reality has often proposed virtual presence (telepresence and/or social presence) as potential facilitator of consumer persuasion. For example, consumers interacting with products in 3D advertising are likely to perceive a sense of telepresence, which results in product knowledge, brand attitude, and purchase intention [4]. An anthropomorphic agent has been shown to increase social presence and telepresence, which in turn have an impact on attitude toward the message transmitted by this agent [8]. Finally, product placements in virtual environments have greater persuasive impacts when telepresence is higher [11]. Telepresence and social presence are commonly conceptualized as multidimensional constructs [12, 13, 14]. However, little is known about the specific contribution of those constructs dimensions on persuasive outcome.

We decide to focus only on telepresence, i.e. the spatial counterpart of the perceptual illusion of nonmediation [12] and not its social counterpart. In this framework, this study aims to examine how telepresence through its dimension (engagement, naturalness and physical space) might enhance consumer persuasion. We first describe the conceptual framework (i.e., the three-factors conceptualization of telepresence, the process of persuasion and the underlying theories). We then formulate hypotheses linking telepresence dimensions and persuasive outcome. Next, we report results from a laboratory experiment. Finally, we propose avenues for further research.
2. Conceptual framework

Our conceptual framework is mainly built on direct experience and source-monitoring error theories. We first define the two core concepts and then present the aforementioned theories.

2.1. Telepresence

We adopt the ISPR (International Society for Presence Research) view of presence: “subjective perception in which even though part or all of an individual's current experience is generated by and/or filtered through human-made technology, part or all of the individual's perception fails to accurately acknowledge the role of the technology in the experience”.

Presence is thus considered as a psychological construct rather than a technological property of the virtual environment. In this framework, all interfaces (Imax, television, desktop screen and so on) are likely to facilitate some sense of telepresence. However, “immersive” virtual environment created by tools such as head-mounted display and data gloves at this point of time can generate highest senses of telepresence [12]. On the other hand, the term “telepresence” is here preferred to “presence” in order to focus only on spatial aspects and exclude social ones, as mentioned before.

More precisely, telepresence is considered as a three-factors construct. These factors are the following ones [13]:
- **Engagement**, which refers to cognitive as well as affective aspect, i.e. “in telepresence, as the user focuses more attention on the virtual stimuli (because of the degree of significance or meaning that the individual attaches to the stimuli, activities or events), and due to the potential intrinsically motivating quality of such experience [15], the individual will become cognitively and affectively more involved in it.”
- **Naturalness**, the telepresence reality judgment component, i.e. “in telepresence, the user will perceive the environment as perceptually (photorealistic) and socially (“true to life”) realist to her/him”.
- **Physical space**, the telepresence spatial-constructive component, referred to the idea of transportation, i.e. “in telepresence the user feels her/his phenomenal body spatially located in the virtual environment and s/he perceives her/himself to be enveloped by, included in, and interacting with the virtual environment and its elements”.

Telepresence is viewed here as a continuous variable in the sense that the degree of telepresence is likely to vary from the peripheral sense of spatial presence to progressively higher levels of telepresence characterized by greater cognitive and affective engagement in the virtual experience.

2.2. Persuasion

In social psychology literature, definitions of persuasion could be either process-oriented or outcome-oriented [16]. The first kind of definitions assumes persuasion as an activity or a process in which a communicator attempts to induce a change in the individual’s belief, attitude, or behavior. Therefore, persuasion can occur regardless of whether anybody is actually influenced by a message. In contrast, the second kind emphasizes successful attempt to influence. It does not make sense to state “I persuaded him but failed” [17]. We rather adopt the second view for integrating impacts on people, positive as well as negative ones. Therefore, we use the following definition [18]:

“persuasion corresponds to a communicational act, that aims at modifying the inner state of the individual and, in fine the individual’s behavior, in a context in which the individual has some degree of free choice”

In order to predict the extent to which attitudes will change, researchers must understand the underlying psychological processes. For doing so, dual-process models are an attempt to integrate conflicting findings in providing comprehensive framework. The Elaboration Likelihood Model (hereafter ELM) [19] is preferred to all other persuasion models because of its integrative perspective, its empirical validity and its theoretical pertinence. The term « elaboration » in the model refers to the extent to which people think about issue-relevant arguments contained in persuasive message. The model’s key hypothesis postulates that the elaboration likelihood moderates the route to persuasion. If the elaboration likelihood is high, the probability that people follow the “central” route to persuasion also will be high. In fact, the ELM outlines two basic routes to persuasion: the central route, along which the person changes her/his attitude on the basis of elaboration on arguments contained in the message and, the peripheral one along which the person may change her/his attitude on the basis of peripheral cues (for instance, receiver attractiveness).

In persuasive outcome, we consider: attitude confidence, attitude richness and purchase intention. Attitude confidence is a cognitive construct that reflects one’s conviction in one’s beliefs, which may be caused by quantity of information, credibility of information, and consistency of information [20]. Attitude richness refers to “the amount of contextual and sensorial information that accompanies one’s attitude in memory” [21], in other words, a knowledge-related construct. Finally, behavioral response is commonly measured through the notion of purchase intention. This measurement is far from perfect because it only captures intentions and not true behaviors, which are very difficult to measure, especially in a laboratory experiment.

2.3. Direct experience effect

In perceiving telepresence, individuals feel their presence in the virtual environment and less and less in
the immediate physical one [22]. Secondly, individuals respond to virtual stimuli like they would do in the reality [12]. For instance, they could have the impression to "really" manipulate a product in a virtual store and thus adopt the same behavior as the one in physical stores. Therefore, telepresence experience is supposed to be located closer to direct experience than to indirect experience on an experience spectrum [1], which might result in the same cognitive and affective consequences on the consumer as those for direct experience.

The way by which attitudes are formed (direct experience or indirect experience) is recognized as an important moderator between attitude and behavior. The distinction between direct experience and indirect experience of an attitudinal object represents a continuum. At one extremity, direct experience designates a behavioral action with the attitudinal object. At the other extremity, indirect experience means the attitude formation through an informational contact with the object such as information reading about this object [23]. For instance, word-of-mouth is very close to indirect experience (i.e., the consumer receives product information but can't touch it) while product trial is a direct experience [23].

Attitudes formed through direct experience have been shown to be more consistent to behavior than those based on indirect experience [19, 24, 25, 26, 27]. Due to effective object manipulation, attitudes based on direct experience are linked with greater elaboration of the merits of the object than attitudes formed through indirect experience like passive exposure [28, 29]. In turn, greater elaboration results in attitude confidence [26, 27]. In the same way (through virtual manipulation and thus elaboration), we assume that both physical space-related and engagement-related components of telepresence are likely to influence attitude confidence. On the other hand, the more the individual will have a "sense of believing" (naturalness dimension), the more s/he will also be confident in her/his judgment toward the environment and/or the objects inside it. In other words, the influence of engagement and of physical space is rather due to elaboration while the influence of naturalness depends rather on familiarity due to manipulation.

2.4. Source-monitoring error

If telepresence experience has the potential to involve users in sensory worlds that are indistinguishable or nearly indistinguishable from the real world, then people have to make sophisticated judgments about what is real and what is not [30], which contributes to source-monitoring error. Source monitoring error assumes that information from fictional source will be encoded with same qualitative characteristics as information from direct experience [20]. In other words, the richness of an attitude formed during a mediated experience is supposed to be high when a source-monitoring error has occurred.

Therefore, the more the user will have the perception of "real" interaction with the virtual environment and/or its elements, and the more the individual will blend information provided by the virtual environment with information from the real world due to her/his impression of naturalness, the more s/he will form attitudes toward the object with rich sensorial and contextual details. In other words, attitude richness is assumed to depend on the telepresence dimensions of physical space and of naturalness. The influence of engagement is not assumed to be explained with source-monitoring error (a more unconscious process).

3. Hypotheses

Based on these theories, we formulate the following hypotheses:

Hyp. 1: Telepresence through its dimension of engagement is likely to positively influence brand attitude confidence and store attitude confidence.

Hyp. 2: Telepresence through its dimension of naturalness is likely to positively influence brand attitude richness, brand attitude confidence, and store attitude confidence.

Hyp. 3: Telepresence through its dimension of physical space is likely to positively influence brand attitude richness, brand attitude confidence and store attitude confidence.

A next study will be needed in order to validate the underlying theories of these relationships.

4. Method

4.1. Experimental design and stimuli

To create variation in telepresence, we produced two commercial virtual stores representing low and high level of social realism (figures 1 and 2). Social realism refers to the extent to which a media portrayal is plausible or "true to life" in that it reflects events that do or could occur in the nonmediated world [12]. Media content and more specifically social realism has been suggested to positively influence telepresence [31, 32, 33]. Social realism is manipulated according to the number of details (not related to the product advertised) and the number of collateral products (18 in the low condition and 69 in the high one) present in the virtual environment.

On the other hand, as there were not time limitations for navigation, and due to the differential level of social realism, we expect that users in the low condition will take less time to visit the virtual environment and thus will have less opportunity to perceive telepresence, in comparison with individuals in the high condition.

The stimulus is a quasi 3-D virtual environment that displays an outside view of a computer store. The store has two floors, the first one showing the stimulus product and the second floor being used for video games. Three criteria were used for selecting the focal product: product with geometric properties rather than material ones (i.e., easier to evaluate the central merits of the brand rather
than smell, taste and hearing) and a technologically oriented product (i.e., in order to have congruency between the product type and its placement in an “innovative” environment). Based on these two criteria, the product selected was a notebook computer. Major collateral products were games-oriented products, and therefore not directly related to the notebook computer. Its brand was unknown (i.e., in order to avoid brand familiarity between subjects).

The procedure was as follows. The subjects participated in the experiment (approximatively 30 minutes) one by one at a time in an “immersive” room (i.e., a small room with a computer surrounded by wood panels). They were told that the study was designed to obtain people’s opinions about computer product presentation in 3D virtual store that was owned by a European company. The subjects were asked to fill in the first part of the questionnaire (demographics and computer usage questions) before training themselves in 3D navigation (in a labyrinth). After visiting the store (without any time limitation), they filled in the questionnaire’s final part to check the manipulation for vividness and to obtain a measure of telepresence, overall attitude toward the brand and the store as well as the confidence in these judgments and finally purchase intention. After one week, they received a short questionnaire by email in order to measure attitude richness concerning the virtual experience.

4.2. Sample

The subjects were 63 undergraduate students of US post-graduate students. They were recruited from communication courses and received extra credit points for their participation. They were randomly assigned to one of the two conditions (in average 30 in each). However a large majority of students who had participated in the experiment came from telecommunication, which imply a very frequent Internet usage, favorable affect toward the Internet and great perceived skills in it. These notions were not significantly different across the groups.

4.3. Measures

Telepresence was measured using a shortened version of the Independent Commission-Sense of
Presence Inventory (ITC-SOPI) [34] were used, adapted to a product and a store. A measure of Internet experience [35] consisted in asking the total number of hours spent on web, how long ago they first accessed the web, how often they accessed it and whether they felt “at ease” on the web. Virtual reality-related item was added to this scale in order to check the familiarity with 3D navigation. Positive and Negative Affect and Internet self-efficacy tests was grouped as a set of 16 questions (5-point) (for instance, “using the Internet is fun”, “using the Internet is frustrating”, “it is easy for me to use the Internet”).

Global attitude toward the brand was measured with 5 bipolar semantics differentials [36] (for instance, bad/good, high quality/low quality). We created an overall brand attitude measure by averaging these scale responses. 6 bipolar semantics differentials measured the attitude toward purchasing the brand [37] (for instance, foolish/wise, harmful/beneficial). Global attitude toward the store was measured with 12 bipolar semantics differentials [38] (for instance, modern/traditional, outgoing/inward). The other persuasive outputs were attitude confidence (2 seven-points items) [39] (adapted for the product and the store), attitude richness (28 bipolar semantics differentials, Memory Characteristics Questionnaire) [17], and purchase intention (4 bipolar semantics differentials) [40].

5. Results

The social realism manipulation was successful because we get: (1) higher scores in the store social realism scale for the high social realism condition (hereafter HSR) than for the low one (hereafter LSR) (mean\_LSR = 2.63 < mean\_HSR = 4.71; t=-11.298, p1=0.000), (2) no significant difference between the product social realism scores which has to be at low level (mean\_LSR = 2.51 ~ mean\_HSR = 2.18), (3) more time spent in the store for the HSR condition than the LSR one (mean\_LSR = 3.94; p1=0.000 / mean\_HSR = 3.94; p1=0.000), and more importantly (4) higher scores in the three dimensions of telepresence for the HSR condition than for the LSR one (mean\_engagement\_LSR = 2.04 < mean\_engagement\_HSR = 3.88; p1=0.000 / mean\_naturalness\_LSR = 1.99 < mean\_naturalness\_HSR = 3.94; p1=0.000 / mean\_physical\_space\_LSR = 2.06 < mean\_physical\_space\_HSR = 3.9; p1=0.000).

Brand attitude confidence only depends on the engagement dimension of telepresence (37% of total variance explained). The influence is negative: the more the user will focus her/his attention in the virtual environment, the less s/he will express attitude confidence. Before the experiment, the individual was asked to explore the environment, which leads to allocation of attention to all content elements, and not specifically to the focal product. As a consequence it could be logical that s/he has relatively less focused on the focal product, and thus are less confident in his/her judgment.

Store attitude confidence is only influenced by the physical space dimension of telepresence (30% of total variance explained). We could here assume that due to spatial construction the individual will more appropriate the space, even virtual (i.e., to transform and personalize the space in her/his mind in order to get the impression of control), that leads to more confidence toward this space. Such effect was expected also on the product side, which was not the case, probably due to the poor product design.

As not expected, we found, a positive influence of the naturalness dimension of telepresence on global brand attitude (16% of total variance explained) and on attitude toward purchasing the brand (15% of total variance explained). We could assume that the individual who considers the environment as believable is likely to form favorable attitudes toward the brand and toward purchasing the brand. The hypothesis of possible influence on attitude extremity has to be clarified.

Therefore, direct experience effect could be proposed as an explanation of the relationships between physical space dimension and store attitude confidence, and between engagement dimension and brand attitude confidence, due to possible higher elaboration. However, probably due to poor product manipulation, we didn’t found any influence of the naturalness dimension on attitude confidence.

No influence of the three factors on attitude richness was found. A reason of that could be the attitude richness scale, normally used to compare fictional memory and real one and not two mediated experiences.

6. Conclusions

The purpose of this study was to deeper understand the specific contribution of each telepresence dimension on persuasive outcome. This could be particularly interesting in a theoretical perspective but also for companies, which desire to have a presence in quasi 3D virtual stores.

Results about the influence of each telepresence dimensions on persuasion are quite mixed. On the one hand, some persuasive outputs are likely to be influenced by a particular telepresence dimension variation. For instance, the extent to which the individual will judge the virtual environment as close to the reality (i.e., naturalness dimension) will positively influence global attitude toward the brand. On the other hand due to the experimental design characteristics, some relationships are not significant (e.g., the relationship between naturalness and brand attitude confidence). Indeed, the virtual store was not perfectly designed in the sense that some graphics were perceptually and socially realistic but not enough to have a significant impact on attitude characteristics.

The insights given by our study have to be validated in another study, which will take into account design recommendations. However, to our knowledge, our research is the first attempt to empirically examine the effects of telepresence dimensions on persuasion and to
bring theoretical explanations of such relationships. Two linkages should be further investigated: among attitude extremity and naturalness, among store-oriented variables and all telepresence dimensions. Moreover, it could be interesting to clarify if there is any halo effect of the store-related variables on the product-related ones facilitated by telepresence.

Acknowledgements

The author would like to gratefully thank the Mindlab (Media Interface & Network Design Lab), Michigan State University, U.S.A, where the data collection was made.

References


Effects of Physical Embodiment on Social Presence of Social Robots

Younbo Jung, Kwan Min Lee
Annenberg School for Communication, University of Southern California,
3502 Watt Way, Los Angeles, CA 90089-0281, USA
{younboju@usc.edu, kwanminl@usc.edu}

Abstract
Two experiments were conducted to investigate the relative effectiveness of physical embodiment on social presence of social robots. The results of Experiment 1 show positive effects of physical embodiment of social robots (PESR) in the feeling of social presence, the general evaluation of social robots, the assessment of public opinion of social robots, and the evaluation of interaction with social robots. The result of a path analysis also provides the evidence of the mediating effect of social presence in people’s general evaluation of social robots. However, the results of Experiment 2 show that PESR without touch-input capability causes negative effects. Implications for the relative effectiveness of PESR, the importance of tactile communication in human-robot interaction, as well as the market potential for social robots in relation to loneliness are discussed.

Keywords---Physical embodiment, touch, human-robot interaction, companionship, social robots, tactile communication.

1. Introduction

In the movie “A.I.” directed by Steven Spielberg, Cybertronics, a firm that manufactures robots, creates a new boy robot to both give and elicit the genuine emotion of love. Its name is David. The major purpose of David is to share emotional reciprocity with human beings, which is totally different from the purpose of utility robots such as cleaning, cooking, or industrial manufacturing. Although we do not have social robots as sophisticated as David in “A.I.” yet, researchers and practitioners have begun to realize that robots can be social actors whose major purpose is to interact with humans in a socially meaningful way.

Lee et al. [7] define social robots as “new types of robots whose primary goal is social interaction with humans.” In other words, a social robot is a robot designed to evoke meaningful social interaction with its users. Given the above definition, social robots do not necessarily need physical embodiment to accomplish their goals unlike other functional robots. Physical embodiment is a mandatory requirement for functional robots because the purposes of them are mostly related to labor-intense works such as cleaning, destroying, or lifting. However, the purposes of socially interactive robots are research platform, toys, educational tools, and therapeutic aids including emotional supports [4]. Most of these goals are not directly related to physical activities per se, thus can be delivered via virtually embodied social robots.

One of the most fundamental questions on social robots is whether or not physical embodiment is required for the successful social interaction between human and social robots. This is a critical question to both researchers and practitioners due to high costs for manufacturing physically embodied social robots, not to mention technical difficulties. Nonetheless, there are only a few empirical studies investigating the effects of physical embodiment of social robots (PESR) in human-robot interaction.

Two experiments were conducted in this study in order to investigate the relative effectiveness of PESR. In addition, the feeling of social presence is examined in a path analysis to investigate its mediating effects in human-robot interaction. Finally, we discuss how lonely and non-lonely people respond differently to social robots such as Sony Aibo or Samsung April.

2. Literature review

2.1. Embodiment

Fong et al. [4] define embodiment as “that which establishes a basis for structural coupling by creating the potential for mutual perturbation between system and environment.” Given the definition, as long as the relationship between a system and its environment is perturbative, physical embodiment is not necessary any more. Thus, this relational definition of embodiment raises an important question whether or not a physical embodiment is essential in designing social robots.

The positive expectation about physical embodiment is that PESR may result in better affordance, thus lead to less frustration from people. In fact, the form and structure of a robot is important because it helps establish social expectations [4]. Therefore, the following hypotheses are proposed:
2.2. Social presence

Lombard et al. [8] define presence as “the perceptual illusion of non-mediation.” Similarly, Lee [5] defines social presence as “a psychological state in which virtual (para-authentic or artificial) actors are experienced as actual social actors in either sensory or nonsensory ways.” Given the above definitions, the feeling of social presence can play an important role in successful social interactions with even non-human beings. When a person interacts with a social robot, the person may respond to the social robot—an artificial social actor—as if it were an actual social actor. For example, although David in the movie, “A.I.,” is not a real boy, the mother is satisfied with her emotional fulfillment from social interaction with David. It means that she feels strong social presence of her real son—an actual social actor—when she interacts with David—an artificial actor—in sensory ways.

Bartneck [2] found social facilitation effect in his study with an emotional robot, eMuu. In the study, participants acquired higher score in the negotiation game when they interacted with a robot character—physically embodied character—than when they interacted with a screen character—physically disembodied character. Although Bartneck did not use a specific term of “social presence” in his study, the finding of social facilitation effect can be explained in terms of social presence. Participants in the experiment put more effort into the negotiation when they interacted with the physically embodied character because they felt strong social presence.

Indeed, the feeling of social presence is highly likely to be related to the richness of sensory inputs because the more a person feels, the stronger the person believes (see [18]). PESR can provide people with richer sensory inputs than physically disembodied social robots (PDSR), which may result in people’s strong feeling of social presence. The strong feeling of social presence, then, may result in positive effects of social robots. Therefore, social presence could be a mediating factor in the effectiveness of PESR. Based on this assumption, the following hypotheses are proposed:

H1: People will evaluate a social robot more positively when they interact with a physically embodied social robot than when they interact with a physically disembodied social robot.

H2: People will assess other people’s evaluation of a social robot more positively when they interact with a physically embodied social robot than when they interact with a physically disembodied social robot.

H3: People will evaluate the interaction with a social robot more positively when they interact with a physically embodied social robot than when they interact with a physically disembodied social robot.

H4: People will feel stronger social presence when they interact with a physically embodied social robot than when they interact with a physically disembodied social robot.

H5: The effects of physical embodiment on general attraction of a social robot, assessment of public opinion of a social robot, and evaluation of the interaction with a social robot will be mediated by users’ feeling of social presence.

2.3. Loneliness

Social robots are similar to pets in a way that both of them provide people with companionship. Similar to the findings that interaction with pets would be good complementary to or even substitute for traditional interpersonal interaction [19], social robots may be able to satisfy one’s needs for social interaction.

We are not surprised to see that Rook [15] found a significant negative relationship between loneliness and companionship. In his study more frequent companionship with other people was associated with less loneliness. Therefore, a lonely person is likely to appreciate the interaction with social robots more positively than a non-lonely person because the former is more in need of social companionship. Based on this assumption, the last hypothesis is proposed:

H6: Lonely people will feel more socially attracted to a social robot than non-lonely people.

3. Experiment 1: Effects of physically embodied social robots

3.1. Experiment design

A 2 (embodiment vs. disembodiment) x 2 (lonely vs. non-lonely) between subjects factorial analysis of variance (ANOVA) design was used to test the hypotheses in a laboratory environment. A total of 36 undergraduate students enrolled at a major west-coast university were participated in the experiment.

3.2. Procedure

The whole experiment process consists of three steps. First, a survey of UCLA Loneliness Scale (Version 3) was administered in a larger data pool of 62 people. Then, 16 participants were selected from each extreme end of the scale based on their survey scores.

Second, 16 participants within each group (lonely vs. non-lonely) were randomly assigned to one of the two different embodiment conditions (embodiment vs. disembodiment). For example, 8 participants in the lonely group were randomly assigned to the embodied condition. The other half (8 participants) in the lonely group were also randomly assigned to the disembodied condition.

Finally, participants were asked to come to a laboratory where they individually interacted with a social
robot, Sony Aibo, for about 10 minutes alone. Then, participants were asked to complete a paper-based survey.

### 3.3. Manipulation

Two conditions of physical embodiment were manipulated for Experiment 1. First, Sony Aibo was selected to represent a physically embodied social robot because Aibo is one of the most successful social robots currently on the market [7]. Aibo contains sensors in its head, chin, and back that enable interactions with people by affectionate pats. For the experiment purpose, Aibo was programmed to perform singing and dancing for 2 minutes and 20 seconds. After its performance, Aibo was also programmed to interact with participants by responding to its sensory inputs in limited and constant ways. Participants in the physically embodied condition interacted with actual Aibo.

Second, physically disembodied Aibo was manipulated by using animation-making software, Director. Aibo’s performances were pre-recorded in a digital camcorder. Then, the digital video files in AVI format were imported to Director for final manipulation. The created program was shown in a shockwave player on a 17 inch flat-screen monitor. When participants in the physically disembodied condition clicked any of Aibo’s sensory input areas, they saw a virtual hand pushing the sensory area on the screen, and then saw a particular response from Aibo just like they would have seen from Aibo in the physically embodied condition.

To summarize, the main difference between two embodiment conditions was whether or not participants interacted with actual Aibo or virtual Aibo that was digitally programmed.

### 3.4. Measure

All dependent measures were based on items from paper-based questionnaires.

UCLA Loneliness Scale (Version 3) was used to measure participants’ perceived loneliness. The scale has been tested in many studies and regarded to be highly reliable in terms of internal consistency (coefficient α ranging from .89 to .94) and test-retest reliability over a one-year period (r = .73) [16].

Seven questions concerning the general evaluation of Aibo were asked using 10-point Semantic-differential scale: bad/good; bitter/sweet; cruel/kind; distant/close; not friendly/friendly; not loving/loving; unpleasant/pleasant (Cronbach’s α = .91). This is a modified measure from the study of Perception Of Pets As A Companion by Poresky et al. [13].

Six questions about social presence were asked using a combination of 10-point Semantic-differential scale and independent 10-point scale: unsociable/sociable; impersonal/personal; machine-like/life-like; insensitive/sensitive; while you were interacting with this Aibo, how much did you feel as if it was a social being?; while you were interacting with this Aibo, how much did you feel as if it was communicating with you? (Cronbach’s α = .920).

Psychological perspective of social presence such as attention and involvement were disregarded because 10 minutes of interaction time was too short to establish involvement. Besides, participants were asked to pay attention to Aibo and allowed to interact with Aibo alone in a laboratory room. Therefore, measuring participants’ attention is considered to be inappropriate for the measure of social presence in this study.

Three questions concerning public opinion of Aibo were asked using independent 10-point scale: People will find it interesting to play with this Aibo; People will find this Aibo attractive; People are likely to buy this Aibo (Cronbach’s α = .79).

Six questions dealing with the evaluation of interaction with Aibo were asked using independent 10-point scale: enjoyable; entertaining; exciting; fun; interesting; satisfying (Cronbach’s α = .88).

Social attraction and physical attraction were measured by a modified version of McCroskey and McCain’s Interpersonal Attraction Scale [9]. Three questions about social attraction of Aibo were asked using independent 7-point scale: I think this Aibo could be a friend of mine; I think I could spend a good time with this Aibo; I would like to spend more time with this Aibo (Cronbach’s α = .92).

Finally, three questions concerning physical attraction of Aibo were asked using independent 7-point scale for the manipulation check: I think this Aibo is quite pretty; This Aibo is very good looking; I find this Aibo very attractive physically (Cronbach’s α = .88).

### 3.5. Results

Table 1 shows a full correlation matrix of the measured variables in Experiment 1. Although personal evaluation and the assessment of other people’s attraction may sound similar, they are different concepts due to the third person effects [12]. A person may not feel attracted to Aibo at all, but may think that other people would. Hypotheses were tested with one-way between subjects ANOVA. Consistent with Hypothesis 1, participants evaluated the physically embodied Aibo more positively (M = 8.14, SD = 0.8499) than the physically disembodied Aibo (M = 7.34, SD = 1.0970), F (1, 28) = 5.141, p < .05, η² = .191.

Consistent with Hypothesis 2, participants assessed other people’s evaluation of a social robot more positively when they interacted with the physically embodied Aibo (M = 7.98, SD = 1.0644) than when they interacted with the physically disembodied Aibo (M = 7.21, SD = 1.0829), F (1, 28) = 4.4, p < .05, η² = .233.

Consistent with Hypothesis 3, participants evaluated the interaction with Aibo more positively when they interacted with the physically embodied Aibo (M = 7.13, SD = 1.1395) than when they interacted with the physically disembodied Aibo (M = 7.13, SD = 0.7245), F (1, 28) = 15.28, p < .01, η² = .599.

Consistent with Hypothesis 4, participants felt stronger social presence when they interacted with the...
Table 1
Correlation Matrix of Six Measured Variables in Experiment 1

<table>
<thead>
<tr>
<th>Measured Variables</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. General Evaluation of Aibo</td>
<td>.228</td>
<td>.258</td>
<td>.374*</td>
<td>.584**</td>
<td>.337</td>
<td></td>
</tr>
<tr>
<td>2. Social Attraction of Aibo</td>
<td>.278</td>
<td>.188</td>
<td>.261</td>
<td>.096</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Assessment of Public Opinion</td>
<td>.404*</td>
<td>.454**</td>
<td>.117</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Evaluation of Interaction with Aibo</td>
<td>.469*</td>
<td>.250</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Social Presence</td>
<td></td>
<td></td>
<td></td>
<td>.469*</td>
<td>.250</td>
<td></td>
</tr>
<tr>
<td>6. Physical Attraction of Aibo</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.343</td>
<td></td>
</tr>
</tbody>
</table>

Note: *p < .05, **p < .01 (2-tailed).

A path analysis was conducted to test Hypothesis 5, which predicted the mediating effect of social presence on the general evaluation of Aibo (see Figure 1).

The physical embodiment was a significant predictor for the general evaluation of Aibo ($\beta = .385$, $p < .01$) and for the feeling of social presence ($\beta = .724$, $p < .001$) when it was the only predictor entered into a simple linear regression. The feeling of social presence was also a significant predictor when the evaluation of Aibo was the only predictor regressed on social presence ($\beta = .641$, $p < .01$). However, the effect of physical embodiment on the general evaluation of Aibo significantly dropped and became non significant when the dependent variable was regressed on both physical embodiment and social presence. Put together, the series of the regression analyses reported in Figure 1 provide a strong evidence for the mediating role of social presence only in people’s general evaluation of social robots.

For three other dependent variables both of the effects of embodiment and social presence became non significant when the dependent variables were regressed on both embodiment and social presence. In conclusion, the current path analysis provides evidence of the mediating role of social presence only in people’s general evaluation of social robots.

Figure 1. Path Analysis of the Mediating Effect of Social Presence in the Evaluation of Aibo: Experiment 1.

Note: Numbers inside arrows are standardized coefficients for each regression. Numbers inside parentheses are standardized coefficients when the evaluation of Aibo was regressed on physical embodiment alone. Two conditions of physical embodiment were dummy coded: 0 disembodied; 1 embodied.

physically embodied Aibo ($M = 7.81$, $SD = .8185$) than when they interacted with the physically disembodied Aibo ($M = 5.62$, $SD = 1.2903$), $F(1, 28) = 32.27$, $p < .001$, $\eta^2 = .545$.

A path analysis was conducted to test Hypothesis 5, which predicted the mediating effect of social presence on other dependent variables (see Figure 1).

The physical embodiment was a significant predictor for the general evaluation of Aibo ($\beta = .385$, $p < .01$) and for the feeling of social presence ($\beta = .724$, $p < .001$) when it was the only predictor entered into a simple linear regression. The feeling of social presence was also a significant predictor when the evaluation of Aibo was the only predictor regressed on social presence ($\beta = .641$, $p < .01$). However, the effect of physical embodiment on the general evaluation of Aibo significantly dropped and became non significant when the dependent variable was regressed on both physical embodiment and social presence. Put together, the series of the regression analyses reported in Figure 1 provide a strong evidence for the mediating effect of social presence on people’s general evaluation of social robots (for the statistical proof of why the above analyses provide a convincing evidence for mediation, see [1]).

For three other dependent variables both of the effects of embodiment and social presence became non significant when the dependent variables were regressed on both embodiment and social presence. In conclusion, the current path analysis provides evidence of the mediating role of social presence only in people’s general evaluation of social robots.

Figure 2. Interaction Effect of Embodiment and Loneliness in Experiment 1.

Note. The evaluation of interaction with Aibo was measured in 10-point semantic differential scale. Numbers above markers are means.
Consistent with Hypothesis 6, participants in the lonely group felt more social attraction to Aibo ($M = 4.57$, $SD = 0.370$) than participants in the non-lonely group did ($M = 3.95$, $SD = 0.8151$), $F(1, 28) = 7.184$, $p < .05$, $\eta^2 = .208$. In addition, a significant interaction effect between loneliness and embodiment in the evaluation of interaction with Aibo was found with two-way between subjects ANOVA, $F(1, 28) = 25.607$, $p < .001$, $\eta^2 = .599$ (see Figure 2). Specifically, under the disembodied condition, participants in the non-lonely group evaluated the interaction with Aibo more positively ($M = 7.64$, $SD = 0.5523$) than participants in the lonely group did ($M = 6.61$, $SD = 0.4600$). Alternatively, under the embodied condition, participants in the lonely group evaluated the interaction with Aibo more positively ($M = 8.86$, $SD = 0.5664$) than participants in the non-lonely group did ($M = 7.35$, $SD = 1.0817$).

Participants did not feel any difference in physical attraction between the physically embodied Aibo and the physically disembodied Aibo, $F(1, 28) = 1.49$, n.s. This non-significant finding suggests that the manipulation of two embodiment conditions is successful. Physical attraction should be same across two embodiment conditions because the physical shapes of Aibo in the two conditions are exactly same.

### 3.6. Discussion

A number of conclusions can be drawn from the results of Experiment 1. First, physically embodied social robots (PESR) are more attractive to people (H1 & H2). This result implies that physical embodiment is an important component in designing social robots although social robots are not particularly related to physical functions. People prefer interactions with physical social actors to interactions with virtual social actors.

Second, social robots are more socially attractive to lonely people (H6). This finding supports more diverse role of social robots and their market potential. Social robots can provide social companionship, thus can be used as therapeutic aids for lonely people. Indeed, social robots are not just toys for kids.

Third, physical embodiment yields higher social presence of artificial social robots than physical disembodiment (H4). The result implies that PESR influences people’s imagination of actual social actors positively. In addition, the social presence is the key mediating variable for the effect of physical embodiment in the general evaluation of social robots (H5). The result also implies that people’s social responses to artificial social robots are oriented toward imagined social actors. These findings support the computers are social actors (CASA) paradigm suggested by Nass and his colleagues [14, 10]. It also replicates the results reported by Lee and Nass [6].

Finally, an interaction effect between embodiment and loneliness was found in the evaluation of interaction with Aibo (see Figure 2 in Section 3.5.). The result shows that lonely people are more sensitive to PESR than non-lonely people. It implies that touch-input capability can be a potential factor that causes the major difference of effectiveness between physically embodied and disembodied social robots. The distinctive sensory difference between two embodiment conditions is touch. Participants in the physically embodied condition could touch and feel actual Aibo unlike participants in the physically disembodied condition where they saw a virtual hand touching Aibo. Participants in the lonely group might appreciate touch-input capability more positively than participants in the non-lonely group due to their relatively stronger needs for companionship.

However, we have to be cautious to draw a conclusion about the positive effects of touch-input capability in human-robot interaction from the results of Experiment 1. Physical embodiment has two major components: (1) visual embodiment; (2) touch. Of course, in future other human senses such as smell and taste can be incorporated into PESR. For now, however, the incorporation of other human senses is not the major concern for the development of social robots. There is no major difference between physical embodiment and disembodiment with regard to audio. Therefore, people can sense the physical embodiment of social robots by simply seeing them or by directly touching them. Due to the nature of "embodiment," the above two factors cannot be separated under a normal condition. In fact, touch is a nesting variable because there is no "touch" for physically disembodied conditions. As a consequence, we could not make separate conclusions on the effects of visual embodiment and the effects of touch-input capability based on Experiment 1.

In order to make a clear conclusion about the effects of PESR and to eliminate an alternative explanation of the effects of visual embodiment, we conducted Experiment 2 by manipulating physical embodiment only as a "visually embodied" factor. In Experiment 2 we focused solely on the effects of touch-input capability in human-robot interaction.

### 4. Experiment 2: Effects of touch-input capability in human-robot interaction

The importance of tactile communication in interpersonal relationship has been addressed in many studies. Nguyen et al. [11] found that “touching larger skin surfaces signified playfulness, warmth/love, and friendship/ fellowship.” Similarly, Burgoon et al. [3] also found that the combination of touch and high communicator valence produced the highest credibility and attraction ratings in their empirical study. These findings from interpersonal communication may hold up in a new type of relationship, human-robot interaction. Therefore, Experiment 2 was conducted to examine the effects of touch-input capability in human-robot interaction.

#### 4.1. Experiment design

The same 2 (embodiment vs. disembodiment) x 2 (lonely vs. non-lonely) between-subjects factorial analysis
of variance design with a total of 36 participants was used in Experiment 2.

4.2. Procedure

The procedures were same as Experiment 1, except the use of a new social robot. April, a prototype robot manufactured by Samsung Electronics, was used instead of Aibo in order to control touch-input capability in the physically embodied condition. April can play music and perform dance as programmed. Unlike Aibo in Experiment 1, we disabled all the sensors of April and instructed participants in the physically embodied condition not to touch April. By doing so, we could eliminate the potential effects of touch from the effects of PESR in Experiment 2. Therefore, the only difference between the embodiment and disembodiment conditions was whether or not participants saw actual dancing April or virtual dancing April.

4.3. Manipulation

Same as Experiment 1, two conditions of embodiment were manipulated. First, April was programmed to play a particular song and to perform a dance based on the song for two minutes. Then, the pre-recorded performance in a digital movie format was shown on a 17 inch flat-screen monitor for the physically disembodied condition.

4.4. Measure

In addition to the previous measures, three questions concerning the evaluation of music were asked using independent 10-point scale: how much did you enjoy hearing this music?; how likely would you be to recommend this music to your friends?; how likely would you be to download this music? (Cronbach’s α = .79).

Same six questions used in Experiment 1 regarding the general evaluation of April were asked (Cronbach’s α = .86). One question eliminated from the index used in Experiment 1 was distant/close. The question was regarded irrelevant because participants were not allowed to touch April in Experiment 2.

Same five questions used in Experiment 1 concerning social presence were asked (Cronbach’s α = .79). Once again, one question used in Experiment 1 was eliminated: impersonal/personal. Because April has an anthropomorphic shape of female body line with futuristic look, asking the question of impersonal/personal seemed to be meaningless.

Same three questions used in Experiment 1 were used to measure the assessment of public opinion (Cronbach’s α = .83).

Same six questions used in Experiment 1 concerning the evaluation of interaction with April were asked (Cronbach’s α = .92).

Same three questions used in Experiment 1 regarding social attraction of April were asked (Cronbach’s α = .92).

4.5. Results

Table 2 shows a full correlation matrix of the measured variables in Experiment 2. Again, hypotheses were tested with one-way between subjects ANOVA.

We are rather surprised to find that the results of Experiment 2 were either opposite to the results of Experiment 1 or not significant. More specifically, participants evaluated music more positively when they interacted with the physically disembodied April ($M = 6.92, SD = 0.8801$) than when they interacted with the physically embodied April, ($M = 5.23, SD = 1.9196$), $F(1, 28) = 11.78, p < .01, \eta^2 = .393$.

There was no significant difference between two embodiment conditions with regard to the feeling of social presence. The non-significant pattern shows that participants felt moderately stronger social presence when they interacted with the physically disembodied April ($M = 5.72, SD = 1.4609$) than when they interacted with the physically embodied April, ($M = 4.75, SD = 1.5117$), $F(1, 28) = 3.75, n.s.$.

Participants assessed other people’s evaluation of April more positively when they interacted with the physically disembodied April ($M = 6.62, SD = 1.2718$) than when they interacted with the physically embodied April, ($M = 5.26, SD = 2.0323$), $F(1, 28) = 7.15, p < .05, \eta^2 = .426$.

Table 2

<table>
<thead>
<tr>
<th>Measured Variables</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Evaluation of Music</td>
<td>.355</td>
<td>.390*</td>
<td>.515**</td>
<td>.602**</td>
<td>.481**</td>
<td></td>
</tr>
<tr>
<td>2. General Evaluation of April</td>
<td>.438*</td>
<td>.609**</td>
<td>.651**</td>
<td>.557**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Social Attraction of April</td>
<td>.529**</td>
<td>.643**</td>
<td></td>
<td>.684**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Assessment of Public Opinion</td>
<td></td>
<td>.721**</td>
<td>.642**</td>
<td></td>
<td>.773**</td>
<td></td>
</tr>
<tr>
<td>5. Evaluation of Interaction with April</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Social Presence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: * $p < .05$, ** $p < .01$ (2-tailed).
Lastly, participants in the lonely group felt more social attraction to April ($M = 3.65$, $SD = 1.4062$) than participants in the non-lonely group, ($M = 2.31$, $SD = 1.3525$), $F(1, 28) = 7.57$, $p < .05$, $\eta^2 = .263$. This finding, which replicates the result of Experiment 1, shows a solid evidence for the strong needs of social robots to lonely people.

4.6. Discussion

We can find a possible explanation for the results of Experiment 2 in post-experiment interviews with participants. Followings are excerpts from the in-depth interview with participants: “I thought it was going to talk to me.”; “I expected interaction such as sensing users’ movement.”; “I want it to have sensors for interaction rather than to do the same thing over and over again.”; “I expected it to talk to me. It appears to have personality but repeats the same thing, unsatisfying.”; “I expected it to say hi and shake my hands…”

As shown above, most of the participants expected to have some level of interactions with April when they first saw it because of its anthropomorphic shape. However, participants could only see April’s performance and were not allowed to touch it even in the physical embodiment condition in Experiment 2. Although a minimum level of interactivity was provided by allowing participants to push a button on a remote control to make April start its dancing performance, participants did not regard it as a meaningful social interaction. The interaction that participants had in Experiment 2 lacked sensory touch despite April’s highly anthropomorphic shape. According to the uncanny valley effect suggested by Mashi Mori, the subtle imperfection of the recreation becomes highly disturbing, or even repulsive (see [4]). Certainly, the anthropomorphic shape of April could set up high expectations [17]. However, the anthropomorphic shape without touch-input capability might lead to the sudden drop of participants’ high expectations to their frustration and disappointment, which, in turn, might result in the general negative effects of physical embodiment.

The results of Experiment 2 show that PESR does not always result in positive effects. We are surprised to find that PESR without touch-input capability causes negative effects. This finding in Experiment 2 suggests that it is important for physically embodied social robots to have touch-input capability. It also implies that the importance of tactile communication in interpersonal relationship holds up in a new type of relationship, human-robot interaction, as well. To put together, the effects of PESR may become synergetic when users are able to fully interact with social robots by touching and feeling them.

5. Conclusions

In summary, the findings of Experiment 1 elucidate the importance of physical embodiment in designing social robots. Physical embodiment usually enhances the feeling of social presence, which results in more positive evaluation of social robots. Furthermore, the findings of Experiment 2 indicate the relative effectiveness of PESR: without the power of touch-input capability and interaction, the effectiveness of PESR diminishes in human-robot interaction. Especially, Experiment 2 helped us to make a solid conclusion about the effects of touch-input capability in human-robot interaction by allowing us to separate two nesting component of physical embodiment: (1) visual; (2) touch. We hope the findings of this study shed light on the design of social robots, the importance of tactile communication in human-robot interaction, and the design of new interfaces for future technologies.

References


Personality-related differences in subjective presence

Jari Laarni¹, Niklas Ravaja¹, Timo Saari¹ and Tilo Hartmann²
¹Center for Knowledge and Innovation Research, Helsinki School of Economics, Finland
²Department of Journalism and Communication Research, Hannover University of Music and Drama, Germany
{lara@hkkk.fi, ravaja@hkkk.fi, saari@hkkk.fi, Tilo.Hartmann@ijk.hmt-hannover.de}

Abstract
Even though user-related variables are important determinants of presence, quite little is known about their role. The main aim of the present study was to examine differences in presence experience as a function of personality. Participants navigated through a multimedia presentation on a desktop computer. Half of them had the opportunity to navigate through the stimulus without interruptions, the other half were asked to generate random numbers whenever an audio signal occurred. Personality-related differences were shown to be an important determinant of presence. Especially, extraversion, impulsivity and self-transcendence were positively associated with presence ratings.

Keywords--- Spatial Presence, personality, emotions, attention.

1. Introduction

As complex media environments and virtual reality systems are becoming more popular, it is important to identify those people who are likely successful users of these systems [1,2]. Since there is evidence that feelings of presence are positively associated with performance in complex media environments, the task is to identify those people who are able to experience a higher sense of presence.

In fact, people seem to differ in their ability to experience presence, and characteristics of the media users are apparently important in determining the degree of presence feelings. However, as Lombard and Ditton [3] noted in their 1998 paper, very little research has been conducted on the question.

Several user-related variables are typically thought to have an impact on presence experiences, such as the user’s willingness to suspend disbelief, her/his knowledge of and prior experience with the medium and adaptability to new circumstances [3,4]. For example, Slater and Usoh [5] found that quick adapters take greater notice of their environment and experience a greater sense of presence than slower adapters. Other possible characteristics of media users that may influence their presence feelings are personality type, domain specific interest, cognitive capabilities, mood, age, gender, social class and culture [1,3,4,6].

There is a lot of evidence that dimensions of personality, such as extraversion-introversion and anxiety-calmness, are related to different components of neural functions, information processing skills, knowledge and real-world adaptation [7]. Personality factors may thereby also exert influence on complex mental states such as perceived workload [8], situation awareness [9], flow experience [10], reality judgment [11] and presence.

Since presence experiences are typically related to the use of complex computer technologies, it can also be asked whether personality characteristics exert impact on the use of information technology. There is controversial evidence regarding this issue. On the one hand, there is evidence that personality factors are not important, and they are, for example, not able to predict computer performance [12]. On the other hand, personality traits, such as introversion-extraversion, are related to many aspects of human-computer interaction (for a review, see [13]). And as Stanney et al. [6] have noted, personality may become more important during more complex interactions such as those experienced in virtual environments.

1.1. Personality-related differences in sense of presence

What is common to most definitions of presence is that, first, presence is related to ‘being there’ in one environment, and second, presence is related to the ‘perceptual illusion of nonmediation’ [14]. Lessiter et al. [15] and Schubert et al. [16] found three factors that were related to the state of presence: 1) Spatial presence which means a sense of physical placement in the mediated environment, interaction and control over different parts of the environment; 2) engagement which consists of a tendency to feel psychologically involved and to enjoy the content; and 3) naturalness which means a tendency to perceive the mediated environment as lifelike or real. Wirth et al. [17] have recently presented a two-stage model of the formation of spatial presence experiences. According to their model, presence experience is based, on the one hand, on the feeling of being physically present in the mediated environment, and, on the other hand, on the ability to act in this environment.

People differ in their ability to get involved and immersed and to feel present in the mediated world [18].
For example, people differ in their ability to orient towards motivationally significant stimuli [19] and to divide attention between different stimuli [20]. Witmer and Singer [18] have presented evidence that these differences exert influence on people’s presence feelings.

Several models of personality have been proposed that could potentially explain the differences in attentional engagement and presence experiences. Gray’s [21,22] three-arousal model proposes the existence of three independent systems, the Behavioural Approach System (BAS), the Behavioural Inhibition System (BIS) and the Nonspecific Arousal System (NAS), each of which has a neurophysiological substrate of its own. The BAS is responsive to conditioned signals of reward and conditioned stimuli associated with the cessation of punishment fostering approach behavior towards motivationally significant stimuli [22,23]. The typical emotions are positive emotions such as hope and happiness. The BIS is responsive to secondary aversive stimuli, and it is also activated by extreme novelty [22,23]. Typical behaviors related to the BIS are stopping actions, scrutinizing the environment, passive avoidance and giving up behaviors that are not readily reinforced. The typical emotion is anxiety.

BIS activity is, thus, associated with negative affect and BAS activity with positive affect [24]. According to Gray [25], individual differences in the activity and responsiveness of the BIS and BAS systems determine two major personality dimensions. Because of higher impulsivity, high BAS people presumably concentrate less on low-immersive media stimuli such as hypertext. They may rush quickly through the site in the pursuit of new possibilities and thus experience a lower level of presence than low BAS individuals. Also, they may engage more in distracting stimuli, and thus attentional distraction and a secondary task may have a larger effect on their feelings of presence.

In fear of punishment, people who have high BIS activity may concentrate more on low-immersive media stimuli, and thus they may experience a higher state of presence than those who have lower BIS activity. On the other hand, attentional distraction and a secondary task may have a larger effect on their sense of presence.

Eysenck [26] differentiated three traits, extraversion-introversion, neuroticism-stability and psychoticism-superego, and considered them as the basic dimensions of individual differences. Extraverts engage in many external activities, because they are able to process many stimuli at the same time. They, thus, possess more processing resources or they are better in allocating their resources than introverts [27]. Because of this, it is possible that extraverts experience higher sense of presence than more introverted individuals. In addition, if extraverts have more processing resources available than introverts, attentional distraction and the need to divide attention between two stimuli may have a smaller effect on their sense of presence.

Also neuroticism-stability may contribute to presence feelings. In Avila’s [28] study neurotics had more problems in shifting attention from new locations to previously revised locations. Neurotic introverts also had more problems in disengaging attention away from motivationally significant stimuli; neurotic extroverts, in turn, had more problems in disengaging attention away from positive stimuli. In general, neurotic people may thus experience a lower level of presence than those who are less neurotic, and a secondary task may have a larger detrimental effect on their sense of presence.

Zuckerman [29] defined sensation seeking as a trait defined by the seeking of novel and intense sensations and experiences and the willingness to take different kinds of risks for the sake of such experiences. Individuals who are high/low in sensation seeking may process information differently because of differences in arousal and attention. Those who score high on sensation seeking have better focused attention than low sensation seekers [31]. High sensation seekers may thus do better in a selective attention task for which certain stimuli must be attended to and others ignored; low sensation seekers, in turn, may do better in tasks that require them to divide their attention to several stimuli [30,31]. Overall, if high sensation seekers are better able to focus their attention to a particular media stimulus than low sensation seekers, they may experience a higher level of presence. Attentional distraction and a secondary task may also have a smaller effect on their sense of presence.

Impulsive individuals, in turn, are better able to shift attention throughout space [32]. Avila and Parcet [33], for example, have found that impulsive participants focused more on reward/expected targets than low impulsive individuals. Therefore, impulsive people may concentrate less on low-immersive media stimuli such as hypertext. Since they may rapidly navigate through the site in the pursuit of new possibilities, they may experience lower states of presence than those who are less impulsive. They may also pay more attention to distracting stimuli and allocate more attention to a secondary task than less impulsive people. As a result, the secondary task may have a larger detrimental effect on their sense of presence.

According to Cloninger et al. [34], self-transcendence refers generally to identification with everything considered as essential parts of a unified whole. This unitive perspective is described as acceptance, identification, or spiritual union with nature and its source. Self-transcendence has multiple aspects or stages: self-forgetfulness, transpersonal identification and spiritual acceptance. Self-transcendence, particularly self-forgetfulness, should be positively associated with dimensions of presence.

1.2. Aims of the present study

The main aim of the present study was to examine differences in presence experience as a function of personality. Another aim of the study was to investigate the relationships between different personality traits and the tendency to get involved and immersed in the mediated world.
It is possible that people who are more prone to experience positive or negative emotions have a tendency to focus attention more tightly to a stimulus that they are inclined to engage in, and then as a result, they will experience a higher level of presence. We hypothesized that experiences of presence, i.e. spatial presence and attentional engagement, are positively associated with high BIS activity, extraversion, sensation seeking and self-transcendence. Presence would, in turn, be negatively associated with high BAS activity and impulsivity. We also expected that the secondary task would have a larger detrimental effect on presence for high BIS and BAS individuals, introverts, impulsive people and high sensation seekers.

In order to experimentally induce different levels of presence a dual-task method was used. We assumed that the secondary task will reduce the amount of attentional resources that subjects can allocate to the processing of the media stimulus, which will then reduce the level of presence they experience.

2. Method

2.1. Participants

Eighty volunteers participated in the experiment (51 females, 29 males). The mean age of the participants was 24 with a range between 18 and 39. They were ignorant of the purpose of the study before participating. Participants were selected in the order of their announcement to an email message. They were paid for their participation (each one received two movie tickets, total value about 13€).

2.2. Stimulus

'The Art of Singing' CD-ROM (Nothing Hill Publishing Limited 1996) based multimedia stimulus was applied. It is a commercial multimedia presentation 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 stimuli were generated on the face of Apple Multiple Scan CRT (17" in diameter) with a Power Macintosh 7200/90 computer. The number of color was set to 256, and the screen resolution was set to 800 x 600. Sounds were presented through Multimedia Speakers (SP-628). A standard computer mouse was used for input.

In the dual-task condition, the distracting audio signals were presented through loudspeakers that were located behind a participant. Five different signals (e.g., alarm, train and school bell) were used. They were presented in a random order with random intervals between the single signals. There were 12 signals in a 10-minute test session.

2.3. Procedure

A between-subjects design was used; half of the participants had the opportunity to navigate through the hypertext presentation without interruptions, the other half of the participants were asked to generate 3-digit random numbers whenever a defined signal occurred during the reception of the media stimulus. This task reduced the amount of attentional resources that participants could devote to the processing of the stimulus.

2.4. Presence measures

2.4.1. MEC-SPQ The MEC Spatial Presence Questionnaire (MEC-SPQ) consists of nine scales that measure the different concepts integrated in Wirth et al.'s [17] theoretical model. It includes process factors (Attention Allocation, Spatial Situation Model, Self Location and Possible Actions), variables relating to states and actions (Higher Cognitive Involvement and Suspension of Disbelief) and trait variables (Domain Specific Interest, Visual Spatial Imagery and Absorption). The items related to spatial presence experiences, i.e., Self Location and Possible Actions, were analyzed, and their scores were summed before they were entered into the analysis. Two of the three trait variables (Domain Specific Interest and Visual Spatial Imagery) were also entered into an analysis that considered the possible relationships between attributes of immersive tendency and other individual traits.

2.4.2. ITC-SOPI Since the initial English version of MEC-SPQ was used in this study, the participants also filled out another presence questionnaire, the Independent Television Commission Sense of Presence Inventory (ITC-SOPI), which has been widely applied in presence studies [15]. The ITC-SOPI measures four dimensions of presence: 1) Sense of physical space which means a sense of physical placement in the mediated environment, interaction and control over different parts of the environment; 2) Engagement which consists of a tendency to feel psychologically involved and to enjoy the content; 3) Naturalness which means a tendency to perceive the mediated environment as lifelike or real; and 4) Negative effects, that is, a tendency to have adverse physiological reactions [15]. The ITC-SOPI consists of 43 items. It is recommended that each scale is analyzed separately, since each scale is differentially sensitive to manipulations of particular determinants of presence. Only Sense of physical space, Engagement and Naturalness subscales were used in this study.

2.5. Trait measures

2.5.1. Immersive Tendencies Questionnaire (ITQ) Wittmer and Singer’s [18] Immersive Tendencies Questionnaire (ITQ) is aimed to examine individual differences in the ability to experience presence. It thus
concentrates on the user characteristics. For example, it aims to measure the capability or tendency to be involved or immersed, and the ability to focus on a particular activity. It consists of three subscales, Focus, Involvement and Games. According to Witmer and Singer [18], the Focus items are related to mental alertness, participants’ ability to concentrate on enjoyable activities and their ability to block out distractors. Involvement items, in turn, are related to the participants’ propensity to get involved passively in some activity, and the Games items are asking how frequently participants play video games and whether they get involved to the extent that they feel they are inside the game. Some studies have found a significant correlation between presence ratings and ITQ scores [18].

2.5.2. BIS/BAS scales The BIS/BAS Scales are a 20 item self-report questionnaire assessing individual reactivity to reward and punishment [35]. It includes four subscales: 1) BIS (fearfulness and reactivity to negative events), 2) Drive (the persistent pursuit of goals), 3) Fun Seeking (the desire for novel rewards and the inclination to eagerly approach such rewards), and 4) Reward Responsiveness (the positive reaction to reward and its anticipation). The global BAS score was the sum of Drive, Fun Seeking and Reward Responsiveness scores. BIS items are aimed to reflect the experience of anxiety in situations in which there are signs of possible punishment. BAS items tap a strong pursuit of goals, responsiveness to rewards, a tendency to seek out new potentially rewarding experiences and to act quickly in the pursuit of desired goals.

2.5.3. Eysenck Personality Questionnaire-Revised, Short Form (EPQ-R Short) The EPQ-R Short [36] is a 48 item self-report questionnaire assessing extraversion (E), neuroticism (N) and psychoticism (P). The EPQ’s scales have been revised several times while attempting to produce scales that assess orthogonal and reliable personality dimensions [37].

2.5.4. Zuckerman-Kuhlman Personality Questionnaire-ImpSS (ZKPQ-ImpSS) The original Zuckerman-Kuhlman Personality Questionnaire (ZKPQ) is a 99-item questionnaire aimed at the evaluation of the five-factor model [29]. The impulsivity and sensation seeking scales comprise 19 items. The impulsivity items describe a lack of planning and a tendency to act impulsively without thinking; the sensation seeking items describe a general need for thrilling and exciting experiences, a preference for unpredictable situations and friends and the need for change and novelty.

2.5.5. Self-Forgetful vs. Self-Conscious Experience scale The self-forgetfulness trait was measured with the 11-item Self-Forgetful vs. Self-Conscious Experience subscale of the Self-Transcendence scale included in the Temperament and Character Inventory (TCI) [34]. The TCI is a self-report personality questionnaire based on Cloninger's psychobiological model of personality [34].

2.6. Data analysis

Since there were both categorical and continuous independent variables, the data were analyzed using the General Linear Model Univariate/Multivariate procedure in SPSS (Statistical Package for the Social Sciences).

Continuous independent variables were used as covariates while Condition was used as a between-subjects variable. MEC-SPQ’s Spatial presence and ITC-SOP’s Sense of Physical Space, Engagement and Naturalness scales were used as dependent measures. Multiple regression analyses were used to investigate the relationships between the Immersion Tendency Questionnaire and different individual traits.

3. Results

To investigate the relationship between presence scales and different personality-related scales a series of multivariate ANOVAs was carried out. The mean scores, standard deviations and Cronbach alpha coefficients for all the predictor and outcome variables are presented in Table 1. The scores for those who participated in the single-task and dual-task condition are presented separately. Except for the Psychoticism subscale (Cronbach’s alpha = 0.43), the alpha coefficients for all predictor variables were high ranking from 0.74 to 0.84. The alpha coefficients for outcome variables were also high ranking from 0.77 to 0.95.

3.1. BIS/BAS scales

BAS had a significant main effect in predicting Engagement, $F(1,75) = 5.3$, $p < 0.05$. BAS was negatively associated with Engagement: the participants who had high BAS scores felt less engaged than those who had lower scores. The Condition x BAS interaction was significant for all scales, Spatial Presence (MEC-SPQ): $F(1,75) = 5.4$, $p < 0.05$; Sense of Physical Space (ITC-SOPI): $F(1,75) = 6.4$, $p < 0.05$; Engagement: $F(1,75) = 12.6$, $p < 0.001$; Naturalness: $F(1,75) = 9.9$, $p < 0.01$. The secondary task had a different effect on high and low BAS participants. For low BAS participants the sense of spatial presence, engagement and naturalness were higher in the dual-task condition, while for high BAS participants feelings of presence, engagement and naturalness were higher in the single-task condition.

BIS had no significant effect in predicting presence, $p > 0.1$. The Condition x BIS interaction was, however, significant for the three scales of the ITC-SOPI, Sense of Physical Space: $F(1,75) = 5.0$, $p < 0.05$; Engagement: $F(1,75) = 11.6$, $p < 0.001$; Naturalness: $F(1,75) = 9.1$, $p < 0.01$. This interaction approached significance for MEC-SPQ’s Spatial Presence scale: $F(1,75) = 3.7$, $0.05 < p < 0.1$. For high BIS participants, the secondary task had
Table 1. The mean scores, standard deviations and Cronbach alpha coefficients for the predictor and outcome variables.

<table>
<thead>
<tr>
<th>Predictor/Outcome Variable</th>
<th>Single Task</th>
<th>Dual Task</th>
<th>Cronbach's Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BIS/BAS Scale</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BIS</td>
<td>2.10</td>
<td>2.10</td>
<td>0.76</td>
</tr>
<tr>
<td>BAS</td>
<td>1.97</td>
<td>1.93</td>
<td>0.74</td>
</tr>
<tr>
<td><strong>EPQ</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psychoticism</td>
<td>2.40</td>
<td>2.40</td>
<td>0.76</td>
</tr>
<tr>
<td>Extraversion</td>
<td>7.58</td>
<td>7.68</td>
<td>0.83</td>
</tr>
<tr>
<td>Neuroticism</td>
<td>5.68</td>
<td>5.10</td>
<td>0.76</td>
</tr>
<tr>
<td><strong>ImpSS</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Impulsivity</td>
<td>2.99</td>
<td>2.85</td>
<td>0.84</td>
</tr>
<tr>
<td>Sensation Seeking</td>
<td>3.16</td>
<td>3.23</td>
<td>0.84</td>
</tr>
<tr>
<td>Self-Transcendence</td>
<td>3.18</td>
<td>3.00</td>
<td>0.84</td>
</tr>
<tr>
<td><strong>ITQ</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Focus</td>
<td>4.91</td>
<td>4.64</td>
<td>0.80</td>
</tr>
<tr>
<td>Involvement</td>
<td>4.80</td>
<td>4.60</td>
<td>0.77</td>
</tr>
<tr>
<td>Games</td>
<td>3.89</td>
<td>2.96</td>
<td>0.78</td>
</tr>
<tr>
<td><strong>MEC-SPQ</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spatial Presence</td>
<td>2.56</td>
<td>2.39</td>
<td>0.95</td>
</tr>
<tr>
<td><strong>ITC-SOPI</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spatial Presence</td>
<td>2.35</td>
<td>2.21</td>
<td>0.92</td>
</tr>
<tr>
<td>Engagement</td>
<td>3.13</td>
<td>2.78</td>
<td>0.92</td>
</tr>
<tr>
<td>Naturalness</td>
<td>2.57</td>
<td>2.51</td>
<td>0.77</td>
</tr>
</tbody>
</table>

3.2. EPQ scales

Extraversion was positively associated with Spatial Presence, Spatial Presence (MEC-SPQ): F(1,76) = 5.9, p < 0.05; Sense of Physical Space (ITC-SOPI): F(1,76) = 7.2, p < 0.01. That is, the extraverts had higher scores on presence scales than those who were more introverted. The association between Extraversion and Engagement and between Extraversion and Naturalness approached significance, Engagement: F(1,76) = 4.0, 0.05 < p < 0.1; Naturalness: F(1,76) = 3.4, 0.05 < p < 0.1. Neuroticism and Psychoticism had no significant effects in predicting presence, all p’s > 0.1.

3.3. ImpSS scale

Impulsivity had a significant main effect in predicting Engagement, F(1,73) = 4.1, p < 0.05, and the association between impulsivity and engagement was positive in nature. The effect of Impulsivity on Spatial Presence (MEC-SPQ) approached significance, F(1,73) = 3.6, 0.05 < p < 0.1. None of the other effects were significant, all p’s > 0.1.

3.4. Self-forgetfulness scale

Self-forgetfulness had a significant main effect in predicting Spatial Presence and Engagement, Spatial Presence (MEC-SPQ): F(1,75) = 8.9, p < 0.01. Sense of Physical Space (ITC-SOPI): F(1,75) = 7.7, p < 0.01; Engagement: F(1,75) = 15.2, p < 0.001. Self-forgetfulness was positively associated with Spatial Presence and Engagement.

3.5. ITQ scales

Focus had a significant main effect on Spatial Presence and Engagement, Spatial Presence (MEC-SPQ): F(1,76) = 16.4, p < 0.001; Sense of Physical Space (ITC-SOPI): F(1,76) =11.7, p < 0.001; Engagement: F(1,76) = 14.4, p < 0.001. The correlation between Focus and presence scales was positive. Involvement had a marginally significant effect in predicting Engagement, F(1,76) = 3.2, 0.05 < p < 0.1.

3.6. Relationships of personality measures with the Immersive Tendency Questionnaire (ITQ)

Multiple regression analyses were carried out separately for the three dependent variables (Focus, Involvement and Games). A stepwise approach that
enters the variables into the equation according to the strength of their association with each primary dependent variable was used. Predictor variables included BIS, BAS, Extraversion, Neuroticism, Psychoticism, Impulsivity, Sensation Seeking, and Self-Forgetfulness. Also, Domain Specific Interest and Spatial Imagery Ability from the MEC-SPQ were entered into the analysis.

The multiple regression analysis showed that the predictor variables together explained 31% of the variance in the Focus, $F(3,74) = 11.0, p < 0.001$ (adjusted $R^2 = 28\%$). Three variables contributed statistically significantly to the equation, Self-Forgetfulness ($\beta = 0.40, p < 0.001$), BIS ($\beta = 0.22, p < 0.05$) and Extraversion ($\beta = 0.22, p < 0.05$).

The predictor variables together explained 33% of the variance in the Involvement, $F(3,74) = 12.3, p < 0.001$ (adjusted $R^2 = 31\%$). Again, the same three variables met the inclusion criteria and contributed significantly to the regression equation, Self-Forgetfulness ($\beta = 0.40, p < 0.001$), BIS ($\beta = 0.35, p < 0.001$) and Extraversion ($\beta = 0.21, p < 0.05$).

Regression analyses predicting the Games dimension showed that the predictor variables together explained 7% of the variance, $F(1,74) = 5.1, p < 0.05$ (adjusted $R^2 = 5\%$). The only variable that met the inclusion criteria was Impulsivity, $\beta = -0.26, p < 0.05$.

4. Discussion

Our results suggest that personality-related variables influence subjective presence. Especially, BIS/BAS scales were quite successful in predicting presence ratings. In accordance with the hypotheses, the BAS scores were negatively correlated with Engagement. That is, those who had higher scores on BAS felt less engaged than those who had lower scores. It was also found that those who had high BAS activity were more engaged in the single-task condition than in the dual-task condition, and the secondary task had a large detrimental effect on their sense of engagement. Contrary to that, those who had lower BAS activity were even more engaged in the media stimulus in the dual-task condition than in the single-task condition. Perhaps, in order to keep the performance at the same level in the divided attention condition, the low BAS participants tried to attend more tightly to the media stimulus and allocated more attentional resources to the task. They were presumably also more worried about their performance than those who had higher BAS activity.

Even though BIS had not a significant main effect in predicting presence, the interaction between BIS and Condition was significant, however. The difference in presence feelings between conditions was larger for those who had a higher BIS activity. Since the high BIS participants presumably felt themselves more anxious when the demands of the task were increased, they reported considerably lower levels of spatial presence and engagement in the dual-task condition.

As hypothesized, extraverts had higher scores on presence scales than those who felt themselves more introverted. If extraverts have more processing resources than introverts, they may process more information per time unit and thus feel more present in a mediated environment. Contrary to the hypotheses, neurotics did not experience a lower level of presence than those who are less neurotic.

Overall, the EPQ scales were not very successful in predicting presence. Impulsivity was, however, positively associated with presence. It is possible that the positive effect of impulsivity is related to the amount of processing resources available. As in case of extraverts, impulsive individuals may have more processing resources than less-impulsive people, and thus, they may be able to engage more deeply in a mediated environment.

The effect of self-transcendence is quite clearcut. Self-Forgetfulness and transpersonal identification seem to be important determinants of subjective presence. Those who get high scores on these scales are better able to identify themselves with a media stimulus.

Witmer and Singer’s [18] ITQ has been developed to examine individual differences in state of presence. If higher scores on these scales reflect a greater tendency to become involved or immersed, then those participants who score high on these scales should report more presence. In fact, Witmer and Singer [18] have found a positive correlation between ITQ and presence. Both the Focus and the Involvement subscales of the ITQ were positively associated with Extraversion and Self-Transcendence. In general, extraversion and some kind of self-Forgetfulness seem to be associated with the ability to focus one’s attention to the mediated environment and the ability to get involved with media stimuli.

There are two types of breaks in presence: First, our attention may be shifted away from the mediated environment towards the real world; second, attention may be shifted away from media stimuli to internal thoughts [38]. In the present study, the participants, thus, had to assign a priority to responding, on the one hand, either to stimuli from the real world or to stimuli from the hypertext, and on the other hand, either to environmental stimuli or to those stimuli that seem to be internal. How much they at every moment paid attention to these different types of stimuli may, at least in part, depend on personality factors. For example, both anxious and impulsive participants may have had difficulties in focusing attention on hypertext stimuli because of greater distractibility by audio signals and peripheral stimuli. Moreover, since it was necessary to succeed in the secondary task, they were perhaps less eager to focus on neutral hypertext stimuli, and instead paid more attention to distracting audio signals.

Attentional processes thus seem, at least in part, play a mediating role between personality and presence: People differ in their ability and willingness to pay attention to the mediated world, and these differences exert influence on people’s presence feelings. The
results, however, do not give a definitive answer to the question by which way the effect of personality factors is mediated. For example, according to the two-stage model of Wirth et al., there are several possible mediated factors that exert influence on the sense of presence [17]. More information can be gathered by investigating the relationship between more objective indicators of presence and personality. In the future, we will more systematically study the effect of personality variables on different dimensions of the MEC-SPQ. Our aim is also to investigate the impact of personality on presence experiences in different types of media environments.

Conclusions

Since presence is a subjective phenomenon, characteristics of individuals apparently influence presence [18]. Our study is one of the first ones to show that personality-related factors are an important determinant of presence. Especially, extraversion, impulsivity and self-transcendence were positively associated with presence. Attentional processes seemed to play a mediating role. People who got high scores on these scales paid more attention to the media stimulus and were also more deeply involved in the mediated environment.

Acknowledgements

This research has been funded by the European Community, FET Initiative “Presence Research Activities”, Project “Presence: MEC (IST-2001-37661). We thankfully acknowledge the EC’s support. We would also like to thank two anonymous reviewers for helpful comments on the manuscript.

References


Dynamics as a common criterion to enhance the sense of Presence in Virtual environments

Annie Luciani
ICA laboratory, Institut National Polytechnique de Grenoble, France
\{Annie.Luciani@imag.fr\}

Abstract
This paper is a theoretical dissertation on the role of the matter and its representations through dynamics in the organization of the multisensory space. We assume that the mechanical matter, which originally produces the multisensory-handled events, is an invariant identified from the spatio-temporal correlation between visual, auditory and tactilo-proprio-kinesthetic perceptions. We illustrate this assumption by several examples using virtual objects modeled by means of physically-based formalism, able to produce dynamically plausible and consistent multisensory events. We then examine frontiers between ambiguous notions as: geometry vs. dynamics in visual believability, optical matter vs. mechanical matter in contour identification, interaction vs. morphology or topology vs. force in complex manipulations. We conclude that, a challenge to enhance the presence of virtual objects is to be able to instill in their models a just sufficient clue of evoked materiality using minimal physically-based representations.

Keywords: Virtual objects. Multisensory simulations. Force feedback manipulations. Physically-based models. Dynamic consistency.

1. Preliminary observations

Despite the huge quantity of developments in Computer Graphics and Computer Sounds and Music, allowing to reach a high degree of visual and auditory realism in synthetic images and sounds, this realism seems not sufficient to trigger spontaneously the feeling of the Presence of such artifacts. The possibility to handle them, as in conventional interaction, i.e. to link action inputs to visual and auditory outputs, improves this feeling. But, even in the best implementation of these types of sensory-motor rendering, allowing to say “Yes, very good technical implementation”, the feeling of Presence will remain “asymptotically unreached”.

Even so, this feeling of “imperfectly reached being here” does not exist in front of recorded sounds and images, even if their quality is bad or degraded as in ancient recordings of singers or political speeches. Listening them on old acetate records bring them immediately and emotionally present in our space.

We will assume here that the sense of Presence emerges from the identification of the material object that is behind its phenomenological expressions. This material object is playing as a non pre-existing invariant that will be identified from experienced specific correlations in the sensible phenomena material object.

This assumption is suggested by the well-known fact that the feeling of a virtual object through a force feedback device during its manipulation convinces us immediately of the reality or the presence of such object. Even in the absence of other sensorial returns, visual and/or acoustical, and even if the object was simply or roughly rendered, suddenly, a strong piece of reality undoubtedly emerges for the experimentalist, during this type of sensory-motor experiment. Here, the consistency between action and perception is represented by forces and supported by the matter of the objects, allowing us to assume that presence cannot emerge without some clue of materiality, in other words, without some clue of energetic consistency or physically – based coherence in synthetic artifacts.

2. Is Presence a new question?

The distinction of what is real and what is non-real is an usual and long lasting question of philosophy as well as of physics. Recently, in his theory of Veiled Reality, Bernard d’Espagnat [1] points out that in Physics, the reality remains intrinsically unknowable in details but the knowledge developed by physicists as description of the phenomena, enlightens the structure of an underlying reality. Remembering that psychology was in the past a part of philosophy and that it joined the fields of experimental sciences recently, with psychologists as P. Piaget, we can answer that the problem of Presence, considered from these points of view, is not a novel question.

No explicit problem of Presence occurs no longer whenever human beings manipulate real objects, directly or indirectly through mechanical instruments. In teleoperation, since objects are mechanically teleoperated, as in the manipulation of blocks of nuclear matter through a mechanical pantograph, feeling it mechanically and seeing it through the glass that separates the two spaces, the immediate and trivial presence of objects continues to be
felt by the experimenter. When this direct physical communication has been replaced by electrical communication between the two spaces, the space of the user and the space of the task, the physical continuity of both has been broken causing the lost of the trivial sense of presence of each space for the other.

Similarly, in the field of sensorial data production, representation and transmission, any explicit problem of presence appears, when the sensorial data were provided by real objects, directly or indirectly through sensors (microphones, telephones, cameras, etc.). Since the 50’s, with the demonstration of Shanon’s theorem and its implementation in digital to analog converters, real sensorial data could be produced, “ex nihilo”, i.e. without any real objects, by abstract and symbolic entities such as numbers and algorithms.

In both cases, the primary properties that have been lost are the same: those that are related to the “materiality” of the manipulated real objects or recorded phenomena produced by real objects.

3. Experimental context, aims and methodology

In the real world, objects are a “single entities” interacting mechanically between themselves, producing correlated changes in each of them, without any input – output causality. Due to the action-reaction principle, it is strictly impossible, to distinguish of whom is acted and whom is acting on. Real objects are not input-output systems. Conversely, there are two cases in which the oriented input-output paradigm is “naturally” implemented: the human machine and the electrical (or electromechanical) machines. In contrast to a physical object, these machines are “broken” into four components: (1) sensors – (2) processing – (3) actuators embedded in a (4) mechanical morphology.

Thus, when a physical object “physically interacts” with a human being, two completely different systems are in vis-à-vis, one that is not a sensor-processing-actuator system (figure 1 on the left) and one that is (figure 1 on the right). As pointed out by the interrogation mark, the observation of such closed-loop system between these two types of machines (mechanical object and sensor-actuators systems) is not trivial.

Faced to this theoretical difficulty, studies in human action-perception system are usually performed by trying to understand how the multisensory signals received by human sensors and emitted from human mechanical and vocal actuators are interpreted and/or correlated in humans through the neuro-cognitive processing, i.e. by considering only the human side of the entire system, via for example the studies on cross-modal transfers or intersensoriality. Such phenomenological approaches are recently improved by the concept proposed by Stoffregen and Bardy [2] of “Global array” as an ambient multisensory space array considered as a whole. Their general hypothesis is that perception consists of sensitivity to patterns in such global array, more than to patterns in single-energy arrays.

The Gibson’s fundamentals [3] of the genuine link between action and perception on the human side could be faced to the Leroy-Gourhan’s works [4] who demonstrated that humans and physical world act reciprocally and are modified simultaneously, one by the other.

Replacing the real world by artifacts as those produced by computerized electromechanical machines (as VR is) allows to elicit this face-to-face situation and to experiment it. As shown in Figure 2 (on the left), computerized electromechanical machines are able to represent a real object by an input-output paradigm, similarly to the human machine. They are composed of sensors – processing - actuators in which the computational process through which sensors and actuators are linked necessarily plays the role of a representation of the physical material object.

Thus we are able to experiment what could be the objective properties of the process that produces sensorial handled data, i.e. what could be the genuine correlation between these data able to trigger the sense of presence and of the believability of virtual or distant worlds. Further, we have thus at our disposal the experimental means to catch the role of the matter as an invariant producing such phenomenological correlations allowing us to address the question of ”how the sense of the tangibility could be cognitively built?".
The methodology we will apply is exclusively qualitative by direct sensible appraisal by users, like that used by a musical instrument maker with instrumentalists. It is more a pragmatic approach of clearly cut user’s agreement and satisfaction than measurements approaches [5], either subjective or behavioral. Our methodology is based on “analysis by synthesis” method proposed by J.C. Risset [6] in musical psychoacoustics, or similarly on the “understanding by building” method as proposed in cognitive and developmental robotics [7]. We will build computerized electromechanical artifacts, by means of physically-based models, a priori able to allow us to functionally manipulate the dynamical features. “Functionally” means that physically-based models are not understood as a model of natural phenomena (Physics for Physis1), with its correlated feature of realism, but as a formal algebra, able to model relevant sensible observable features of classes of dynamic phenomena in a consistent way.

The methodology of designing such models is based on phenomenological top-down analysis. The first stage of the modeling process is to specify these relevant dynamic features.

The first stage is then to specify a minimum set of sensible patterns that could be relevant to address the tangibility of the underlying material causality from the observation of sensorial events. The second stage is to generate this set of patterns by a simulated physically-based model and to verify objectively if the produced artifact renders them correctly. Finally, we verify qualitatively, through unanimous answers, if these patterns play cognitively and perceptively the same role as analyzed previously in reality, during the specification stage.

4. Role of the matter in images, sounds and actions

4.1. Geometry vs. Dynamics in visual events

Except in images produced exclusively by geometrical processes (geometrical drawings or synthetic 3D images), in all other cases, images engrave motions. Even if objects are at a greater scale of time than usual evolving phenomena such as mountains, trees, etc., expressing the immobility, we can remark that the morphological features (the shape, the texture, etc…) contain more or less explicitly the trace of the evolution. Look at the figure 3 that shows (1) a fossil shell, (2) a geometrical representation of the fossil shell with a logarithmic spiral. It appears clearly that in the photograph representing the shell, the trace of the time is explicit and is a relevant feature in the interpretation of the object as a real object. The immobility itself has to be seen as a state of the motion.

The major difference between the photograph of the shell and its geometrical representation is that the time is completely absent in the last. This difference sparkles to our eyes by means of very small details as the local variations of the shape that point out the fact that, engraved in this shape, a physical evolution (the fossilization) happened with its spatio-temporal random features. Thus, if we want to produce a believable virtual fossil, must we emphasis the geometrical organization of the shape or the features that address the labor of the time?

![Figure 3. Dynamics vs. Geometry](image)

From this observation, it appears that the critical frontier in visual representation is not the distinction between morphology (shapes) and rendering (light) as usually considered in Computer Graphics, but between optical matter, represented by electromagnetic field, and mechanical matter represented through forces, in which the first produces pure visual features (color, shadows, etc.) and visual shape and the second produces mechanical shapes and motions. Visual features are then related more to the geometry of the space, since mechanical shapes and motion have to be represented by dynamics2.

4.2. Optical shape vs. mechanical shapes

In the previous paragraph, shapes are on the two sides of electromagnetic features and of mechanical features. This points out the underestimated ambiguity of the notion of shape. Shapes are usually considered in their geometrical features and so speaking they can be extended to all the spatial properties exhibited by an object: shape, size, orientation, and texture considered as micro-local properties of the contour of an object. There is a lot of work addressing the question of the recognition of such parameters, the considered senses being the sight and the touch. It is often considered that except the texture that is sensed equally by the touch and the vision, the others are more reliably encoded by the visual than the haptic system [8]. Developmental psychology points out other results as those in very young infants, when transfer from touch to vision and not from vision to touch is observed in the recognition of shapes (prism or cylinder).

Does it mean that there are two notions of shapes, one purely geometric, more related to vision, and another “physical”, more related to resistant matter, the texture?

Footnotes:
1 “Physis”, in ancient Greek, means Nature, as “Being given (to the humans)”. “Mathematê” means “being done (by the humans)”.
2 In ancient Greek, “Dynamê” means “forces” and “Kinema” means “motion”. Dynamics is a representation of systems that generates kinematical behaviors.
being the frontier between the two spaces? Indeed, shapes have, as the Janus figure, two faces or two determinants. They emerge from two completely different processes, optical and mechanical, and thus, a single object can paradoxically exhibits several shapes, or several “contours”: the visual shape and the mechanical shapes.

More, the visual shape and the mechanical shapes of a single object have no reason to be always identical. Several situations illustrate this paradox. A rainbow or the mirage of an oasis in the hot desert has a visual shape but doesn’t have mechanical contour. We can traverse them or walk through them. Conversely, a perfectly transparent door has not a visual contour but has a hard mechanical shape. The visual shape is sensed by eyes whereas the mechanical shape is sensed by the body.

Basically, the visual features are nothing else but the singularities of the interaction between photons and electromagnetic matter. The visual shape (the visually detected flatness, the visually spherical shape etc…) is the geometrical locus of the spatial singularities of the interaction light – optical matter. Thus, visual events are intangible. Other classical examples could be geometrical drawing and synthetic 3D images produced by pure geometrical representations.

In usual rigid objects, the visual shape seen by the eyes is at the same spatial location as the mechanical shape “seen” by the body. Although these objects are usual, nevertheless, they represent specific cases where the matter is 100% (99,99%) mechanically rigid and simultaneously 100% (99,99%) electromagnetically rigid (opaque). But what about flames, rainbow, water, fluids, translucent pastes, glasses etc?

Furthermore, what about objects like cat fur or hair, that are not 100% (99,99%) mechanically rigid, and thus exhibit several mechanical contours. In other words, and in a funny way, all what it is happening in terms of “contour” as a primary cue of space organization, depends on the percentage of the optical and of the mechanical rigidity.

More, a thing that could be considered as a single object can exhibit several mechanical contours. If you put a force sensor on the palm of the hand when stroking your cat, the force detected will be very low when the hand is in the fur, higher when it is on the deformable skin and higher when it is touching the skeleton. This means that a single entity - your preferred pet - may exhibit several mechanical contours, described by several thresholds in the singularities of the physical interaction.

4.3. And what about sounds in thebelievability of virtual objects?

Audition is, as vision, exteroceptive perception. It is why, in Virtual environments, the sounds are often used for the localization of sound sources through 3D auditory representations. But sounds convey another basic function: the identification of the mechanical properties of the sounding object (heavy, light, resistant, deformable, metal, wood, paper sheet, etc…) and of its interaction with the physical environment (more or less hard shocks, sticking collisions, friction, etc…), in the aim of recognizing the objects as well as their evolution (triggering, stopping, pursuing, etc…).

This function is widely studied in computer music and in digital musical instruments [9][10]. It is underestimated in Virtual Environments as well as in psychological studies, which prefer to point out the role of the audition in abstract activity as the discrimination of numerosities [11][12], and consequently its link with the vision.

Sound does not exist without mechanical matter with inertia (not neutrino) and minimal rheological properties, as at least elasticity. The minimal system to produce sound is a second order differential equation system. Sound is the mechanical behavior of an inertial and rheological matter and its morphological and topological organization. It encodes in a single signal all the properties of the material object. In addition, it conveys all the properties of this mechanical matter: material, structure, etc… on the larger spectra of temporal characteristics (from some Hz to several tens of KHz). From this point of view, it is the best “distant” and “diffuse” representation of what the material object is: a kind of “exteroceptive sense of touch”.

4.4. Interaction as exchanged actions

Action is the modality that had been deeply transformed with the electrification of instruments. In computerized environments, action started with interactivity. Since Evans & Sutherland who introduced the manipulation of virtual objects in 1963 [13], the computer interactivity has had huge developments. Nevertheless, since the introduction of force feedback devices, the action has been restricted only to one type, called by C. Cadoz “semiotic function” [14][15], in his typology of gestures. This typology analyses the gesture according to three functionalities:

• The semiotic function: A pure semiotic function appears in action during which the tactile-kinesthesia perception of an external object can be neglected, as it is the case in : free gestures, pointing, sign language, gestures which accompany the speech, the gesture of musical conductor, etc…

• The epistemic function: a pure epistemic function appears in the pure tactile activity to know objects: contours, texture, orientation, temperature, etc. The associated action is mainly an exploratory action, characterized by the fact that the energy produced and exchanged with the explored object is negligible [16].

• The ergotropic function appears during a sensory-motor activity in which the physical energy exchanged between the two interaction bodies (human and manipulated objects) (1) is not negligible in the performance and (2) is engraved in the produced phenomena. For example, in the pointing gesture or in typewriting, that are pure semiotic actions, the dynamic of the gesture is not usable for the aim of the action (pointing, typewriting). Similarly, the palpatation of an object does not transform this object. Conversely, when we are molding a paste, the dynamic of the molding is engraved in the paste. When we are plucking a guitar string or when we are bowing a cello string, the produced sounds
engrave (or encode) the dynamic of the action as well as the name exchanged between human and objects, as a main feature of the sounds. Differently from the epistemic and semiotic functionalities, ergonomic functionality cannot be implemented without mechanical matter.

These three functionalities – semiotic, epistemic, ergonomic - are more or less merged in usual actions. Nevertheless, in computer environments, the most part of the developments are devoted to the development of devices and processes either for the semiotic function (pure sensors as sticks, keyboards, mouse, motion capture devices, etc…), or for the epistemic function (tactile devices, haptic display, data haptization). In the hypothesis of the central role of the matter (or the material features of the world) in the recognition and in the mental reconstruction of the tangibility of virtual or represented objects, the ergonomic function will have to play a central role. The transducers able to convey this ergonomic interaction are only the force feedback gestural transducers.

4.5. Sounds, Shapes and Image relations

And what about sounds, shapes and image relations? Having in mind that sounds represent better than vision and like touch, the inertial and rheological properties of the matter, it can be assume that a genuine link between sounds and action may obviously exist. They address directly the same mechanical matter. Such genuine link is more difficult to eliciting in visual events and action, due to the nature more immaterial, and perhaps then more symbolic (as the structural topology identification) of the optical cues.

5. Experiments by means of Virtual Physically-based Objects

5.1. Methodology in sound believability : the basic experiment of the “little coin”

To illustrate the methodology, we describe only an example, made by Claude Cadoz in 1978, and called the little “coin”. The first physical particle model we designed is composed of a small 1D punctual mass falling on a simple ground modeled only by a 1D fix point and a visco-elastic buffer, that is a physically-based representation, of a second order collision. The simulation ran at 20 KHz. When the mass shocks the ground, the last deforms at the acoustical frequencies and we expected that the produced sound was the simple sound of rhythmic bounces. But surprisingly, this sound was composed not only of the expected series of the auditory rhythms. At the end of bounces, the sound exhibits a frequency modulation, impossible to obtain with only temporal sequences of bounces. This frequency modulation is provided by the fact that when the deformation of the ground is at a similar amplitude of the bounces (very small, the two objects are sometimes stuck, constituting a single object with other vibrating mode. This phenomenon is a very discreet feature, which exists in real sounds, and that plays the role of the subtle signature of the interacting objects. Several times, we played this sound surreptitiously when people were in the office, and all of them spontaneously were looking at a coin falling in a glass bowl.

5.2. Experiments in the believability of multisensory-handled virtual objects

5.2.1. Experiments in haptics and sounds

The two following experiments (Figure 4) show a minimal model, which allows to produce in real time believable glass-finger friction sounds and bowed strings sounds on all the range of the manipulation. There is a lot of acoustical models for this type of phenomena. They focus on the complex rosin material that regulates the sticking between the two objects. These models are very complex and they are used, as the physically realistic models of strings, to design real objects. Nevertheless, they have never been able to render all the sensible and complex modulations appearing during performance: timbre modulation, pizzicati, way of attacks, creaking, etc. A simple model as been designed, not focused on the realistic reproduction of the morphological properties of the objects themselves but on the interaction between the surface (resp. the string) and the hand, through a simple non-linear friction model modulated by velocity and pressure. The force feedback device is a 2D stick, 1D for the velocity and 1D for the pressure. It moves only on about 5 cm. The surface (resp. the string) is simply represented by a very few number of uni-dimensional masses linked by simple uni-dimensional visco-elastic constraints.
5.2.2. Experiments in haptics and vision

- **Paradoxical matter**

  The following simulations (Figure 5) illustrate that “an impossible matter” (i.e. a matter having rheological parameters not possible to implement in the real work), is considered as a true matter, when the physico-visual experiences exhibit consistent physical behaviors in time.

  On the upper row, simulations are of a matter that is too hard to really go through the bottleneck (the forces felt by hands are very high) although the deformations can be very large. On the lower row, simulations are of material that is too soft, too much fluid to be felt by fingers. The feeling is very delicate. According to the energetic consistency that is clearly revealed by the visualization as well as by the feeling of the force, this non-realistic object seems unanimously possible and “real”.

![Figure 5. Paradoxical matter](image)

- **Obstacle guiding and avoidance**

  In this experiment (Figure 6), a user is guiding a small physical simulated train, with a lot of DOF, by pulling it from its “nose”, in a labyrinth composed of rigid obstacles and a straight free way.

![Figure 6. Vision as a sense of topology and force as a sense of physical global state.](image)

5.2.3. Experiments in haptics, audition and vision

- **Anamorphosis of the action and force feedback**

  The following experiments show that the morphology of the manipulation and of the visual space can be different according to the presence (or not) of a force feedback. The two simulations represented on the left and on the right of the figure 7 are the same. They are composed of small sharp pyramids moving in a ball manipulated by hand with force feedback. Only the morphology of the manipulation differs.

  On the left, the co-ordinates (x, y) of the sphere are manipulated by two independent keys that are displaced vertically. The motion of manipulation is non-usual and it is very different of the visual motion of the sphere. On the right, the manipulation is by means of a 2D stick and the motion of manipulation is similar to the visual motion of the sphere. Without force feedback, the first manipulation (left) is impossible, as in the game in which we try to draw by manipulating two independent knobs. But, when we added a little drop of consistent force feedback, all the experimenters perform accurate manipulation of the ball and of the pyramids.

![Figure 7. Two morphologies of manipulation: different (left) and homothetic (right) to the visual motion](image)

More surprisingly, in such situation, the manipulation is more accurate than with the second (right) with usual morphology: in the first case, persons are able to control the shocks of the pyramids on the ball producing expected auditory rhythmic sequences.

This means that in the presence of a sufficient energetic consistency between all the multisensori-motor events, the manipulation that allows to feel it more precisely, leads to more accurate manipulation. That is the case when the two coordinates are manipulated separately, the two components Fx and Fy of the force being also felt separately.
The little bouncing grains

The following pictures (Figure 8) show two similar experiments in which little objects are moving inside a ball manipulated by hands with force feedback. The shocks of the grains on the ball produce sounds. On the left, the simulation was made in 1989. The visual quality of the image and the acoustical quality of the sound are low. On the right, the simulation was made ten years after in 1999 with a better rendering of image and sound. These experiments show that the quality of the sound and of the image does not increase the believability of the represented scene. More, the new visual rendering (on the right) underlines the synthetic images, revealing the artificial process and thus, the scene is said technically speaking better than the first but not more believable.

![Figure 8. Believable dynamics vs. realistic rendering](image)

The force feedback multisensory nanomanipulator

The experiment shows in the figure 9 is a manipulation of a complex phenomenon that occurs during the approach-retract of the atomic force microscope probe to a nanoobject. This approach-retract is a non-linear hysteretic phenomenon, with two zones of instabilities. It is difficult to manipulate this type of phenomenon without sensory feedbacks.

![Figure 9. Multisensorial physically - based artifact to improve real tele-manipulation](image)

5.3. Come back to visual representations

As stated at paragraph 4.1, even in pure visual representations, the question of what is the features able to trigger the sense of Presence. As a result of the extensive developments in computer graphics and computer animation of the last 20 years, we understand today that a complete physical reproduction of reality by means of Computers is unreachable and in addition cognitively unsatisfying. Consequently, works in Computer Graphics were thus oriented in non-photorealistic representations. Our assumption being that believability does not reached without motion (even if in its immobility stage), we have to design models in which the consistency of the dynamics appears clearly and optimally (economically).

The following pictures (Figures 10 and 11) show minimal physically - based particle models of complex physical phenomena: granular materials (figure 10, up), turbulent fluids (figure 10, down), pastes (figure 11), which are usually considered as difficult to implement and to calculate. These models have been designed by firstly taking into account only the dynamics features, as dynamic of piling, of avalanches, of collapses, of turbulences, of curling etc. [18]. For this purpose, physical particles models have the strength (that it is obviously considered in shape modeling as a weakness) to implement minimal geometry (points only without neither volume and nor rotations). The animations as well as the static minimal non-photorealistic representations (only by points) illustrate clearly that the motions are convincing in all the range of the phenomena (they are not one shot phenomenon rendering), and that they reveal an underlying plausible object.

![Figure 10. Dynamics vs. geometry (2)](image)

All these models are (1) able to run on real time, and (2) to be controlled consistently by gestures, which increase the believability of what it is represented.

In addition, the more photorealistic rendering shown on the right cell of the figure 11 illustrates that the plausibility is not drastically increased by the visual realism: the image is obviously better but the plausibility of the underlying object (i.e. the presence of it) is of the same level.
6. Conclusion

This paper presents an approach of the Presence concept, based on the assumption that the dynamics, representing the matter’s behaviors, that produces multisensory events under manipulation, plays the role of the invariant identified during action-perception experiences of the objects. This leads to put the emphasis more on specific consistent correlation between the multisensory-handled events than (1) in each sensory feedback or (2) in arbitrary purely phenomenological correlation between them. It points out the role of new computerized electromechanical technology as those implemented in Virtual Reality, that allows to represent real physical objects as an input – processing – output system in which the computational process can play the role of the representation of the physical object. The methodology applied is this “understanding by building”, allowing to create physically plausible artifacts representing a priori hypothesis on the relevant features able to trigger the sense of presence. Physically-based models of multisensory-handled events have been made to explore (1) the genuine link between gesture and audition through dynamics and mechanical matter (2) the believability of impossible material, (3) the robustness of the manipulation to the morphological changes in presence of force feedback and energetic consistency, (4) the bias introduced by synthetic images representations in a coherent multisensory (vision-audition-touch and action) situation and (5) the possibility to build sufficiently believable multisensory-handled artifact to drive complex teleoperated tasks.

In conclusion, the core minimum criterion to trigger the feeling of presence of virtual or represented worlds is probably the existence of matter, responsible for the energetically consistent dynamic correlation between all the multisensory-motor events produced by the object. Thus, the primary property to be instilled in our virtual representations, whatever they are, should be a drop of “evoked matter” using dynamic models designed to represent the just sufficient relevant physical consistency in the produced multisensory sensible phenomena.

7. Acknowledgments

The physically-based models and simulations presented in this paper have been designed with my colleagues Claude Cadoz and Jean-Loup Florens, and with an invited artist David Pryrtech, a post-doctoral researcher Daniela Urma and the PhD students Claire Guilbaud and Sylvain Marlière. This work has been supported by the French Ministry of Culture and by the FP6 Network of Excellence IST-2002-002114 - Enactive Interfaces.

8. References

Abstract

This paper describes the application and results of two studies using a qualitative tool designed to examine sense of place in relation to presence research in the BENOGO project. The ‘place probe’ incorporates a range of techniques aimed at articulating a person’s sense of place, these include: verbal descriptions; sketch maps; salient features; semantic differentials and six words that best describe the experience of being in a place. The paper will describe the application of the probe in both a physical place and a virtual representation of that place. The Analysis of the results reveals a similarity of reported experience in both places, but that the extremes experienced in the physical place were more flattened in the virtual representation.

Keywords--- Presence, Place, Probes, Subjective measures, Empirical studies of place.

1. Introduction

BENOGO is a research project funded under the European Community’s Future and Emerging Technologies “Presence” initiative that aims to investigate concepts of presence within the technologies of real-time image-based rendering. One aim of the research is to develop new tools for empirical and theoretical studies of presence based on the concept of the observer’s embodiment in the computationally created virtual environment. To do this we are developing new tools for empirical and theoretical studies of presence based on the concept of embodied cognition, or embodiment [4]. As real places (possibly known to the observer) with man-made and/or organic objects (like trees, foliage etc.) are hard to represent in a graphically constructed virtual environment, another objective is to bring about new insight into presence through comparison with the sense of presence experienced in the real world. The new technology of Image Based Rendering does not require a reconstructed geometrical model of the scene [3, 16]. It bypasses an important technological problem and presents a breakthrough, but large amounts of image data need to be stored, and recalled for real-time visualization. Thus a third objective of the project is to find ways in which the key aspects of the place can be communicated to engineers, so that technological constraints do not undermine the subjective feeling of place.

It is contended that the philosophical background of embodiment has two important consequences for the research. Firstly, the sense of presence requires a body; it is not solely cognitive. A body is clearly missing from the experience of many virtual environments (such as those rendered using a head mounted display). Secondly, the requirement to articulate what people are thinking and feeling in order to compare experiences across the real and the virtual. This raises the issue of how to probe people’s thoughts about a situation without interfering with the concepts and processes that underpin the experience.

For these reasons and for the more pragmatic reasons of informing the design of our photo-realistic virtual environments, the decision has been made to distance the research from attempting to understand the whole of the complex concept of presence. Instead the work has focused on a key-contributing factor of presence, namely place. The thrust of the BENOGO project is ‘to be there without going’; that is to provide people with a realistic sense of being somewhere else. Accordingly the research has focused attention on capturing the essential features of places and finding ways in which to communicate these to the designers and engineers of virtual environments.

This paper reports on the experience of conducting empirical work into understanding and representing a sense of place. A qualitative tool entitled the place probe is introduced and its application in both a real and virtual environment is discussed. The paper presents results from both environments and draws conclusions as to the utility of the probe as a means of articulating discussions concerning the sense of place.

2. Contributing studies

The approach adopted by the BENOGO team from the outset of the research has been to understand place in relation to presence, from a qualitative perspective. As a means of grounding this approach, a range of previously used techniques were explored. A number of these methods were utilized during the initial phase of the BENOGO project.
The Immersive Tendencies Questionnaire (ITQ) was developed to identify real world tendencies (e.g. using computer games) that may affect a person’s sense of presence [23]. The ITC-SOPI was developed for the UK’s Independent Television Commission. It is a cross media questionnaire, which explores: spatial presence, levels of engagement, sense of naturalness and negative aspects that effect presence [10].

We have also used more open-ended methods such as talk-aloud methods [13], semi-structured interviews and written descriptions of places [18].

Sketch maps [21] are a technique used to elicit people’s understanding of spaces, key landmarks and relationships between salient features of an environment.

Finally we have explored the use of repertory grids as a means of gaining access to the meaning a person has attached to their experience and/or properties of the environment. In the repertory grid technique [9] people are asked to describe their experiences according to a number of dimensions supplied either by the people themselves or by the experimenter. Relationships between these dimensions are grouped together and used to arrive at key concepts that describe a domain.

The effectiveness of these techniques has been explored through a variety of data analysis techniques such as grounded analysis, peer reviewing, semiotic analyses and various forms of coding. Furthermore, the utility of these techniques has been studied in a variety of settings such as environmental architecture [17], real and virtual environment representations of botanical gardens a university stairwell and a city view of Prague, all rendered in an HMD [11,12,13].

One of the motivations underpinning the research is to compare virtual representations against real places. Such comparisons can be made at many different levels of abstraction. The comparison of the Prague botanical gardens and the Edinburgh botanical gardens [13] was considered to be valid at the level of general characteristics – the virtual gardens did not feel hot and humid which was a key characteristic of the real gardens. However, many of the plants were of similar shapes and sizes. They were not the same plants, but they provided an overall feel for ‘being in’ a botanical garden.

The research reported in this paper has provided the best opportunity to ‘benchmark’ a place and to enable the comparison between a real place and a virtual representation of that place. The place chosen was a city viewpoint in Prague. Of course it is very difficult for the real and the virtual places to ever be exactly the same. In this case, they were experienced at different times of day and with different weather conditions. This impossibility of making a detailed and exact comparison must be born in mind when considering the ‘benchmarking’ data. However, for the purposes of informing the design of a virtual place that is as faithful as possible to the real place, focusing on the key salient aspects does appear useful.

3. The place probe

The experience of using a variety of data capture techniques in our previous studies indicated that no single questionnaire, or set of questions or other unitary method was going to provide the rich variety of data required to understand the key features of a place. To address this lacking it was proposed to utilize a variety of data capture methods used in conjunction with a range of data analysis methods. Accordingly a ‘Place Probe’ was created.

Probes have been used recently in two main contexts. Cultural probes [6] for the capture of rich data about the context in which a technology was being used, or where it was likely to be used. The probes were not intended to simply elicit some objective data they were intended to provoke responses. In a similar vein Technology probes [22] have been used particularly in the domestic setting to explore new uses of technologies. These probes used methods that informed the design of a prototype system for a common interface where different generations within a family could communicate irrespective of physical location. Information was captured about the nature of communications between family members. The probes contained a communication diary, notebook, two disposable cameras, address envelopes and a pen. Probes have also been used in other design and evaluation situations [5] and [2] to explore domestic environments.

The probe reported here differed in that it was designed to capture experiences at a specific point in time or a particular place rather than over an extended period of time. Drawing on the experiences of the previous empirical studies it was decided to include the following instruments within the probe.

3.1 Probe part 1: the visitors book

Research undertaken by Turner and Turner [20] has highlighted the written reports contained in visitor’s books as a source of rich data about place. Indeed such reports have the advantage that they are often ask open-ended questions e.g. ‘Please tell us about your experience’ rather than ‘Tell us about the lighting’, hence they do not prompt people to provide answers on specific topics.

3.2 Probe part 2: sketch maps

Sketch maps are a technique used to elicit people’s understanding of spaces, key landmarks and relationships between salient features of an environment. In this case accuracy of the map is not of prime concern, rather it is the depiction of those aspects of the place that people remember for example a tree, building or seating area. They can also be used to provide additional information such as where people are standing or their paths through the environment.
3.3 **Probe part 3: salient features**

This section of the probe asks for participants to rate the three most salient features of the environment. The aim of this is to establish the most important visual benchmarks in the real world that should be identifiable in the virtual one.

3.4 **Probe part 4: semantic differentials**

This part of the probe combined Osgood’s semantic differentials [14] and Relph’s three aspects of place (physical features, activities afforded and affect engendered). The objective is to gauge people’s feelings towards the environment based on their responses to the differential scale. For example people are asked if the environment is attractive or ugly, with the rating ranging from Very Attractive, through neither (i.e. not attractive or ugly) to Very Ugly. Responses to this section result in a gauge of how strongly or weekly participants rate their experiences of the environment within the scales provided.

3.5 **Probe part 5: six words**

The final part of the probe asked people to write down six words which best described their experience of being in a particular place. These were used to establish themes that emerged across participants based on the relationships between the words and their synonyms.

4. **Data analysis methods**

A grounded theory approach to data analysis was adopted, the objective being to uncover common aspects of the scene from the varying sources of data. Having uncovered the common themes these were categorized in terms of models of place derived from environmental psychology offered by Relph [15], Jorgenson [8] and Gustafson’s [7], particularly Relph and Gustafson. This allowed us to study the utility of these models in the context of the current research [18,19].

4.1 **Inter-rater reliability**

The range of comments found within the descriptive paragraph resulted in it being open to a high degree of interpretation by each evaluator. To combat this, random samples of comments were analyzed by all three evaluators, during which time common themes were found to be emerging.

One evaluator analyzed the sketch maps for salient aspects such as named locations/objects, or easily identifiable locations. There was obviously some degree of interpretation within sketch maps, as a result a random sample of ten were given to two other evaluators to check for inter-rater reliability within the data. Maps contained only basic drawings of the cityscape, therefore rather than attempting to categorize each building this was placed within the ‘general city or cityscape’ category.

For the semantic differentials it was decided not to calculate medians or means as they provide only a very crude indicator. Rather it was decided to count the number of responses in relation to each point within the differential scale. This data was used to corroborate data found within other data capture methods. However the main objective was to capture data that could be used to benchmark the real and virtual scenes.

The place probe has been used in three real world environments and two virtual environments. There is no doubt that it does indeed capture important characteristics of places. In the following section we report on a comparison of a real city viewpoint and on a photo-realistic image rendered through a head mounted display with full head-tracking and simulated stereo vision.

5. **Study 1: the viewpoint, Prague**

The first use of the probe was in Prague during December 2003. The viewpoint is one of Prague’s largest single green spaces (with many trees) and it is ideal for quiet walks and breathtaking views over the city of Prague (Figure 1).

![Figure 1. View of Prague from the Hillside Viewpoint](image-url)
At the hilltop view, members of the public who were visiting the area were asked to complete the presence probe. Each person was told it would take around 15 minutes and they would be paid 150CZK (around 5 euros). Participants were informed that the study was interested in their experience of the place.

5.1 Descriptions

Recall that in this section, participants were asked to write a free-form paragraph describing their experience. They were not prompted to focus on anything in particular. Below is a typical example of the data gathered from the description section of the probe:

“A gray mist is covering Prague, my fingers are bitterly cold but still the magnificence of Prague cannot be covered. I was standing on the hill, which the castle is on, and it was facing the new town, the Volta River and the old town in the distance. All the most beautiful landmarks are visible and I felt totally happy and contented to be there”. (22,female, Singaporean, first time at the viewpoint)

It is interesting to note that the three levels of description from Relph’s model of place [15] occur quite naturally in such descriptions. There are comments on the physical aspects, the activities offered and affective characteristics of the environment. Other descriptions typically followed this pattern.

5.2 Sketch maps

The sketch maps drawn by the participants in the Prague study contained a great deal of detail. The buildings, landmarks and other physical aspects of the environment were very descriptively drawn with the shapes and types of buildings made evident. Particularly most of the maps highlighted the cathedral, cityscape, statue, walkways/paths, monastery, hillside, platform and trees, along with labels identifying their names. Figure 2 is an example of the kind of sketch maps that were drawn by the participants in the Prague study.

![Figure 2: A typical sketch map from the Prague Study](image)

5.3 Salient features

In this section of the probe, the participants seemed to rank the view, the castle and their feelings/impressions of the place (i.e. quiet, peaceful) as the features of most importance. Overall, the castle featured more dominantly than any other feature (12 times out of a total of 30). The view was mentioned on ten occasions with references to it ranging from a ‘fantastic’ view to a ‘wide’ view to a good ‘view’. The platform, river, nature and trees were also frequently mentioned. While the hill, Charles Bridge, the statue, St. Nicolas church, buildings and monastery were less frequently mentioned. Feelings and expressions of openness and niceness were noted throughout the responses and the overall impression was quiet, peaceful and clean with noted sounds of dogs and birds.

5.4 Semantic differentials

Below is a table indicating the participants’ responses to the semantic differentials. The numbers are given as a direct count from the probes. This section of the probe allows participants to rate their experiences of the environment in relation to scales that are based on data from our previous rep-grid studies and from Relph’s model of place:

<table>
<thead>
<tr>
<th></th>
<th>Very</th>
<th>Quite</th>
<th>Neither</th>
<th>Quite</th>
<th>Very</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attractive</td>
<td>23</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Big</td>
<td>7</td>
<td>12</td>
<td>6</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Colourful</td>
<td>5</td>
<td>12</td>
<td>8</td>
<td>4</td>
<td>Colourless</td>
</tr>
<tr>
<td>Noisy</td>
<td>4</td>
<td>8</td>
<td>5</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>Temporary</td>
<td>1</td>
<td>7</td>
<td>6</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>Available</td>
<td>4</td>
<td>11</td>
<td>10</td>
<td>4</td>
<td>Permanent</td>
</tr>
<tr>
<td>Versatile</td>
<td>2</td>
<td>11</td>
<td>8</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>Interactive</td>
<td>5</td>
<td>8</td>
<td>5</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>Pleasant</td>
<td>23</td>
<td>5</td>
<td>2</td>
<td>Unpleasant</td>
<td></td>
</tr>
<tr>
<td>Interesting</td>
<td>19</td>
<td>6</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Stressful</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>20</td>
</tr>
</tbody>
</table>

5.5 Words

The words were categorized in terms of semantic similarity, where a grounded approach was used to identify emerging themes that appeared again and again. Interestingly, the strongest themes bear some similarity to those used in the semantic differentials while others are more specific to the environment. Below are examples of responses together with the themes under which they were categorized.

Peaceful/Relaxing: Quiet, relaxing, peaceful, calm, tranquil, silent.

Interesting: Interesting, educational, enthusiastic, curious, engaging, absorbing, captivating.

Beautiful: Beautiful, nice, pleasant.

Cityscape: View, cityscape, scenic, outlook, sky, open, endless remote, distant, roofs, castle, town, people, noise, monuments, atmosphere.
Enjoyment: enjoyment, contentment, happy, satisfaction, worthy, fulfilled.
Amazement: surprised, overwhelming, amazing, powerful, impressive, mighty.
Cold: cold, damp, autumn, gray, cool.
Nature: nature, birds, life, environment, trees, river.
Refreshing: fresh revitalizing, bright, freedom, recovery, escape, vibrancy, clean, fresh air.

5.6 Prague viewpoint results

Taking all of this evidence together, what emerges from the data in the Prague study are a set of themes that can be grouped in accordance with both Relph's [15] and Gustavson’s [7] models of place. At the most general level these are the physical attributes, activities and the emotional or affective aspects of the place.

The physical attributes of the environment included: The Statue, The Benches, Paths and stairs, Platform, Plaque, Background buildings, Trees, Birds, Hillside, Cold, Local Sounds, Distant Sounds, Cityscape.

The activities that the participants engage in while they were in this environment are generally: Looking at the cityscape, exploring, and moving about.

The affect that the environment had on the participants i.e. the types of feelings, emotions, memories and other experiences that the participants offered in relation to their experience of the immediate environment were: Enjoyment/Contentment, Refreshing/Re-vitalizing, Interesting/Engaging, Peaceful/Relaxing, Amazement, Beautiful.

6. The BENOGO virtual viewpoint

The BENOGO VR (Figure 3) environment was a representation of the viewpoint in Prague rendered using BENOGO’s real-time IBR (image based rendering) software. The system in itself comprises six networked computers running as a cluster, a head mounted display (HMD) and eight speakers providing surround sound. The system is set up in a darkened room where all of the participants take part in the studies.

Figure 3: Viewpoint Mosaic created for the Head Mounted Display

BENOGO IBR offers the unique advantage of real-time rendered images for a moving observer in a virtual environment. This requires a lot of computing power. While the cluster did provide this, it came with restrictions caused by the bandwidth of the network, that at the time unfortunately reduced the performance of the IBR technology slightly (more recent demo’s of BENOGO IBR have resolved this issue).

While the real-time IBR technology offered very realistic parallax motion, the environment had an overall feel that lacked detail and had a tendency to jitter, this was partially due to the restrictions previously mentioned and partially because of the way the images are rendered and ‘stitched’ together in real-time. The images themselves were of high quality and depicted the viewpoint scene in Prague on a sunny spring day with a clear blue sky and no leaves on the trees. The buildings close by, the statue, the hedges and the platform itself were all clearly identifiable. However the resultant lower resolution of the real-time IBR meant that the images in some way looked less detailed, particularly the distant buildings of the cityscape.

Those who took part in the study in virtual environment were instructed on how to use the HMD and then told to explore and experience the environment. They were made aware of the movement restrictions of the HMD but encouraged to move freely within the available space. They were also told that they would be immersed in the virtual environment for a period of time and that they would be informed when that time was up. When the time was up they were helped out of the HMD and taken through to a separate room where they undertook the probe questionnaire. After this they were offered some light refreshment as reward for taking part in the study.

6.1 Descriptions

Below is a typical example of the data gathered from the descriptive section of the probe. In general the responses to this section identified aspects of the virtual environment as well as distortions, glitches and resolution problems with the IBR rendering. Interestingly, there was a distinct lack of emotional responses to this section of the probe compared to the real viewpoint:

“The view was from a hill overlooking a city in southern Europe on a sunny day with birds and some church bells in the distance. I was standing on a paved circle with a statue behind me. The picture was a bit blurry especially the trees. Perspective seemed natural. There were camera/lens reflections hanging in mid air behind me, though it was not possible to see the sun only a white sky”. (24, male, Danish, First time user of HMD technology).

6.2 Sketch maps

The sketches drawn by the participants in the VR study were markedly simpler than those in the real environment. While key features were still identified and
arranged in a similar way, in most of them, a large amount of detail remained lacking (Figure 4).

Figure 4: Sketch map of the Virtual Viewpoint

6.3 Salient features

In the virtual studies, the participants seemed to rank the church, castle, city and statue as the most important features. And overall, the statue/sculpture was most frequently noted (15 times), then the church (11 times) and the castle (8 times)- bearing in mind there has been some discrepancy over what people where calling each of these buildings (i.e. some called the church a castle and visa versa). The view was mentioned on six occasions as was the large white house /monastery. To a lesser extent birds, platform, city, trees, bench, camera stand, roofs and sunshine were noted. Some participants made a note that it was difficult to see the details in the environment and there was a mention on two occasions about the feeling of quietness.

6.4 Semantic differentials

Below is a table indicating the participant’s responses the semantic differentials. Again, this section of the probe allows participants to rate their experiences of the environment in relation to the scales provided. The numbers are given as a direct count from the probes.

<table>
<thead>
<tr>
<th></th>
<th>Very</th>
<th>Quite</th>
<th>Neither</th>
<th>Quite</th>
<th>Very</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attractive</td>
<td>7</td>
<td>15</td>
<td>7</td>
<td></td>
<td>7</td>
</tr>
<tr>
<td>Big</td>
<td>2</td>
<td>11</td>
<td>13</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Colourful</td>
<td>14</td>
<td>5</td>
<td>9</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>Noisy</td>
<td>3</td>
<td>7</td>
<td>9</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>Temporary</td>
<td>3</td>
<td>6</td>
<td>8</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>Available</td>
<td>1</td>
<td>9</td>
<td>11</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Versatile</td>
<td>2</td>
<td>6</td>
<td>12</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>Interactive</td>
<td>1</td>
<td>7</td>
<td>6</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>Pleasant</td>
<td>4</td>
<td>16</td>
<td>5</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Interesting</td>
<td>8</td>
<td>11</td>
<td>5</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Stressful</td>
<td>5</td>
<td>5</td>
<td>11</td>
<td>8</td>
<td>5</td>
</tr>
</tbody>
</table>

6.5 Words

In a similar manner to the Prague study, a number of themes emerged from the words used by participants to articulate their experience of the virtual viewpoint. Below are examples of responses together with the themes under which they were categorized.

- **Peaceful/relaxing**: peaceful, serene, quiet, relaxing, harmony, calm, tranquil.
- **Grainy**: blurry, hard grained, coarse grained, fragmented, grainy, unclear, low resolution.
- **Weather**: summer, warm, sunny, cold, windy.
- **Mediterranean**: Greek, Spanish, vacation, holiday, south, family.
- **Beautiful**: beautiful, nice, pleasant
- **Restrictive**: want to move, inability to join in, non-explorative, restricted, captured, limited, stuck, static, locked up.
- **Sounds**: birds, crows, ambulance, sound fall, church bells, church time.
- **Interesting**: interesting, curious, exciting, fun.
- **Realistic**: realistic, real, semi real.
- **Faded**: faded colours, lack of colour, not colourful.
- **Lonely**: lonely, loneliness, solitude alone.
- **Stressful**: stressful, annoying, confusing, frustrated.
- **Natural**: natural, blue sky, trees, outdoor.
- **Viewpoint**: viewpoint, view, skyline.

While some of the categories are the same as the real environment the most notable difference is the emergence of themes that refer directly to technical aspects of the VR experience.

6.6 Results from the VR viewpoint

Considering all of the sections of the probe together, themes emerge from the data in relation to Reph’s and Gustafson’s models of place mentioned earlier.

In a similar way to the Prague viewpoint study the predominantly physical attributes of the environment include: The Statue, The Benches, Paths and stairs, Platform, Plaque, Background buildings, Trees, Birds, Hillside, Grainy/blurry, Realism, Sounds, Cityscape, People, Weather.

The activities that participants engaged in while immersed in the VR environment were, to a certain degree, either static or restricted in some way by the technology itself or by the quality of the images: Standing, restricted movement, restricted looking, blurry images.

The affect that the environment had on the participants offered a great deal of interesting themes in relation to both the emotional affects and other experiences evoked by exposure to the VR environment. In particular references to Mediterranean countries and vacations were quite different from those in the real Prague study: Interesting/Exciting/Fun, Nice/Pleasant/Beautiful, Peaceful/Relaxing, Stressful, Restricted Movement, Mediterranean, Holiday, Dizziness, Loneliness.
7. Comparing the results

The Themes that emerged across both environments were predominantly those ones related to the physical/visual make up of the environment, with some of the affective ones also translating well. While the quality of images in the VR environment was clearly an issue participants still identified and located the buildings in and around the view. However due to the lack of resolution and also the fact that they had not actually been visiting Prague they were unable to name the specific buildings and tended to generalize about the nature of the city rather than ‘pick out’ specific locations.

Themes such as Peaceful, Interesting, Beautiful and Nature also appeared in both studies, as did Cold and Viewpoint to some extent. The themes that appear in the real world but not in the virtual are: Enjoyment, Refreshing and Amazing. This shows the lack of a positive emotional affect on the users of the virtual world that was a particular aspect of the real viewpoint.

Themes that appear in the virtual world but were not part of the real world were Grainy, Restrictive, Faded, Realism, Stressful, Lonely and Mediterranean/holiday. These themes often point to the technical aspects of the VR environment highlighting problems that interfere with a sense of place. More significantly however they highlight some emotional responses to the VR viewpoint that were very different to the experiences of people in the real environment. These responses may be directly linked to the technical aspects of the environment such as restricted movement and low resolution that combined to frustrate participants. Also, responses that resulted in the emergence of the Mediterranean/holiday theme show the impact of participants not knowing the specific location of the place they were visiting. This highlights the connotative aspects of the IBR rendered environment.

Looking at the two semantic differential tables specifically, what is immediately noticeable is the shift from the extremes in the real environment towards the middle for the virtual environment. This suggested that the real environment had a much more powerful affect on participants than the virtual one. Breaking it down into the three main connotative sections it becomes obvious that the real environment was considered to be more attractive, felt bigger and was more colourful than the virtual one. There is no clear response to the noise differential or the perceived permanence of the place in either environment, although there was a slight trend towards the more permanent in the real world.

With regards to the participants responses to the activities section of the differentials, there was a shift, if however slight, away from the extremes. In the real environment the trend was towards experiencing the environment as quite available, quite versatile and quite interactive. In the Virtual environment there was a move towards neither available nor unavailable, neither versatile nor limited. There was also an opposite trend towards passive rather than interactive, which was a very different response to the real world.

In the last section, as in the first there were more positive responses. All three factors show similar positive trends towards experiencing the environment as pleasant, interesting and relaxing. However there was a difference between them in that the responses to the real world were much more positive than in the virtual one.

What the data from the two studies appears to show is that the participant’s experience of the virtual representation are similar to those of the participant’s in the real environment but that the responses are less positive. That is to say that the virtual world does not have the same intensity of affect on the participants as the real environment does.

Although participants clearly responded to parts of the probe in similar ways in each study, there are clear differences between the two. While the virtual environment was able to recreate a sense of peacefulness, beauty and interest similar to the real place, it was unable to recreate the refreshing sense of enjoyment and amazement that the real place engendered in the participants. Indeed it would appear that this type of experience in the real world has been replaced in the virtual one by a sense of stressful loneliness brought on by the lack of realism or intensity in the faded, grainy images and restrictive nature of the virtual environment.

8. Conclusions

This paper has discussed the application of ‘The Place Probe’ during the evaluation of a real and virtual environment. The aim of the probe was to capture data about the real world that would be of use to the designers of virtual environments attempting to recreate real places. An explicit intention of the probe was to find out what was missing from any experience of a virtual environment when compared to its real world counterpart, rather than simply provide a quantitative score for place or presence. It is contended that such an approach when combined with traditional methods such as ITC-SOPI and ITQ will provide a greater insight into level of presence experienced by people, and how this is affected by their sense of place. The method is qualitative in nature and that of course introduces a series of issues with data interpretation, capture and reporting. However it is believe that by using appropriate methods of inter-rater reliability and that the multiple sources of data within the probe overcome some of the issues with qualitative data.

Analysis of the data generated by the probe suggested that people were having broadly similar experiences in both the real and virtual environments of the same location. Furthermore the data highlighted both technical and non-technical issues that affect the desired objective of re-creating a real place in virtual reality. In that sense the subsequent analysis was able to report to
the technologists/designers of the virtual place about what was lacking in the IBR environment (e.g. sufficient resolution through to other issues such as lack of other people, and how this created a sense of loneliness). Furthermore a series of attributes have been identified which should be addressed in future designs. Moreover, the types of issues found in many cases can be expanded upon and it is our intention to use these themes as templates that can be provided to the developers. Templates are envisioned as early patterns, in a manner similar to Alexander [1] that can provide information on features of a space and will provide information on the technical aspects that may alter their affect and implementation.

In conclusion, the ‘The Place Probe’ provided a method of exploring real and virtual places and demonstrated the potential to make a valuable and timely contribution to the design of virtual places. The technique has been successfully used to compare the similarities and differences between a virtual environment and its real world counterpart, as well as to identify areas of both strength and weakness within a virtual space. From a technical standpoint it is a technique that can highlight areas demanding technical solutions, for example the requirement to increase the resolution of some objects, as they are deemed more important to that scene. While the probe continues to be refined the initial indications are positive in terms of its potential for contribution to the design, implementation and evaluation of photo realistic virtual environments.

9. Acknowledgements

The authors would like to thank all members of the BENOGO project for their input and co-operation with the studies mentioned in this paper. http://www.benogo.dk

10. References

Spatial Presence and Emotional Responses to Success in a Video Game: A Psychophysiological Study

Niklas Ravaja1, Jari Laarni1, Kari Kallinen1, Timo Saari1, Mikko Salminen1, Jussi Holopainen2, and Aki Järvinen3
1M.I.N.D. Lab/CKIR, Helsinki School of Economics, Finland, 2Nokia Research Center, Finland, 3Veikkaus, The Finnish National Lottery, Finland
{ravaja@hkkk.fi, laarni@hkkk.fi, kallinen@hkkk.fi, saari@hkkk.fi, misalmin@hkkk.fi, jussi.holopainen@nokia.com, aki.jarvinen@veikkaus.fi}

Abstract
We examined the relationship of self-reported Spatial Presence with emotion-related psychophysiological responses to success (reaching the goal) in a video game among 36 young adults. Event-related changes in facial electromyographic (EMG) activity (an index of emotional valence), electrodermal activity (EDA; an index of arousal), and cardiac interbeat intervals (IBIs) were recorded. The results showed that Spatial Presence was associated with an increase in zygomatic and orbicularis oculi, and a decrease in corrugator supercilii, EMG activity in response to reaching the goal. Spatial Presence was also positively related to EDA, but unrelated to IBI. High Spatial Presence may result in a greater increase in positive affect, and a greater reduction in negative affect, in response to a positive event. Spatial Presence exerts an influence on both the valence and arousal dimensions of emotions. Phasic corrugator supercilii EMG activity may be a particularly useful measure in Presence research.

Keywords--- Spatial Presence, emotions, facial electromyography, electrodermal activity, interbeat interval, video games.

1. Introduction
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]. It has been suggested that Spatial Presence may have important emotional consequences. That is, media presentations, such as video games, engendering a strong sense of Spatial Presence have been suggested as eliciting higher arousal and enjoyment [1, 3, 4]. People also differ in their ability to experience Presence [5], and these differences may be associated with the emotional responses to the mediated environment. Examining the relationship of Presence with emotional responses is of import both in its own right and because Presence may exert an effect on other important variables and processes (e.g., attention, memory, entertainment, and persuasion) through the mediating influence on emotions [see 1, 6].

Emotions are biologically based action dispositions that have an important role in the determination of behavior [e.g., 7]. Most theorists endorse the view that emotions are constituted by three aspects or components: subjective feeling, expressive behavior, and physiological arousal; others add motivational state or action tendency and/or cognitive processing [7, 8]. A dimensional theory of emotion holds that all emotions can be located in a two-dimensional space, as coordinates of valence and arousal (or bodily activation) [e.g., 7, 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.

The influence of Spatial Presence on arousal is relatively well established by studies using psychophysiological measures [10, 11]. Electrodermal activity (EDA), commonly known as skin conductance, is the primary psychophysiological index of arousal [6]. As people experience arousal their sympathetic nervous system (SNS) is activated, resulting in increased sweat gland activity and skin conductance. EDA has recently been used in a couple of studies examining Presence in virtual environments (VEs). These studies showed that EDA was positively associated with self-reported Presence during exposure to VEs depicting an airplane flight [10] and a pit room with an unguarded hole in the floor leading to a room 20 ft. below [11]. Meehan et al. also found that EDA was higher during exposure to the frightening (i.e., arousing) virtual height situation compared to a non-frightening virtual room [11]. Although these studies showed that EDA is positively related to Presence when the media content is
arousing, there is no reason to expect that EDA would increase with increasing Presence when the content of the mediated environment is non-arousing (e.g., a deserted beach of a Caribbean island).

Some studies have associated Presence also with heart rate (HR) changes. Meehan et al. showed that the aforementioned stressful VE depicting a pit room evoked notable HR acceleration (i.e., a decrease in interbeat intervals, IBIs) [11]. In addition, changes in HR correlated positively with self-reported Presence. Apparently, the VE (pit room) was stressful enough to elicit arousal-related HR acceleration that is mediated by the SNS [see 6]. In contrast, Wiederhold et al. found that changes in HR correlated negatively with self-reported Presence when participants were presented with the aforementioned VE depicting an airplane flight [10]. This VE (airplane flight) is likely to have prompted increased attention resulting in HR deceleration that is mediated by the parasympathetic nervous system (PNS) [see 6]. Thus, there are obvious interpretative difficulties associated with HR responses: HR may be either positively or negatively associated with Presence depending on the content of the mediated environment and task demands (e.g., attentional demands).

The relationship of Presence with the valence dimension of emotions is not well established, and psychophysiological studies on this issue are lacking. Although some studies using self-report measures have suggested that Presence may result in greater overall enjoyment [3, 4], the relationship of Spatial Presence with positive and negative emotional responses to specific events in the mediated environment has not been investigated. According to Frijda, events appraised as real elicit emotions, and their intensity corresponds to the degree to which the events appear real [12]. That being so, high Spatial Presence might result in more positive emotional responses to a positive event, for example.

In the present study, we used facial electromyography (EMG) to examine the relationship of Spatial Presence with positive and negative affective responses to success (reaching the goal) in a video game (Super Monkey Ball 2). The facial EMG provides a direct measure of the electrical activity associated with the facial muscle contractions underlying emotional expression, and is the primary psychophysiological index of hedonic valence [6, 13]. 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 media [14, 15, 16, for a review, see 6]. In addition, tonic activity at the orbicularis oculi (periocular) muscle area has been associated with positively valenced high-arousal emotions [6].

The preceding discussion provided the foundation for the following hypotheses. Reaching the goal (i.e., a positive event) in a video game would be expected to elicit an increase in zygomatic (Hypothesis 1) and orbicularis oculi (Hypothesis 2) EMG activity for participants reporting high Spatial Presence during the game, but not for participants reporting low Spatial Presence. In contrast, reaching the goal would be expected to prompt a decrease in corrugator EMG activity for participants reporting high Spatial Presence, but not for participants reporting low Spatial Presence (Hypothesis 3). In regard to emotional arousal, there is likely to be high anticipatory arousal before attaining the goal, after which arousal diminishes. However, it would be expected that this reduction in arousal, as indexed by EDA, would be more pronounced for participants reporting low Spatial Presence compared to those reporting high Spatial Presence (Hypothesis 4). We also hypothesized that Spatial Presence would not have an influence on IBI response to reaching the goal (Hypothesis 5). This is because (a) the video game used is likely to elicit both emotional arousal (cardiac sympathetic activity increases) and attentional engagement (cardiac parasympathetic activity increases) and (b) the SNS and PNS have opposing effects on cardiac IBI [6].

2. Methods

2.1. Participants

Participants were 36 (25 male and 11 female) Finnish undergraduates with varying majors (1 participant was about to apply to the University), who ranged from 20 to 30 years of age. All participants played video or computer games at least once a month. They participated in return for three movie tickets.

2.2. Video game

In the present study, we used a video game called Super Monkey Ball 2 (Sega Corporation, Tokyo, Japan). The game was played with the Nintendo GameCube (Nintendo Co., Ltd., Kyoto, Japan) and presented on a screen using the Panasonic PT-LC75E Multimedia Projector (Matsushita Electric Industrial Co., Ltd., Osaka, Japan). The image size was 114 cm (width) × 85 cm (height), and the distance between the player’s eyes and the screen was about 200 cm.

The game 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. 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 to the deep. The player needs to avoid obstacles and pick objects as the monkey rolls around the stages. The aim was to clear each stage with as high a score as possible. The player had 1 min to clear each stage and earned extra points for clearing the stage in 30 s or less. The practice session was played at the Beginner level, and the actual two play sessions, easy and difficult, were played at the Beginner and Advanced levels, respectively. Super Monkey Ball 2 is a relatively nonviolent game with happy background music and atmosphere. It requires fast reflexes and some strategy.

2.3. Procedure

When arriving to the laboratory, the participant returned a number of questionnaires (e.g., gaming habits,
personality, temperament) that had been sent to him or her beforehand. After a brief description of the experiment (the participant was told that the researchers were interested in measuring his or her psychological processes and physiological activity during playing video games), the participant filled out an informed consent form. Electrodes were then attached and the participant was seated in a comfortable armchair, followed by a rest period of 7 min. The participants played four different video games in a random order. There were three 5-min game sessions for each of the four games; that is, a practice session and two actual play session (i.e., easy and difficult). In the present study, we used only data from Super Monkey Ball 2 played at the Beginner level. The participant was told that the three best male and female gamers would be awarded one movie ticket as a bonus. After each of the actual game sessions, the participant rated the sense of Presence experienced during the game and his or her emotional responses to the game on several dimensions (the items were presented on a computer screen). The room was dimly illuminated during the game and his or her emotional responses to the game environment.

**2.4. 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 [17]. 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 items addressing the Spatial Presence factor. 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. Physiological data collection**

Electrocardiogram (ECG) was recorded 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.

Facial EMG activity was recorded from the left corrugator supercili, zygomaticus major, and orbicularis oculi muscle regions as recommended by Fridlund and Cacioppo [18], using surface Ag/AgCl electrodes with a contact area of 4 mm diameter (Med Assoc. Inc., St. Albans, VT). Electrodes were filled with TD-246 electrode gel (Med Assoc. Inc.). The raw EMG signal was amplified, and frequencies below 30 Hz and above 400 Hz were filtered out, using the Psylab Model EEG8 amplifier. The raw signal was rectified and integrated using the Psylab INT8 contour following integrator (time constant = 50 ms). EDA (skin conductance level, SCL) was recorded with the Psylab Model SC5 24 bit digital skin conductance amplifier that applied a constant 0.5 V across Ag/AgCl electrodes with a contact area of 8 mm diameter (Med Assoc. Inc.). Electrodes were filled with TD-246 skin conductance electrode paste (Med Assoc. Inc.) and attached to the middle phalanges of the first and second fingers of the subject’s nondominant hand after hands were washed with soap and water.

The digital data collection was controlled by Psylab7 software, and all physiological signals were sampled at a rate of 500 Hz.

**2.6. Video recording of the game**

During the game, the output signal (video and audio) from the GameCube was stored as digital video (25 frames per second) with the V1d Random Access Video Recorder/Player (Doremi Labs, Inc., Burbank, CA). Psylab7 software was used to trigger the V1d Disk Recorder to start recording the game screen at the same time when the physiological data collection started. After calibrating the timing, the recorded video image of the game screen was in time synchrony with the physiological data with a one-frame (40-ms) accuracy.

**2.7. Event scoring**

The exact onset time of reaching the goal was determined by examining each of the played games, frame by frame, using V-ToolsPro 2.20 software. The onset times were saved as CSV files that were then converted with special software (V1 Clip Converter for Psylab) and imported into Psylab7 software. The participants reached the goal, on the average, 6.6 times (range: 4 to 9).

**2.8. Data reduction and analysis**

Mean values for the psychophysiological measures (i.e., facial EMG, SCL, IBI) were derived for two 1-s epochs before reaching the goal (this provides a local baseline; seconds 1 and 2) and for six 1-s epochs after the attainment of the goal (seconds 3 to 8).

The data were analyzed by the Linear Mixed Models procedure in SPSS with restricted maximum likelihood estimation and a first-order autoregressive covariance structure for the residuals. Participant ID was specified as the subject variable, and the sequence number of an event and second (seconds 1 to 8) were specified as the repeated variables. We created seven (orthogonal) contrast variables that compared the different seconds; for example, Contrast 1 was seconds 1 and 2 (local baseline = Time 1) vs. seconds 3, 4, 5, 6, 7, and 8 (time after event onset = Time 2). These contrast variables and Spatial Presence were entered as covariates. This results in an analysis that is identical to an analysis with second as a factor, but enables the.
examination of interactions between second (or the contrast variables) and the (continuous) Spatial Presence score. We specified a fixed-effects model that included (a) the main effects of the seven contrast variables, (b) the main effect of Spatial Presence, and (c) the Contrast \(1 \times \) Spatial Presence interaction. When interpreting the results, high and low Spatial Presence was defined as scores 1 \(SD\) above and 1 \(SD\) below the mean on the Presence measure, respectively.

3. Results

The average Spatial Presence rating was 2.7 (\(SD = 0.8\)). In agreement with Hypothesis 1, the Linear Mixed Models procedure showed that the Contrast \(1 \times \) Spatial Presence interaction was significant in predicting zygomaticus major EMG activity, \(F(1, 166.12) = 9.47, p = .002\). Inspection of the estimated marginal means revealed that reaching the goal elicited an increase in zygomatic activity for participants reporting high Spatial Presence during the game (\(Ms = 2.69 \ln[µV]\) and \(2.76 \ln[µV]\) for Time 1 and Time 2, respectively), whereas it prompted a decrease in zygomatic activity for participants reporting low Spatial Presence (\(Ms = 2.97 \ln[µV]\) and \(2.93 \ln[µV]\) for Time 1 and Time 2, respectively).

In addressing Hypothesis 2, it was found that the Contrast \(1 \times \) Spatial Presence interaction only approached statistical significance when predicting orbicularis oculi EMG activity, \(F(1, 171.84) = 3.71, p = .056\). Attaining the goal tended to elicit a more pronounced increase in orbicularis oculi activity for participants reporting high Spatial Presence (\(Ms = 2.42 \ln[µV]\) and \(2.59 \ln[µV]\) for Time 1 and Time 2, respectively) compared to those reporting low Spatial Presence (\(Ms = 2.15 \ln[µV]\) and \(2.26 \ln[µV]\) for Time 1 and Time 2, respectively).

As hypothesized (Hypothesis 3), there was a significant Contrast \(1 \times \) Spatial Presence interaction in predicting corrugator supercilii EMG activity, \(F(1, 145.64) = 17.07, p < .001\). Reaching the goal elicited a decrease in corrugator activity for participants reporting high Spatial Presence (\(Ms = 2.88 \ln[µV]\) and \(2.80 \ln[µV]\) for Time 1 and Time 2, respectively), but not for participants reporting low Spatial Presence (\(Ms = 3.70 \ln[µV]\) and \(3.71 \ln[µV]\) for Time 1 and Time 2, respectively).

In agreement with Hypothesis 4, the Contrast \(1 \times \) Spatial Presence interaction was significant in predicting SCL, \(F(1, 35.67) = 7.48, p = .010\). Attaining the goal prompted a more pronounced reduction in SCL for participants reporting high Spatial Presence (\(Ms = 0.640 \log[µS]\) and \(0.637 \log[µS]\) for Time 1 and Time 2, respectively) compared to those reporting low Spatial Presence (\(Ms = 0.678 \log[µS]\) and \(0.677 \log[µS]\) for Time 1 and Time 2, respectively).

As expected (Hypothesis 5), the Contrast \(1 \times \) Spatial Presence interaction was nonsignificant in predicting IBI, \(F(1, 200.35) = 0.01, p = .930\).

4. Conclusions

This study is the first to show that Spatial Presence is related not only to the arousal dimension of emotions, but also to the valence dimension as measured by facial EMG. The results showed that self-reported Spatial Presence was associated with an increase in zygomatic and orbicularis oculi EMG activity in response to reaching the goal in the video game (i.e., a positive event; in the case of orbicularis oculi activity, the association narrowly failed to reach statistical significance, however). That is, given that zygomatic and orbicularis oculi activity index positive emotions [6], high Spatial Presence resulted in more positive emotional responses to a positive event.

The results also showed that Spatial Presence was strongly associated with a decrease in corrugator supercilii EMG activity in response to reaching the goal. That is, given that corrugator supercilii EMG activity indexes negative emotions [6], high Spatial Presence resulted in a greater reduction in negative affect in response to a positive event.

The bipolar valence dimension has been suggested to represent the integration of two separable and partially distinct components of the affect system, that is, positive affect and negative affect [16, 19]. It is somewhat unclear, however, whether zygomatic and corrugator activities are uniquely sensitive to positive affect and negative affect, respectively, or whether positive and negative affect have reciprocal effects on facial EMG measures [16]. Thus, the stronger relationship of Spatial Presence with corrugator compared to zygomatic activity may (a) indicate that Spatial Presence is more strongly related to negative affect compared to positive affect or (b) be due to the previously observed stronger linear effect of valence on activity over corrugator supercilii versus zygomaticus major [14].

Several authors have suggested that psychophysiological parameters might be used as measures of Presence [11, 20]. The present study provides evidence for the validity of facial EMG, particularly corrugator supercilii activity, as a measure of Presence. An apparent limitation of the present study was, however, that we did not compare different mediated environments putatively eliciting varying levels of Presence [cf. 11]. Therefore, sensitivity of facial EMG as a measure of Presence remains to be established. It should also be emphasized that facial EMG is primarily a measure of hedonic valence (or positive and negative emotions) rather than a direct measure of Presence [6]. This is the case for all psychophysiological measures, however (i.e., they are primarily measures of arousal or attention) [6].

It is also of note that we measured phasic EMG responses to a specific event in the mediated environment. There is no reason to expect that tonic EMG activity averaged across a longer exposure to a mediated environment involving both positive and negative events would be related to Spatial Presence. That is, when examining tonic EMG activity, Presence-related increases in positive and negative responses would be expected to cancel each other out. Thus, in most connections, only
phasic EMG responses may be useful as measures of Presence.

In agreement with prior studies [10, 11], the results showed that Spatial Presence was positively related to arousal as measured by EDA. Although all participants experienced a reduction in (anticipatory) arousal after attaining the goal, this reduction was less pronounced for participants reporting high Spatial Presence compared to those reporting low Spatial Presence. An apparent advantage of EDA as a measure of Presence-related emotional arousal is that, physiologically, it is influenced only by the SNS (this results in high diagnosticity) [6].

In disagreement with some prior studies [10, 11], we found that Spatial Presence was unrelated to cardiac IBI changes. This is probably because the video game (i.e., an active coping task requiring sensory intake) may elicit both emotional arousal and attentional engagement. That is, in this connection, high Spatial Presence may prompt heightened (a) emotional arousal accompanied by SNS activity and (b) attention accompanied by PNS activity. Thus, given the opposing effects of the SNS and PNS on IBI [6], Spatial Presence is likely to be unrelated to IBI when the mediated environments prompts both arousal and attentional engagement (sensory intake). However, IBI (or HR) may be a useful measure of Presence when the mediated environment elicits only (or mostly) arousal or only (or mostly) increased attention, Presence being negatively and positively related to IBI, respectively.

In interpreting the present findings, the correlational nature of the present study should be allowed for. That is, in principle, it is possible that emotional reactions exert an influence on Spatial Presence rather than vice versa. For example, it is possible that emotional arousal influences the distribution of attention, thereby exerting an effect on Spatial Presence [see 6]. However, we feel that it is unlikely that the valence dimension of emotions would have a causal effect on Presence.

In sum, the present study 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. Spatial Presence was negatively and particularly strongly related to activity over corrugator supercilii. Spatial Presence was also related to higher arousal as measured by EDA, but it was unrelated to IBI. Spatial Presence may exert an influence on both the valence and arousal dimension of emotions.

References


Using Goffman’s Frameworks to Explain Presence and Reality

Ruth Rettie
Kingston University
RM.Rettie@Kingston.ac.uk

Abstract
This paper defines presence in terms of frames and involvement [1]. The value of this analysis of presence is demonstrated by applying it to several issues that have been raised about presence: residual awareness of non-mediation, imaginary presence, presence as categorical or continuum, and presence breaks.

The paper goes on to explore the relationship between presence and reality. Goffman introduced frames to try to answer the question, “Under what circumstances do we think things real?” Under frame analysis there are three different conditions under which things are considered unreal, these are explained and related to the experience of presence. Frame analysis is used to show why virtual environments are not usually considered to be part of reality, although the virtual spaces of phone interaction are considered real. The analysis also yields practical suggestions for extending presence within virtual environments.

Keywords--- presence, frames, virtual environments, mobile phones, Goffman.

1. Introduction

The conventional conceptualization of (tele)presence is ‘being there’ [2]. This creates a paradox because one cannot be in two places at once; to avoid this paradox the virtual environment is seen as unreal, however realistic the experience [3]. Goffman [1] recognized that different ‘worlds’ compete for our attention, and that there is a relationship between focus of attention and our experience of a world as ‘real’. His frame analysis is designed to explain the circumstances in which we consider an environment real. Its application to virtual environments is useful in two ways: firstly, it helps to conceptualize the experience of presence in a virtual world, and secondly, it provides practical guidelines for increasing the perceived reality of virtual environments.

This paper extends Goffman’s framework theory to presence, in two different ways. Firstly, presence is defined in terms of two concepts: frames and involvement. This analysis of presence is applied to the literature on presence, comparing it to other definitions and applying it to several contentious issues of presence.

The second part of this paper uses frame analysis to explore the relationship between presence and reality. Frame analysis attempts to explain why certain experiences are considered real and others are not. Two different contexts, virtual reality environments and phone interaction, are used to clarify the relationship between presence and reality.

2. Frameworks

Goffman [1, p. 21] introduces the concept of frames as follows: "When an individual in our Western society recognizes a particular event, he tends, whatever else he does, to imply in this response (and in effect employ) one or more frameworks or schemata of interpretation ... [which] is seen as rendering what would otherwise be a meaningless aspect of the scene into something that is meaningful." In other words, we use frames to interpret our experience. Frames answer the question, "What is going on here?" [p. 46]. Frames are socially shared and culture specific.

Frames organize experiences; they provide assumptions about what is going on. Frames are not mental objects, but concepts used to decipher what is happening around us, "observers actively project their frames of reference on the world immediately around them" [1, p. 39]. Frames provide contexts which enable our interpretation of events. The same section or ‘strip’ of experience can take different frames. For instance, something taken at first as a marriage ceremony may be reframed as a rehearsal. What is perceived to be ‘going on’ depends on the frame applied.

This analysis is related to Sheridan's view of a virtual environment as a "mental model" that represents a physical environment [4,5]. Lackner and Dizio [cited by Schuemie et al., 6] suggest that when people have difficulty forming a mental model of a space, they report a loss of a sense of presence. However, ‘mental model’ suggests a mental picture; frames are not additional mental entities but just the way that experience is conceptualized. Goffman’s concept of frame was derived from the work of Bateson [7]. The concept is related to, but is less prescriptive than, Minsky’s concept of frame [8] and Schank and Abelson’s concept of script [9].
3. Presence

Presence is the phenomenological experience of being in a situation or environment, the sensation of "being there" [10, p. 3959]; telepresence is the experience of being somewhere other than where one physically is. Presence can be analyzed in terms of two concepts used by Goffman [1], involvement and frame: presence is engrossing involvement in a spatial frame. If presence is 'being there', involvement relates to 'being' and the frame explains what is meant by 'there', it defines the situation or environment. The term 'involvement' is not used in the sense of interest in the content of an experience, but to describe the allocation of attention. For example, the experience of presence in a theatrical performance means that one is focused on an experience that is framed as being a play. When someone is engrossed in a situation or experience they feel present; frames define the nature of this presence.

This does not mean that the frame is predominant, as one becomes more engrossed there is less awareness of the frame. However, the frame continues to define how the experience is interpreted, for instance, seeing a murder on stage one does not feel an obligation to call the police, as one might if the same experience is seen on the street. One may be present in the ordinary world, a daydream, an imaginary world, or a mediated experience, etc. The extent of involvement in the frame relates to the degree of presence experienced.

3.1. Presence as framed experience

Frames explain what is meant by 'being in' a mediated environment. The mediated environment may be framed as a space or a place. Through the use of a physical metaphor or frame, the virtual environment becomes a space where presence is experienced. The frame both constructs and makes sense of the experience. The frame provides the context, as Heeter notes, "presence requires a context" [11, p. 339]. Schroeder [12, p. 10] uses frames in his analysis of virtual environments, "...VEs have a different kind of 'bandwidth' from real world frames for presenting the self to the other. When we enter a VE, a shift in the 'frame' takes place..." In a shared virtual environment the frame is shared by the participants in the interaction.

Following Giddens [13] it is useful to distinguish places from spaces. Places are physical settings for social interaction, they contain social norms; we are located in space but we act in place. Places are spaces, but not all spaces are places. Electronic mediation enables interaction without being in the same physical place; mediated action and interaction occur in spaces, but not in places. Virtual reality environments may be framed as 'places' or 'spaces'. Participants in a phone call frame their interaction as occurring in a separate 'space', not a place. This helps to explain the common practice of holding private conversations in public places, the participants are 'on the phone', which is perceived as a private space. In a mediated environment, the frame allows one to feel as if one is 'there' and/or to feel as if one is 'meeting with' the other interactant(s). If the environment is framed as a space rather than a place, the paradox of being in two places at one time is avoided.

3.2. Presence as involvement

Involvement [1, p. 346] is a "psychobiological process in which the subject becomes at least partly unaware of the direction of his feeling and his cognitive attention." For Goffman attention or 'involvement' is shared between the ordinary world and 'aways' such as day dreams, hallucinations, self involvements [14, p. 243]. Involvement or attention is divided between these different 'involvements', resulting in different degrees of involvement in each 'realm'. Updating this theory, it is clear that mediated communication can engage our cognitive attention, reducing our presence in the immediate environment. Furthermore, attention is focused on the experience per se rather than on the mediation of the experience: "Speaking to someone on the telephone, is so natural, that we almost forget about the intervening medium" [15, p. 109].

"Engrossables" are experiences that are so involving that one can be "caught up in or carried away by [them]" [1, p. 46]. This paper contends that this creates a feeling of presence in the environment; presence is engrossing involvement in a spatial frame. This is involvement in the framed experience as a whole, not on the contents of the frame, or in the frame itself. In fact, when one is 'carried away' by the experience the frame becomes less apparent Being involved in an experience means that a considerable share of mental awareness or attention is devoted to this experience, this does not imply a focus on the direction of attention. 1.

Involvement here is about allocation of attention to the experience as a whole, and is neither about involvement in some element of the experience nor about involvement in the framing concept. As Slater [17] points out, one can be present even if the content of an experience is uninteresting and un-involving. He comments that it is possible to listen to a quadraphonic sound system and feel present as in a theater before an orchestra, despite feeling uninvolved in the music played. Slater argues that what is relevant is the 'form' of the experience, not the content. Frame analysis suggests that form and content interact and are not independent: the content, or what is understood as 'going on', depends on the frame applied. Using Slater’s example, if the music is not consistent with the frame of being in the theater, in front of an orchestra, the frame will break, consequently the illusion and the feeling of presence will be lost. In addition, if the music is so uninteresting that engrossment is lost and attention is reallocated to the physical environment, then the illusion, and the feeling of presence, will be lost. However, the experience as a whole, rather than the just the music, may be sufficiently absorbing to retain involvement, so that presence can be experienced despite boring music.

1 In the extreme case of flow [16] almost all attention is focused on a single frame.
In the frame analysis developed here, what is important is engrossment in the experience itself and the conceptualization of the experience within a spatial frame, as a coherent space. The ‘form’ of the experience is given by the nature of the frame used to conceptualize the experience. The involvement is in an experience framed as being in a theater; not in the music per se, nor in the idea of being in the theater, but in the experience as a whole.

The concepts of involvement and frame are related. Frames organize involvement; involvement is never total, it is restrained by "a measure of cognitive reserve ...a wisp of doubt concerning framework and transformations" [pp. 379-80]. Frames can break, because of sudden awareness of incoherence; the frame is no longer relevant to the experience, 'flooding out' occurs and there is an immediate loss of involvement in the frame. Conversely, 'flooding in' occurs when someone suddenly becomes involved in a frame, e.g. someone may become suddenly engrossed in a television program. The transfer of involvement between frames can also be a gradual shift of attention, and consequently of presence, between different worlds.

There is some empirical evidence of a relationship between the experience of presence and allocation of attention. Hoffman et al. [18, p. 1245] subjected research participants to moderate pain while exposing them to a virtual reality environment and found that "the more attention drawn into virtual reality, ...the less pain patients experience." Attention to pain (measured by subjective reports of the time spent thinking about pain) was significantly reduced when participants were exposed to a virtual reality environment. Moreover, magnetic brain imaging showed a reduction of pain-related brain activity of over 50% during exposure to a virtual reality environment.

4. Presence as framed involvement

In this section the analysis of presence in terms of framed involvement is related to the literature on presence, comparing it to other definitions and applying it to contentious areas of presence. The concept of presence is ill defined, as Waterworth and Waterworth [19, abs.] comment, "Progress in understanding presence is inhibited by the fact that we are unable to agree what it is we are talking about."

4.1 Presence as non-mediation

Lombard and Ditton [20, para. 33] define presence as "the perceptual illusion of nonmediation". This corresponds with deep involvement, one is so "caught up in or carried away by" the mediated environment that there is an illusion of non-mediation, i.e. loss of awareness of the ‘mediation’ frame.

Although this definition of presence is helpful in identifying the engrossment of presence, there are several problems with it. Firstly, it only applies to presence in mediated environments; this prevents its application to physical and imaginary environments. Secondly, the illusion of non-mediation either occurs or does not occur; this implies that presence is not a matter of degree. The authors explain the subjective experience of degrees of presence in terms of the number of instants during which the illusion of nonmediation occurs. This suggests a perpetual mental assessment of which one is unaware, which is epistemologically problematic. A further problem with this definition is that even if the feeling of presence in the mediated environment is very strong, at some level one is aware of mediation and misperception.

This mental equivocacy is reflected in the definition, which was refined in the Presence-L Listserv [quoted by 6, p. 185]. "Presence (a shortened version of the term "telepresence") is a psychological state or subjective perception in which even though part or all of an individual's current experience is generated by and/or filtered through human-made technology, part or all of the individual's perception fails to accurately acknowledge the role of the technology in the experience. Except in the most extreme cases, the individual can indicate correctly that s/he is using the technology, but at *some level* and to *some degree*, her/his perceptions overlook that knowledge and objects, events, entities, and environments are perceived as if the technology was not involved in the experience."

There are two reasons for this residual awareness of non-mediation. Firstly, involvement is allocated between different realms; there is always some residual involvement with the physical body and with the physical world. Secondly, engrossment is never total, it is restrained by "a measure of cognitive reserve ...a wisp of doubt concerning framework and transformations" [1, p. 379-80]. This is because we are aware that the application of a frame is contingent, there is always the chance that the frame will no longer fit the experience. When this happens, involvement changes suddenly, there is a frame break, which results in a break in presence.

4.2 Elements of presence

Lombard and Ditton [20] review the literature and distinguish six related concepts of presence: realism; transportation; immersion; social richness; social actor within medium, medium as social actor. The last three relate more to social presence rather than locational presence [10], but the first three are key elements of presence: realism, is the extent to which the experience is realistic and like the real world, transportation is the sensation of being in a remote physical environment and immersion is the extent to which the senses are engaged in the mediated communication. Slater [21, p. 261] adds dominance, response and memory. "The extent to which the VE becomes the dominant one i.e., that participants will tend to respond to events in the VE rather than in the 'real world'...The extent to which participants, after the VE experience, remember it as having visited a 'place' rather than just having seen images generated by a computer." Draper, Kaber and Usher [22, p. 366] emphasize attention, "The more attentional resources that a user devotes to stimuli presented by the displays, the greater the identification with the computer-mediated environment and the stronger the sense of telepresence.” Witmer and Singer
These elements can all be covered using the analysis of presence in terms of frame and involvement. If the virtual environment is framed as a place, this explains the feeling of transportation (movement to a different place), located memories and response to an environment. The remaining elements of presence, that is, immersion, engagement, dominant focus, and attention, are all aspects of involvement or engrossment. This indicates the robustness of the analysis of presence as framed involvement. The analysis can also be applied to several areas of interest in this field: the experience of presence in imaginary environments, presence as a continuum or categorical concept, and presence breaks.

4.3. Presence in imaginary environments

Biocca [24, para. 6.1] claims that "at one point in time, users can be said to feel as if they are physically present in only one of three places: the physical environment, the virtual environment, or the imaginal environment. Presence oscillates among these three poles." Heeter [11] and Blascovich [25] also refer to presence in imaginary environments. However, Slater [17], restricts presence to physical and virtual environments, arguing that, while we may be engaged or involved with imaginary worlds, this is not presence. Waterworth and Waterworth [19, para. 1.4], also dispute presence in imaginary environments, arguing that presence involves a focus on external stimuli, and that imaginary environments do not occur in real time.

Analysis of presence as involvement and frame extends the experience of presence to imaginary worlds; they are engrossing, and they deflect attention from the everyday world. However, although couched in terms of presence, the underlying issue here is the question of reality. Whether or not we choose to apply the word 'presence' to imaginary worlds, the question highlights two underlying distinctions: that between reality and fictitious realms, and that between shared and private worlds. This former is discussed in section 5, which discusses the relationship between presence and reality, and explains how we can feel present in realms that we do not consider to be real.

4.4. Presence as a Continuum

Draper, Kaber and Usher [22] claim that presence is a scale of varying intensity, but Biocca's three poles model assumes that presence in the different modes is exclusive. Slater and Steed [26, p. 419] attempt to resolve the conflict between presence as a continuum, and presence as categorically physical or virtual. "We can think of presence as a selector amongst environments to which to respond, which operates dynamically from moment to moment. If it were possible to 'freeze time' at a specific instant, then the individual would be paying attention and responding to a set of stimuli corresponding to one environment, not paying attention to all the other stimuli ... the set of stimuli of the 'present environment' forms an overall gestalt, providing a consistent believable world in itself." This conceptualizes presence rather like the editing suite of a live TV feed; the 'live' camera corresponds to presence; it relates to Goffman's 'engrossables' and the way one is 'carried away'. However, it is problematic in its suggestion of mental processes of which we are unaware. The approach also conflicts with the residual awareness of mediation experienced in a mediated environment. Ijsselsteijn [27, p. 253] makes this point, arguing that unlike the familiar duck/rabbit gestalt, "Both medium and physical environment are distinct entities which may be perceived at the same time. ... Rather, a break in presence may be conceived of as an attentional shift away from the mediated environment and towards the physical environment, but with the possibility to still feel a sense of presence in the mediated environment, albeit to a lesser extent."

The analysis of presence as framed involvement portrays presence as a continuum rather than an all-or-none concept of presence. However, frame breaks help to explain the attraction of the categorical interpretation. When a frame breaks, "...the individual's situation can collapse, disintegrate, go up in smoke..." [1, p. 302]; involvement changes suddenly rather than gradually, creating the impression that presence in different environments is exclusive.

4.4. Breaks in presence

The concept of 'breaks in presence' was operationalized by Slater and Steed [26]. Breaks in presence are transitions between absorption in different spheres, e.g. between the virtual and everyday world. Slater and Stead note that reporting breaks in presence is not possible on transition into the virtual world: "The reporting of transitions into the state of absorption is impossible without undermining the absorbed state itself." [p. 14]. However, transition to the physical world can be reported, and they developed a 'virtual presence counter'; this enables research respondents to signal when the laboratory, rather than the virtual environment, is dominant. The authors report a significant negative correlation between presence breaks and reported presence in the virtual environment.

If presence is construed as framed involvement, a break in presence may consist simply of a reallocation of involvement and a focus on a different frame, or it may reflect the dissolution of a frame. Return to the ordinary world may involve a sudden transition, when absorption in a virtual or imaginary frame is suddenly fractured. Schutz describes these as 'shock experiences' as we move from one world to another [cited by 1, p. 4]. Goffman [1] writes of 'containment' within a frame, this is disrupted by inconsistencies between the frame and the experience. Social interaction in a shared frame can reinforce containment within the frame, but if one person loses involvement this may break the involvement for others. When frames break, involvement and presence is dissipated. This relates to the claim by Walker and Davide [28] that it is breaks in presence that are experienced, rather than presence itself; they define presence as the "absence of breaks in presence".
5. Presence and reality

Brandt and Metzger [29] distinguish four meanings of reality. Reality1 is the objective world that is presumed to exist beyond our subjective experience. The subjective, phenomenal experience of the world is reality2. Reality3 is that which we directly encounter as opposed to that which is merely represented in the mind (e.g. thoughts). Reality4 is the extent to which something is experienced as real.

Reality1 and reality2 relate to the positivist and phenomenological ideologies. Positivists believe that there is an objective knowable reality (reality1), while phenomenologists focus on the world of subjective experience or reality2. Different philosophies pertain to the relationship between reality1 and reality2; social constructionists [30] contend that reality is constructed and there is no reality1, social constructivists [31] merely contend that reality1 is unknowable. Mantovani and Riva [3] claim that most scientists take the rationalist position of ‘ingenious realism’. This involves a concept of reality as both external and knowable; that is, knowledge relates to reality1, which is knowable through reality2. Mantovani and Riva maintain that this ontological position creates a problem for virtual reality, but not for teleoperations systems. The latter involves remote control using robotic controls, i.e. it involves remote presence in the ordinary physical world. However, virtual reality is problematic under ingenious realism because it relates to things which do not exist, but which create sensory experience; it is therefore akin to hallucination. This dilemma, they contend, is created by ingenious realism, which assumes a dichotomy between mediated and unmediated experience, whereas all experience is mediated. The authors adopt a social constructionist view, where reality is defined within a cultural framework by negotiation of action and meaning. Under this perspective, presence depends not on fidelity but “on the capacity of simulation to produce a context in which social actors may communicate and cooperate.” [p 538]. This social constructionist view of reality distinguishes shared worlds, such as virtual environments and the physical world, from imaginary private worlds, such as day dreams and hallucinations, where there is no collaborative construction.

Reality3 is concerned with distinguishing things in themselves from their representations e.g. the physical world is contrasted with what is imagined or thought. Applying Brandt and Metzger’s definition, Heeter [11] concludes that reality3 applies to virtual environments because they impinge directly on the senses, as opposed to simply being represented in the mind. However, the definition of reality3 is unsatisfactory; using the authors’ example, under this definition one’s liver is not real because it is not encountered, whereas dreams and hallucinations are real.

Frame analysis is also a form of social constructionism, in that frames are culturally relative and confer meaning; the theory means that reality2 does not correspond directly with reality1. However, this paper, and Goffman’s frame analysis, is more concerned with reality4, the extent to which we consider our experiences real or genuine. This philosophy of language approach explores the use of the word ‘real’, rather than the metaphysical question of what is reality. This approach has practical application and can be applied to the design of virtual environments, encouraging participants to consider them real. This is discussed in Section 5.5. Frame analysis also introduces a further meaning of reality, analogous to reality3, which usefully distinguishes between experiences in different realms, e.g. between the ordinary world and hallucinations.

5.1. Immersion and reality

Slater defines ‘immersion’ [17,32] as the extent to which a virtual reality environment shuts out the ‘real world’, and has rich representational capability. This use of immersion contrasts with that of Lombard and Ditton discussed in Section 4.2. For Slater immersion is an objective quality of the environment corresponding to its fidelity to the real-world sensory experience. Immersion is not the same as presence, which is a subjective, conscious experience that can occur in unrealistic fantasy worlds. However, Slater [17] states that there is probably a strong empirical correlation between immersion and presence. Immersive environments are likely to be more engrossing, promoting the experience of presence. More immersive environments are more realistic or more ‘real’, but to say that something is more, or less real, is actually to deny that it is real. Immersion makes an environment more realistic, but it does not make it real.

5.2. Reality and frame analysis

In his introduction to Frame Analysis [1, p.2] Goffman positions this work as an ontological enquiry, in the tradition of William James; it is an attempt to answer James’ question, "Under what circumstances do we think things are real?" Frame analysis attempts to identify the elements of situations that convince us they are genuine.

The concept of frame introduced in Section 2 is a simplified version of Goffman's frameworks; it ignores the difference between primary frames and their transformations. To accommodate imaginary and illusionary experiences, Goffman uses two frame ‘transformations’: keyings and fabrications. Fantasy, daydreaming and various forms of drama are 'make-believe' keyings. On the other hand, hallucinations and dreams are fabrications because they involve (self) deception. Transformations are cues that something is not real. Transformations can relate to whole realms, e.g. an imaginary world, or to experience within the ordinary physical world, which is not real in the sense that it is not what it appears to be, e.g. a play fight. Recognition of a
transformed frame is one way of deciding that something is not real. Goffman did not specifically consider mediated environments as a category; most virtual reality environments have the special status Goffman gives to theatrical performances, where we are willingly "transformed into collaborators in unreality", these are "voluntarily supported benign fabrication(s)" [p. 136].

For Goffman, what makes something real relates to both involvement and frame. He notes that it is their potential for inducing engrossment that makes other worlds seem ‘real’ [1, p. 347]. Worlds where we become engrossed or carried away seem real; it is this same involvement which induces a sense of presence in that world. Engrossment or involvement is therefore one aspect that makes things seem real. The reality of an experience also depends on how it is framed.

There are two ways in which an experience may be deemed unreal in virtue of the frame applied. Firstly, the type of frame is relevant, whether something is real or unreal depends on the frame used, under one frame it may be real and under another not real; without a frame it is meaningless. If the frame is a keying or fabrication the experience framed is deemed unreal. For example, a theatrical play is not real because drama is keyed as make believe. In other words, application of a transformed frame automatically implies that the experience is not considered real. Secondly, an experience may be unreal because the frame applied does not fit the experience. So an event framed as murder is shown to be unreal when the victim gets up. When we perceive that the frame does not ‘contain’ the experience, the frame ‘breaks’, we realize that the experience is not real. “Experience ... finds no form and is therefore no experience. Reality anomalically flutters.” [1, p. 302]. If the scene is framed as street theater, the recovery of the victim does not disrupt the frame; both involvement and presence are sustained. However, the event would still not be real because the frame, ‘street theater’ is a ‘make believe’ frame; this makes the framed experience unreal.

Imaginary worlds are not part of the ordinary world; framed as imaginary they are unreal because of the transformed frame, framed as part of the physical world they are not real because the experience does not match the frame. Frames and transformations thus help to distinguish what we can call real.

The concepts of presence and appearance of reality are related in that both are induced by engrossment, and disrupted by frame breaks. However, presence can occur in worlds recognized as unreal, for instance, in the theater or in a day dream.

5.3. Reality and mediated communication

In phone calls there is both remote presence and virtual presence. Phone conversation occurs simultaneously in two physical locations, remote presence of the other person is experienced at each location. In addition, there is the shared virtual space, the participants meet ‘on the phone’. In phone calls sound is produced sequentially and is transient, so that interpretation requires continuous attention, furthermore, silence is not socially acceptable. Consequently, phone calls require a large allocation of attention. A phone call is a frame in which we become engrossed, but phone calls are not framed as places. This may be because they are not visual and have few spatial cues. A phone call is conceptualized as a space where people meet, the meeting is real but there is no meeting place. This avoids disruption of the frame; it is perfectly possible to feel present in the phone call and in a physical location at the same time. Videophones show the physical location of each participant; there are actually three spaces, that of each participant and the phone space. However, there is no inherent inconsistency because each participant is only deemed to be in one place at a time.

Virtual reality environments are often framed as places. They may include visual sensory experience and incorporate three dimensional locational cues, with separate areas and movement between different locations. These increase immersion (representational fidelity) and presumably presence. However, this encourages use of a ‘place’ frame which creates problems. The physical location of a person is given by that of his physical body, and it is impossible to be in two places at the same time. The frame ‘place’ therefore includes an inherent incoherence, which reduces containment in the frame. There are various ‘self-involvements’ such as stomach rumblings which automatically refocus attention on the body creating awareness of conflicting locations.

5.4. Staying in the frame

Containment in the frame depends on coherence between the experience and the frame, but this coherence, in turn, depends on the frame employed. For example, if a make believe frame is applied, for example, ‘theater’, then coherence between frame and experience is easier to achieve. This is because the frame includes a convention of ‘suspension of disbelief’ and imposes fewer constraints on the types of experience that are consistent with the frame. On the other hand, if a ‘real world’ frame is applied this requires ‘real world’ fidelity to prevent dissolution of the frame. The frame applied is not completely dependent on the participant, because it is cued by the experience itself. For instance, in the theater the stage and curtains signal the relevance of the theater frame.

Frame theory can be applied to the design of virtual environments and used to facilitate containment in the frame, thus reducing frame breaks and prolonging presence in the environment. There are a number of factors that encourage containment within the frame. These include consistency between frame and experience, social interaction with others who appear to apply the same frame, emotional involvement in the framed experience, and
consistency with other frames. Goffman also describes how the Shakespearian device of a play within a play has the effect of increasing involvement and containment in the theater frame. The actors and audience share the play-within-a-play frame; this makes the actors seem more real and encourages engagement.

Locational cues that promote the use of a place frame create potential inconsistency between the physical location of the participant and his apparent location in the virtual world. This conflict may be exacerbated by a virtual reality helmet; ideally participants should be physically comfortable to minimize self-involvements, which draw attention to the physical world.

The phone space in a phone call illustrates how a virtual space allows participants to be comfortable in two spaces at one time. Alternatively, the virtual element may be seamlessly integrated within the physical environment, e.g. in augmented reality. On the other hand, locational cues encourage action within the virtual environment, this may promote presence, especially if the action is cooperative, in that it reinforces the frame.

Involvement and containment in a theater frame is supported by conventions which clearly signal the ‘make believe’ frame, for instance the use of a narrator, and the conventions of dramatic discourse, etc. In addition, the stage and curtains separate the performance place from the audience, preventing conflict between the two spaces. Adopting a frame that signals unreality may reduce conflict and increase presence. Virtual environments can develop and adopt conventions which signal virtuality, for example, the use of avatars. While this also signals that the environment is unreal or fabricated, it may nevertheless promote involvement and containment in the frame, and therefore prolong presence. Perversely, in this way, a lack of fidelity may improve presence by signaling the virtual frame, presence is extended at the expense of reality.

6. Conclusions

Frame analysis helps to clarify the concept of presence and its relationship to reality. Experience, contents and frame interact in the construction of the perceived experience. Presence in the virtual environment occurs when the experience is engrossing but is lost when there is inconsistency between the frame applied and the contents. Paradoxically this can mean that if the contents are too realistic they may conflict with the virtual frame. Locational cues which suggest that the user is physically in the environment are particularly liable to create conflict with awareness of actual physical location. For some applications of virtual environments unreality may be a benefit, for example, in the treatment of phobias the patient’s awareness of the virtual frame enables a controlled level of engagement with the object of his phobia.

Presence is engrossing involvement in a spatial frame or transformed frame. Frames explain what is meant by being in a mediated environment. They resolve the paradox of how one can experience different environments or realms at the same time, without being in two places at one time. Involvement allocation explains the degree of presence experienced. The flexibility of framework theory accommodates imaginary, virtual and physical worlds and the many nuances of these basic modes. The theory can also be used to elucidate the concept of reality, and its relationship to presence.

There are three different grounds for considering an experience real: engrossment, containment within a frame, and use of an untransformed frame. The first two also promote presence, but the unreality of a transformed frame does not prevent the experience of presence, and may extend containment.

In terms of mediated communication, a phone call is clearly part of reality; although the phone environment where the interactants meet is a virtual space, it is considered real. On the other hand, while playing a virtual reality game is a real activity, the game environment, however engaging, is not. This is because it is framed as a place, when it is clearly not physical; the place frame conflicts with residual awareness of their actual physical location undermining containment in the frame.

What makes something seem real is engrossment, the relationship between the frame and the content of the experience, and the type of frame applied. Frame analysis identifies as unreal both experiences which are misframed and those where although the correct frame is applied, the type of frame indicates that, in some sense, they are not real. Goffman’s use of ‘real’ is rather extreme, because although daydreams and virtual environments are not part of the ordinary everyday world, they are real experiences. Similarly, if I watch street theater this is a real experience, but it consists of two different frames, ‘being in the street’ and ‘street drama’; the keyed frame of the latter indicates that what is seen cannot be taken at face value. In both these situations there is an element of unreality, things are not quite what they seem, and in this way they are ‘unreal’; paradoxically this makes reality consistent with a degree of unreality.

References

Abstract

The visually induced illusion of ego-motion (vection) is known to be facilitated by both static fixation points [1] and foreground stimuli that are perceived to be stationary in front of a moving background stimulus [2]. In this study, we found that hardly noticeable marks in the periphery of a projection screen can have similar vection-enhancing effects, even without fixating or suppressing the optokinetic reflex (OKR). Furthermore, vection was facilitated even though the marks had no physical depth separation from the screen. Presence ratings correlated positively with vection, and seemed to be mediated by the ego-motion illusion. Interestingly, the involvement/attention aspect of overall presence was more closely related to vection onset times, whereas spatial presence-related aspects were more tightly related to convincingness ratings. This study yields important implications for both presence theory and motion simulator design and applications, where one often wants to achieve convincing ego-motion simulation without restricting eye movements artificially.

Keywords— Vection, psychophysics, spatial presence, virtual reality, illusions, motion simulation.

1 Introduction

The typical apparatus to investigate this striking phenomenon consists of a rotating drum which is painted with simple geometrical patterns like black and white vertical stripes. Investigating vection, however, can have important implication for the emerging field of virtual reality (VR) and multi-media applications, where one would often wish to convey a convincing sensation of self-motion to the user by just presenting a visual motion, without having to physically move the observer.

To bridge the gap between fundamental research and recent computer-based applications, a high-end virtual reality setup was used to present the moving visual stimuli. This had the additional advantage of enabling us to easily present not only simple and abstract geometrical patterns, but also photorealistic renderings of a natural scene. Using such stimuli for vection research has, to the best of our knowledge, hardly been investigated so far, except for recent studies by Steen & Brockhoff [5] and Riecke et al. [6][7].

So, how can we link vection research to computer-mediated applications? One of the unsolved challenges of many VR applications is to prevent users from getting lost in virtual environments; this happens much more often in virtual reality than in comparable real world situations. It seems reasonable to assume that good spatial orientation in VR critically depends on convincing ego-motion perception, which in turn requires effective ego-motion simulation. That is, any measure that increases the convincingness and intensity of visually induced ego-motion illusions without restricting the user unnecessarily would be beneficial. Furthermore, reducing the time needed until users begin experiencing ego-motion (“vection onset time”) when presented with a moving visual stimulus would be advantageous.

We know from the literature that fixating on a stationary object increases the ego-motion illusion [1], especially if the fixation point is perceived as being stationary and in the foreground of a moving (background) stimulus [2][8]. Conversely, stationary objects behind the moving stimulus decrease vection [9]. In those studies, however, observers were explicitly instructed to focus and fixate on those targets. This would be comparable to fixating on some stains on a train window or the windshield of a car. From an eco-
logical perspective, such fixation seems rather unnatural. This is especially true when one is the driver, where it might even be dangerous to pay attention to, for example, some dirt on the windshield instead of the road you want to follow.

1.1 Vection facilitation without fixation point

In this paper, we investigated whether a vection-facilitating effect might also occur under more natural conditions, i.e., under free viewing conditions and without any need to fixate on any nearby object. Furthermore, we attempted to facilitate vection in an unobtrusive manner, such that most participants would not even notice the manipulation.

To study this, the surface and reflection properties of the video projection screen that was used to present the vection stimulus were modified for one experimental group to include some subtle scratches and marks (see Figure 2). Even if participants did notice these marks, we believe that they would most likely oversee it as being an accident instead of an experimental manipulation. Furthermore, participants in our study were asked to view the stimulus in a normal and relaxed manner. They were further instructed to neither stare through the screen nor to fixate on any position on the screen (i.e., not to suppress the optokinetic reflex (OKR)).

1.2 Relation between vection and spatial presence

A recent study in VR demonstrated that vection can be closely related to spatial presence (i.e., the feeling of “being there”) and involvement in the simulated scene [6]. In that study, spatial presence was manipulated by gradually scrambling a photorealistic image of a natural scene. In the present study, however, there seems to be no obvious reason to expect that the additional marks on the projection screen should improve spatial presence and involvement. Intuitively, one might even expect a degradation of presence due to the decreased simulation fidelity. We believe, however, that it is possible that having the marks on the screen will be accompanied by higher spatial presence: As illustrated in Figure 7, the marks on the screen may enhance the sensation of ego-motion (just as a fixation point would), which in turn might lead to an increase of spatial presence and involvement.

1.3 Benefits

Any success in increasing vection and spatial presence through subtle modifications would be of considerable interest both for fundamental research and ego-motion simulation applications: To the best of our knowledge, facilitating vection and spatial presence in an unobtrusive manner and without restricting eye movements has never been shown in the literature. From an applied perspective, subtle vection-facilitating measures that allow for unrestricted eye movements could have important implications for the design of lean and elegant ego-motion simulators both for industry and consumer market.

2 Methods

2.1 Experimental design

Twenty-two naive participants were paid to participate in the study. All participants had stereo vision and normal or corrected-to-normal visual acuity. These 22 participants were randomly assigned to two groups: Twelve participants were presented with the unmarked screen, and ten participants were presented with the marked screen. For the latter condition, a different projection screen of identical size was used that contained hardly noticeable marks in the periphery of the projection screen (see Figure 2). Marks were located at the upper left part of the screen. Apart from the marks, the screens were identical in terms of material (Forex), size, and reflection properties. That is, any potential difference in results between the two screens should be attributed to the minor scratches on the screen. Each participant performed 16 trials, consisting of 4 repetitions of two different rotation velocities (20°/s vs. 40°/s, randomized) x 2 turning directions (left vs. right, alternating). This experimental design is summarized in Table 1.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Varied between participant</th>
<th>Varied within participant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Screen homogeneity</td>
<td>Unmarked vs. marked</td>
<td>20, 40</td>
</tr>
<tr>
<td>Turning direction</td>
<td>Left/right</td>
<td>N.a.</td>
</tr>
<tr>
<td>Repetitions</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 1: Experimental design

2.2 Stimuli and Apparatus

Participants were comfortably seated at a distance of 1.7m from a curved projection screen (1.68x1.42m, curvature radius: 2m) on which the rotating visual stimulus was displayed (Figure 3). The image was projected using a JVC D-ILA DLA-SX21S video projector with 1:1 optics. The projected image had a resolution of 1400x1050 pixels and was corrected for the curved screen geometry. Due to the D-ILA projector, the pixel rasterization was practically invisible. The visual stimulus consisted of a photorealistic view of the Tübingen market place that was generated by wrapping a 360° roundshot around a virtual cylinder (Figure 1). Circular vection was induced by rotating the stimulus around the vertical axis of the subject. Visibility of the surrounding room was prevented using black curtains. Furthermore, spatial auditory cues were masked by the sound of several layers of flowing water that was played through active noise-canceling headphones (Sony HMEC 300) that participants wore throughout the experiment. Responses were collected using a force-feedback joystick that was mounted in front of the participants at a comfortable distance.
Figure 1: Top: 360° roundshot of the Tübingen Market Place. Left: Roundshot model of the Tübingen market place, wrapped around a virtual cylinder. Middle: For the experiments, the simulated viewpoint was centered in the cylinder, yielding an undistorted 54°x40.5° view of the Tübingen market place. Right: Bird’s eye view of market place. The viewpoint is indicated by the cross.

Figure 2: Top left: View of the projection screen displaying the market scene. The marks are located at the upper-left part of the screen, as illustrated by the close-ups (top right and bottom). Bottom: Close-up of the same region as above (right), but illuminated with plain white light to illustrate the marks. Left: The original photograph demonstrating the unobtrusive nature of the marks (diagonal scratches). Right: Contrast-enhanced version of the same image to illustrate the marks.

Figure 3: Participant seated in front of curved projection screen displaying a view of the Tübingen market place. The simulated FOV matched the physical FOV of 54°x40.5°.
2.3 Procedure

Experimental trials were initiated by participants’ pressing a button on the joystick, upon which the static image started rotating clockwise or counterclockwise around the vertical axis with constant acceleration for 3s. Maximum rotational velocities were 20°/s and 40°/s. The maximum duration of constant velocity rotation was 60s after which the stimulus decelerated at a constant rate for 6s. Participants were instructed to pull the joystick in the direction of their perceived self-motion as soon as it was sensed.

Vection was quantified in terms of five dependent variables: The time interval between the onset of stimulus rotation and the first deflection of the joystick indicated the **vection onset time** and was the primary dependent measure. Participants were also asked to pull the joystick more the stronger the perceived self-motion was; this allowed us to record the time course of **vection intensity** (joystick deflection). This continuous recording allowed us to collect two more dependent measures: The time when 50% of the maximum joystick deflection (vection intensity) was reached (named **50% vection onset time**), and the **time between vection onset and maximum vection** reported by the participant in each trial.

The rotation of the stimulus stopped automatically if maximum joystick deflection was sustained for 10s (otherwise it continued for 60s) to reduce the potential occurrence of motion sickness. Finally, at the end of each trial participants were asked to provide a “**convincingness**” rating of perceived self-motion by moving a lever next to the joystick to select one of the 11 possible values of a 0-100% rating scale. The value of 0 corresponded to “no perceived motion at all” (i.e., perception of a rotating stimulus and a stationary self) and that of 100 to “very convincing sense of vection” (i.e., perception of a stationary stimulus and a rotating self).

Between trials, there was a pause of 20 seconds to reduce potential motion aftereffects. In order to familiarize participants with the setup, a practice block containing 4 trials preceded the 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 ego-motion.

Participants were instructed to watch the stimuli “as relaxed and naturally” as possible. They were also told to neither stare through the screen nor to fixate on any position on the screen (in order not to suppress the optokinetic reflex (OKR)). Instead, they were instructed to concentrate on the image in the central part of the projection screen. The marks on the screen were not mentioned to them until after the experiment. In fact, only one of the participants reported having noticed the marks in a post-experiment interview.

We did not use any fixation point, even though it is known that a fixation point reduces vection onset times [1]. The main reason was that from an applied perspective for ego-motion simulation, we were interested in investigating how one can induce convincing ego-motion sensation under natural viewing conditions. Moreover, not fixating also reduced the perceived flicker and ghost images due to the 60 Hz projection: Fast motions like rotations above 60°/s produce strong flicker and ghost images if the eyes fixate one point and do not follow the image motion. For example, a single vertical line translating sideways is seen as multiple flickering lines as it moves across a fixation point.

3 Results

3.1 Vection measures

The behavioral data for the five dependent variables are summarized in Figure 4. A first glance at Figure 4 reveals a considerable influence of the marks on the screen for all dependent measures: Vection onset times and 50% vection onset times were both about 2-3 times smaller with the marked screen. The time between vection onset and maximum vection reached was also considerably decreased due to the additional marks on the screen. Furthermore, both convincingness ratings and vection intensity were clearly increased.

To quantify these effects, a set of repeated-measures ANOVAs was computed for the five dependent variables using a 2 (rotation velocity: 20° vs. 40°/s, within-subject) x 2 (screen type: unmarked vs. marked, between-subject) factorial design. The results of the ANOVAs are summarized in Table 2. The vection-facilitating effect of the additional marks on the screen proved highly significant for all dependent measures. Likewise, increasing the stimulus velocity from 20°/s to 40°/s resulted in a highly significant increase in vection (reduced onset times and increased intensity and convincingness ratings). The effects of the marks on the screen were more pronounced for the slower rotations, as indicated by significant interactions (see Table 2).

3.2 Presence Questionnaires

At the end of the experiment, each participant rated the presence in the simulated scene using the 14-item Igroup presence questionnaire (IPQ) by Schubert, Friedmann, & Regenbrecht [10]. The questionnaire results are summarized in Figure 5. In our sample, the IPQ showed high reliability (Cronbach’s $\alpha = .939$).

Furthermore, participants completed a simulator sickness questionnaire (SSQ) before and after each session [11]. As expected, the simulator sickness ratings were somewhat higher after the experiment, but all participants felt comfortable finishing the experiment.

An additional presence susceptibility questionnaire (unpublished), which is supposed to measure a person’s general susceptibility to presence, did not show any clear results or correlations with any of the vection measures. In the following, only the results form the IPQ presence questionnaire will be discussed.
Figure 4: Plotted are the means of the five dependent variables for the four experimental conditions. The left and right pair of bars in each plot represent the low and high velocity conditions (20°/s and 40°/s, respectively). Data for the unmarked and marked screen are represented by the darker left and lighter right bar of each pair of bars, respectively. Boxes and whiskers depict one standard error of the mean and one standard deviation, respectively. Note the considerable vection-facilitating effect of the additional marks on the screen.

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Marked vs. Unmarked screen</th>
<th>Stimulus velocity</th>
<th>Interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vection onset time</td>
<td>F(1,20) 14.45 0.001**</td>
<td>F(1,20) 38.07 0.000***</td>
<td>F(1,20) 6.01 0.024*</td>
</tr>
<tr>
<td>50% vection onset time</td>
<td>F(1,20) 22.05 0.000***</td>
<td>F(1,20) 60.12 0.000***</td>
<td>F(1,20) 4.62 0.044*</td>
</tr>
<tr>
<td>Time between vection onset and maximum vection</td>
<td>F(1,20) 5.15 0.034*</td>
<td>F(1,20) 14.40 0.001***</td>
<td>F(1,20) 0.001 0.972</td>
</tr>
<tr>
<td>Vection intensity</td>
<td>F(1,20) 20.41 0.000***</td>
<td>F(1,20) 51.19 0.000***</td>
<td>F(1,20) 5.75 0.026*</td>
</tr>
<tr>
<td>Convincingness</td>
<td>F(1,20) 9.95 0.005**</td>
<td>F(1,20) 44.24 0.000***</td>
<td>F(1,20) 0.007 0.933</td>
</tr>
</tbody>
</table>

Table 2: ANOVA results for the two factors rotation velocity (20°/s vs. 40°/s, within-subject) and screen type (unmarked vs. marked, between-subject). Both factors showed significant main effects for all dependent measures.

Individual aspects of presence were investigated by analyzing the sum score (mean over all 14 items) as well as the individual subscales. Those consisted of an involvement/attention subscale (4 items), and the subscales “being there” (1 item), space (5 items), and realism (4 items). Motivated by the results of a factor analysis by Riecke et al. [6], the latter three subscales were combined to form a spatial presence compound scale (10 items) (see Figure 5).

3.3 Correlations between presence ratings and vection measures

To investigate the relation between the subjective presence ratings and the vection measures, a set of paired-samples correlation analyses were performed between the three main vection measures (vection onset time, convincingness, and vection intensity) and the presence scores. Table 3 summarizes the paired-samples correlations (r) and the corresponding p-values. The convincingness ratings showed significant positive correlations with the overall presence score as well as all subscales. Vection onset time, however, correlated significantly only with the involvement/attention subscale, and marginally with the presence sum score (p=0.078). The negative sign of the correlation means that higher attention and involvement in the simulated scene was associated with shorter vection onset times.
Neither spatial presence nor any of its subscales showed any clear correlation to the vection onset time. Vection intensity was only marginally positively correlated to the presence ratings ($0.05 \leq p \leq 0.1$), with a tendency towards higher correlations for the intensity/attention subscale than for the spatial presence-related items.

### 4 Discussion & Conclusions

#### 4.1 Vection-facilitating effect of marks on the screen

The comparison between the marked and unmarked screen showed a clear vection-facilitating effect for the marks in the periphery of the projection screen. The magnitude of the effect is rather striking and comparable to results obtained by a fixation point [1]. How can the effect found in this study be explained?

It has been known that both static fixation points and static foreground stimuli can facilitate vection [1][2]. This has been explained by an increased relative motion on the retina. The novel finding from our study is that this effect occurs even if the stationary objects (or marks) are hardly noticeable - only one participant reported having noticed the marks when participants were asked whether they noticed anything special about the experimental setup. In addition, observers in our study were instructed to view the stimulus in a normal and relaxed manner, without staring through the screen or fixating on any static point. Even though we did not record eye movements, we can infer that participants did in fact not fixate on the screen or stare through it; if they would have done so, they would have perceived multiple images due to the digital projection, which they did not. Hence, our result cannot be simply explained by an increase in the relative motion on the retina.

Nakamura and Shimojo [2] showed that vection can be facilitated if the moving visual stimulus is being perceived as background motion, that is, as being behind a static foreground object that participants fixated on. In the present study, however, there was no physical foreground-background or depth-separation between the static marks on the screen and the moving scene presented on the same screen. Furthermore, participants did not focus on any static object. Instead, they typically followed the moving stimulus via smooth pursuit. Hence, the foreground-background explanation cannot account for the vection-facilitating effect of subtle marks on the screen as it does not apply to this study.

So far, we can only speculate about the underlying processes. We propose that the hardly noticeable marks might provide some kind of subtle stable reference frame with respect to which the moving stimulus is being perceived. Even though there was no physical depth separation whatsoever between the marks on the screen and the visual motion stimulus presented on the same screen, participants might somehow have attributed the marks to the foreground, much like stains on a cockpit window, and the projected stimuli as moving with respect to that cockpit.

A study by Lowther and Ware [12] reported similar vection facilitation due to a stable foreground stimulus. In a vection study in VR, they overlaid a 5x5 grid on a large flat projection screen which was used to present the moving stimuli. The additional grid reduced vection onset times by almost 50%. Note that Lowther and Ware’s grid extended over the whole screen and was clearly visible, which is a major difference to the marks used in the current study: They were barely noticeable (only one participant was able to report them) and covered only a small portion of the peripheral FOV. Nevertheless, these subtle marks facilitated vection consistently in all dependent variables, and the effect size was even stronger than for the clearly visible grid by Lowther and Ware [12].

Further studies are needed to corroborate the proposed explanation that the marks on the screen might provide some kind of subtle foreground reference frame that influences self-motion perception. If that was true, it would have important implications for the design of convincing egomotion simulators, especially if participants would not have to be aware of the manipulation. However, further studies are needed to better understand this phenomenon and corroborate the main findings.

#### 4.2 Velocity effect on vection

The vection literature reports typically no systematic influence of stimulus velocity on vection onset time. For example, Brandt and colleagues [13] reported no systematic effect of stimulus velocity on vection onset time. Kennedy et al. [11] reported no overall velocity effect, only a small increase in vection onset time for the slowest velocities used (20°/s). Those studies used optokinetic drums with full-field stimulation.

The present study, however, demonstrates a clear vection-facilitating effect of increasing the velocity. This is interesting, as doubling the velocity also doubled the acceleration, yielding a larger visuo-vestibular conflict, which is typically assumed to decrease vection. Many factors might have contributed to the different findings, including the photorealistic scene used, the relatively small field of view, and the usage of VR technology. One could for example argue that for the higher stimulus velocities, attention was
drawn more towards the edges of the display. As a stable foreground stimulus as well as occluding foreground edges are known to enhance vection [2][3][9][12][14], this might explain the observed velocity effect. This would also explain why typically no velocity effect was found for full-field stimulation in optokinetic drums that do not have any screen boundaries.

4.3 Relation between vection and presence

The observed pattern of correlations between presence and vection measures points towards an interesting asymmetry: While the online measures of vection onset time (and to some degree also vection intensity) were more closely related to the involvement/attention aspect of overall presence, the subjective convincingness ratings which followed each trial were more tightly related to the spatial presence-related aspects of overall presence. This differential interrelation suggests a two-dimensional structure of presence with respect to ego-motion perception: Attentional aspects or involvement (e.g. awareness of real surroundings of the simulator vs. the simulated environment) on the one hand and spatial presence on the other hand.

In a recent vection study in a comparable VR setup, we attempted to modulate presence in the simulated scene directly by reducing the scene consistency gradually through scene scrambling [6]. This manipulation also decreased vection considerably, and a clear correlation between the degradation of presence and vection was found. One could hypothesize that presence (or other top-down factors) in that study actually mediated vection, in the sense of high presence improving vection (see Figure 6).

The observed pattern of correlations between presence and vection measures points towards an interesting asymmetry: While the online measures of vection onset time (and to some degree also vection intensity) were more closely related to the involvement/attention aspect of overall presence, the subjective convincingness ratings which followed each trial were more tightly related to the spatial presence-related aspects of overall presence. This differential interrelation suggests a two-dimensional structure of presence with respect to ego-motion perception: Attentional aspects or involvement (e.g. awareness of real surroundings of the simulator vs. the simulated environment) on the one hand and spatial presence on the other hand.

In a recent vection study in a comparable VR setup, we attempted to modulate presence in the simulated scene directly by reducing the scene consistency gradually through scene scrambling [6]. This manipulation also decreased vection considerably, and a clear correlation between the degradation of presence and vection was found. One could hypothesize that presence (or other top-down factors) in that study actually mediated vection, in the sense of high presence improving vection (see Figure 6).

This might be understood in the context of the “presence rest frame hypothesis” proposed by Prothero and colleagues [15]. According to this hypothesis, “the sense of presence in an environment reflects the degree to which that environment influences the selected rest frame”. The selected rest frame is defined as the chosen subjective coordinate system with respect to which positions, orientation, and hence also motions are judged. Thus, motion of what is considered to be the selected rest frame may result in illusory self-motions (vection). According to this hypothesis, the consistent scene in [6], which induced higher presence ratings than any of the scrambled (inconsistent) stimuli, might have been more readily accepted as the selected rest frame. Consequently, the effect on the self-motion illusion should be stronger, which was confirmed by the experiment by Riecke et al. [6].

Figure 6: The study by Riecke et al. [6] suggests that top-down factors like spatial presence and involvement can mediate the visually-induced ego-motion illusion.

In the present study, there seems to be no theoretical reason to expect that spatial presence and involvement should be directly enhanced by the additional marks on the projection screen. Rather, one might expect a degradation due to the decreased simulation fidelity. Nevertheless, the additional marks on the screen did considerably increase spatial presence and involvement. It seems possible that the increase in spatial presence was mediated or caused by the increase in the ego-motion illusion (see Figure 7).

This sheds a novel light onto the nature of spatial presence and involvement and its relation to ego-motion perception: The study by Riecke et al. [6] suggests that spatial presence and involvement can mediate the visually-induced ego-motion illusion (see Figure 6), whereas the current results suggest that the opposite might also occur: That spatial presence and involvement can themselves be mediated by the visually-induced ego-motion illusion, as is sketched in Figure 7.

4.4 Outlook and applications

Compared to the traditional approach in circular vection studies that uses rotating optokinetic drums, the present study demonstrates that vection can indeed be reliably induced and studied using a virtual reality setup. The relatively large vection onset times without the additional marks on the screen can probably be explained by the rather small field of view used in this study, compared to full-field stimulation in optokinetic drums. Using digital display technology for presenting moving stimuli has the drawback of presenting individual images with a limited update rate (here with 60Hz) instead of a truly continuous motion like with optokinetic drums. This might impose problems and artifacts like flicker and ghost images if participants would focus on any stationary fixation point. As this study demonstrates, however, convincing ego-motion illusions with quick vection onset can indeed be reliably induced in a VR simulator in a non-obtrusive way, without any explicit fixation and under natural, relaxed viewing conditions.

This finding yields important implications for motion simulator design, because from an applied perspective, one wants to achieve realistic ego-motion simulation without...
restricting eye movements. Further research will investigate how such subtle, unobtrusivevection-facilitating measures can best be included in a consistent and ecologically valid motion metaphor. One application could be to design a motion simulator with stains and dirt on the (physical or simulated) windshield. As this study strongly suggests, such a simple and subtle modification would increase the strength and convincingness of the self-motion illusion as well as presence and involvement in the simulation without imposing unnatural constraints on users’ behavior.

Prothero [15] proposed in the context of his presence rest frame hypothesis that vection and presence in the simulation should be tightly intertwined, if not causally linked. This was supported by the current study and other recent vection experiments by Riecke et al. [6][7]. This suggests that the joint measurement of vection and presence is a promising approach towards a deeper understanding of the underlying factors for both vection and presence. Further studies, however, are required to provide a conclusive answer about the exact relation between the two phenomena.

5 Acknowledgements

This research was funded by the European Community (IST-2001-39223, FET Proactive Initiative, project “PO-EMS” (Perceptually Oriented Ego-Motion Simulation, see http://www.poems-project.info)) and the Max Planck Society. The authors would like to thank Jan-Oliver Hirn for his valuable help during data collection and data analysis.

6 References

Virtual Ba and Presence in Facilitating Learning from Technology Mediated Organizational Information Flows

Timo Saari, Jari Laarni, Niklas Ravaja, Kari Kallinen, Marko Turpeinen
M.I.N.D. Lab/CKIR/HSE & HIIT, CKIR/Helsinki School of Economics, Helsinki Institute for Information Technology
{saari@hkkk.fi, laarni@hkkk.fi, ravaja@hkkk.fi, kallinen@hkkk.fi, marko.turpeinen@hiit.fi}

Abstract
In this paper we present an overview to facilitating learning from technology mediated organizational information flow via high presence states that may lead to better engagement, quality of communication and processing of information. First, the concept of presence is introduced and some performance effects of high presence states are discussed. Second, key concepts of knowledge and a knowledge-based view of organizations are presented. Then a processual model of knowledge creation and sharing inside organization, the SECI-model, is discussed. Next, Virtual Ba, a mental state of sharing, discussing and communicating via communication technology inside organizations is introduced as a link to presence research. A technique called Psychological Customization is briefly presented to support adaptive information presentation and personalization to users of communication technology in organizations in order to facilitate learning from perceived information via high presence states. Examples of using Psychological Customization as part of Virtual Ba to create high presence are presented. Finally, some implications for future system designs and research are discussed.

1. Introduction
The use of technology and tools that support learning (i.e. knowledge construction and creation) from technology-mediated organizational knowledge flows is not a well-developed area of study from the point of view of understanding the experience and state of mind of users of such technology.

Little work has also been done to utilize automatic adaptation and personalization of information in work-related group communication. The use of dialogical tools in team communication is further poorly understood from the point of view of user experience. This paper will explore the design space of personalization of technology-mediated group communication flows and the use of dialogical interaction tools that are based on understanding the state of mind of the users when involved in technology-mediated knowledge construction and knowledge creation.

When perceiving information via media and communications technologies users may have a feeling of presence, the perceptual illusion of nonmediation [16]. There have been several attempts to define the presence concept [e.g. 8, 9, 16, 32, 36]. There are some aspects that are common to most of these definitions: Presence is a multi-dimensional phenomenon that is related to ‘being there’ in one environment and to the ‘perceptual illusion of nonmediation’ [9]. The term presence is used here to mean both virtual presence and telepresence. Virtual presence means that the person feels present in a computer-mediated world; telepresence, in turn, means that he/she feels present in a remote but real environment. Both of these terms refer to experiences that involve displacement of the person’s self-perception into a computer-mediated world.

The state of presence apparently has useful effects. It has been argued that presence is necessary for effective performance in a computer-mediated world: With an increased feeling of presence the user is attending more intensively to the task at hand, and thus his/her performance is improved [3]. There is, for example, some evidence that presence improves efficiency or reduces workload [16].

When using collaborative technology for computer-mediated social interaction, the users experience a state called social presence during which users may, for instance, experience intimacy of interaction or a feeling of togetherness in virtual space [16]. During social presence users also experience various other types of emotional and cognitive effects, such as interpersonal emotion, emotion based on being successful in the task at hand, and learning from shared activities or shared information. It is likely that, when receiving and reading messages via technology, members of a working team may experience presence in response to the message content and social presence with the sender/s of the message. During and after the feeling of presence, users may then experience emotions and moods based on the message as well as learn and initiate problem solving processes and creation of their own content, i.e. sending messages back to the system to be passed on to various recipients. It may be that high presence states in organizational technology users may facilitate better performance and learning from organizational information flows.
This article explores these issues and attempts to provide a preliminary approach to understand the use of communication technologies in knowledge-based organizations from the point of view of the potentially beneficial performance effects of presence or social presence when using such technologies.

2. The Knowledge Based Organization

2.1. The Nature of Knowledge

Learning from information delivered through technology inside organizations is related to the knowledge-based view of the organization. This view entails for instance that shared knowledge inside an organization is key to innovation and success in the marketplace [19].

First, the conception of knowledge is discussed. Nonaka and Takeuchi [17] have introduced their own definition of knowledge and especially the differences of information and knowledge. First, they claim that knowledge is about beliefs and commitment: a function of a particular stance, perspective or intention. Second, knowledge deals with action, since it is always constructed to serve some end. Third, knowledge contains meaning; it is context-specific and relational. They claim that information can be seen as either syntactic (volume) or semantic (meaning) information. Syntactic information would mean for example just the amount of information, regardless of meaning, sent through a communications channel. Semantic information focuses on the mediated meaning with that information and is thus of more interest when discussing knowledge production. [17].

In other words: information is a flow of messages, while knowledge is created from that very flow of information via active interpretation of the perceiver, anchored in the beliefs and commitment of the perceiver. This understanding emphasizes that knowledge is essentially related to human action and interaction.

Nonaka and Takeuchi [17] have further argued for the existence of a special kind of knowledge: tacit knowledge, based on Polanyi [20], which "carries" the creative acts and creative insights of the individual. The conversion of this tacit knowledge into the sphere of linguistic, distributable knowledge (and information) can be done in an organization in various ways. In this sense the sharing of the results of individual creativity is one of the key processes of an organization. However, it is difficult to bring out tacit knowledge and turn it into communicable information easily to be learnt by others.

In this context Nonaka and Takeuchi [17] discuss the distinction between tacit knowledge and explicit knowledge, drawing from Polanyi, [20]: Tacit knowledge is personal, context-specific, and therefore hard to formalize and communicate. It is literally "hidden" or tacit. Explicit or "codified" knowledge, on the other hand, refers to knowledge that is transmittable in formal, systematic language." [17]

Knowledge, just like information, is context-specific and relational, because it depends on a certain situation and setting and is created dynamically in social interaction among people [17]. Thus there is a social aspect to knowledge construction and learning.

Berger and Luckmann [1] emphasize the social aspect of knowledge and state that people in a certain historical and social context share information from which social knowledge is constructed. The social knowledge is then the basis of their common "social reality" or "everyday reality" and it influences their attitudes, judgment and behavior. [1]

Nonaka and Takeuchi (1995) then see that organization is based on socialization, communication and dialogue, not on the efficient distribution of information. At the heart of this approach there has been an emphasis on the special nature of individual knowledge vs. organizational knowledge. The special nature of the knowledge of the individual comes from the fact that it is the individual who creates and that this creation and experience can be shared with others in the organization, i.e. it can be transformed into a force that will propel the organization forward towards innovation, for instance.

Hence, in organizational studies and the knowledge-based view of the organization knowledge has been seen as i) individual, ii) social and iii) tacit and explicit.

This has implications for use of this view in communication studies also. First, the individual focus emphasizes that the individual is the core of knowledge generation or roughly speaking; learning and generation of new ideas. Second, the social aspect emphasizes team and group-level communication and its effectiveness, as the quality of social knowledge may also be dependent upon to the technologies used in social interaction and communication. Third, the most important aspect may be that with technology one is able to transmit “only” information, i.e. explicit knowledge, not tacit, hidden knowledge. Hence, it may be sensible to look for ways to make information more “tacit” via different techniques to enhance organizational performance.

These points may have important consequences for presence research as it may be that for instance intensive presence or social presence would be a more optimal “state of mind” to be able to communicate also the “tacit” level of knowledge, rather than only information. Naturally, a high presence state when perceiving technology mediated information, may also be beneficial to learning and performance.

2.2. The SECI-process and Ba

In the context of seeing interactions of individuals and groups of people in corporations as key parts of platforms for knowledge creation and innovation the influential concept of Ba has been introduced [e.g. 18,19]. In the knowledge-based view of the firm key processes of innovation are seen as linked to the SECI-process. This implies a complex mechanism for creating and sharing knowledge inside the organization, focusing on the interaction between the explicit and tacit level of knowledge, called “knowledge conversion”. The underlying assumption is that as tacit knowledge is valuable to the success of the organization, ways of creating, sharing and communicating it are important to understand.
The SECI-process stands for Socialization (from tacit knowledge to tacit knowledge), Externalization (from tacit knowledge to explicit knowledge), Combination (from explicit knowledge to explicit knowledge) and Internalization (from explicit knowledge to tacit knowledge) of Knowledge. [18]. This is illustrated in Figure 1.

Figure 1. The SECI-process. Adapted from Nonaka, Toyama and Konno [18].

Socialization is a process of converting new tacit knowledge through shared experiences. Traditionally this takes place in apprenticeship where apprentices learn tacit knowledge needed through hands-on-experiences rather than written manuals. Here two or more people work in a close collaboration and “tacit” knowledge of the other, for instance a master, is indirectly learned via practice. Externalization is a process of articulating tacit knowledge into explicit knowledge. Concept creation in a new product development is a good example of this. This means formalizing new, creative ideas into communicable information. Combination is a process of converting explicit knowledge into more complex and systematic sets of explicit knowledge. For instance, explicit knowledge, i.e. communicable information, is collected from inside or outside the organization and is then edited and distributed throughout the organization. This distribution may take place in interpersonal or group meetings or via technology. Internalization is a process of embodying explicit knowledge into tacit knowledge. This allows the creation and sharing of information or explicit knowledge throughout the organization. Examples are product concepts or manufacturing procedures that have to be actualized through action and practice. Such knowledge can be obtained by reading manuals or documents and reflecting upon them to understand the procedures and actions needed, i.e. one can create one’s own tacit knowledge from explicit knowledge. [18].

The SECI-process involves an interaction between all the parts. This interaction has been described as an opening spiral, not a circle or loop. The spiral signifies the creation of new knowledge and ideas through the model.

In the SECI mechanism Ba is used to refer to a sophisticated shared mental, physical and social context of knowledge creation in which new knowledge is created, shared and utilized in dialogue with other people. Ba may be thought of as a physical space, like a meeting room, but it may also be seen anywhere where individuals and groups are interacting to create knowledge and share meanings, such as in email groups or other technology mediated communication, presented as virtual ba. [18, 19]

Ba is an influential concept because it provides a holistic view on knowledge-related activities of a corporation.

Ba is in essence a shared context between two or more individuals. It can be a mental context in the form or shared memories, mental models or perhaps cultural values and norms. It can be also physical, a shared space or virtual as in the case of using communication technology to communicate. The essence of Ba is that this shared, somewhat undefined, yet rich concept, enables in-depth interaction, understanding and generation of “the new” in the form of ideas between two or more people. [see 19]

There are four types of virtual Ba as seen in Figure 2.

Figure 2. Four types of ba. Adapted from Nonaka, Toyama and Konno [19].

*Originating ba* is the sphere of individual and face-to-face interactions. It offers a context for socialization, as it is thought that human-human interaction is the only way to provide the “full range” of bandwidth of communication between people. It is also the place for transcending self in dialogue and interaction with others, a sort of existential state. *Dialoguing ba* is characterized by collective and face-to-face interactions. It is a way of sharing individual mental models and skills that are articulated as concepts. This then mainly offers a context for externalization of information between people. *Systemizing ba* is reflected in collective and virtual interactions through communication technology. This mainly offers a context for the combination of existing explicit knowledge that can be easily communicated and transmitted via technological means via online networks, groupware and websites. *Excercizing ba* is defined by virtual and individual interactions. This is the area of internalization. Individuals can hence embody explicit knowledge. 

PRESENCE 2004

135
knowledge that is received via technology and communication media, such as written manuals or simulations programs. [19]

The concept of virtual ba, often described as complex and somewhat undefined, has not been sufficiently elaborated from the psychological point of view of the experience of users when using such a system. It is evident that when engaged in more synchronous or asynchronous technology-mediated interaction such as email, videoconferencing, collaborative tools or browsing the corporate intranet, the technology, the user interface and the message content sent/read by users becomes the “interface” to Ba. As high presence may be related to high performance with computer aided tasks, such as communication and learning, presence may be seen in this context as a key component of virtual ba.

3. Systems to Support Virtual Ba

3.1. Psychological Customization

Hence, virtual ba is partly a technology-mediated phenomenon even though it is closely linked to personal and team-oriented processes in knowledge construction and creation. The conditions of what it means for Ba-type of concept to be technology-mediated lead to issues such as what is the influence of synchronicity in collaboration, or how does the level of dialogicalness and locus of power of a collaborative tool relate to knowledge creation and construction. One further issue is the use of particular types of user interfaces, modalities of information and particular technologies to support certain knowledge creation and construction processes within teams and corporations. Briefly, the conditions of perception of information produced by others via virtual ba as well as the proactive tools available for dialogue and production of user’s own messages and content into the virtual ba are at the core of the domain of interest when examining the relationship of ba and technology.

If virtual ba can be divided into its technological and user-centered “experiential” elements and their interactions can be explained and predicted, one may offer some design guidelines or other aid into concrete applications of technology to be better facilitating knowledge creation. However, clearly there is a possible gap between philosophical concepts such as ba and real-life case studies of users of information systems. Despite this, an attempt is made in this paper to shed light on this issue.

One starting point may be that presence and social presence occurring in team work processes related to knowledge building, facilitation and sharing may also be thought of as related to ba, a mental state related to intensive common problem solving and knowledge creation.

A key technology to support virtual ba may be individual- and group-centered personalization. Personalization and customization entails the automatic or semi-automatic adaptation of information per user in an intelligent way with information technology [e.g. 26, 33].

One may also vary the form of information (modality for instance) per user or community profile, which may systematically produce, amplify, or shade different psychological effects, such as presence, social presence, emotion, mood, learning and persuasion [27, 28, 29, 30]. This means that the actual content or substance of the message is not altered, merely the way of presenting the message and its various sub-factors.

One may speak of systems that facilitate the emergence of desired user experiences, such as presence or social presence, when dealing with knowledge construction and knowledge creation oriented teamwork. In this manner, one may speak of technologies that efficiently create a Virtual Ba. One approach to describing such systems is Mind-Based Technologies [see 27].

Briefly said Mind-Based Technologies facilitate desired psychological states related to information processing. The key idea is to match significant individual and group-related differences in perceptual processing to facilitate the emergence of for instance efficient learning or intensive emotions or in-depth presence. One operationalization of the idea of Mind-Based Technologies is Psychological Customization. [see 27, 29, 30, 31, 34]

Psychological Customization includes modeling of individuals, groups, and communities to create psychological profiles and other profiles based on which customization may be conducted. In addition, a database of design rules is needed to define the desired cognitive and emotional effects for different types of profiles. Once these components are in place, content management technologies can be extended to cover variations of form and substance of information based on psychological profiles and design rules to create the desired psychological effects. [31, 34]

Even though no actual system has been implemented yet for Psychological Customization related to facilitating learning from technology-mediated organizational information flows, available indirect empirical evidence supports the feasibility and validity of the idea that varying the form or design of information may be efficient in producing various types of psychological effects:

- Individual differences have a considerable effect on computer-based performance [e.g. 5]. For example, individual differences in memory capacity have an effect on people’s behavior in many types of activities [35].
- Varying the form of information creates various emotional and cognitive effects [e.g. 12, 13, 15].
- Different modalities, such as visual and auditory, may lead to different kinds of psychological influences and the valence of a preceding subliminal stimulus influences the subsequent evaluation of a person evaluated [4, 11].
- Different ways of processing information influence learning and emotion of stimuli with certain modality [25].
- Subliminal exposure to happy affective primes in connection with video messages presented on a small screen has several putatively positive influences (i.e., increased pleasure, perceived message trustworthiness, and memory) [23].
- Media messages can be modified in terms of audio characteristics to induce attention, emotion and presence [10, 22].
• Presence of image motion to meet the personality (as defined in terms of dispositional behavioral activation system sensitivity) of the user may enhance his or her attentional engagement, information processing, and enjoyment [21].

Accordingly, Saari [27] has grouped the clusters of information form or design related variables relevant to emerging transient psychological effects of processing mediated information as: i) hardware layer (size, proximity, fixed place/carryed by user), ii) code layer (way of interaction and degree of user control, ways of presenting visual-functional controls in the user interface) and iii) content layer (substance: essence of the event described in the message, form: modalities, visual layouts and temporal structures).

3.2. Examples of Virtual Ba

Prior to gathering empiric evidence of the phenomenon, this paper will present some examples to create intensive virtual ba informed by psychological theory and indirect empiric evidence concentrating on the influence of the way of presenting information (type of end-user device, ways and modalities of interaction, type of user interface, information modality, temporal narrative sequences and structures).

The key idea is that providing an intensive and in-depth virtual ba may facilitate actual knowledge construction and creation processes and communication flows through the system. It is also thought that personalizing the form of information received and sent through such a system may help in creating the desired state of mind in the users, perhaps aiding in knowledge creation and construction processes.

First, the example of Excercising ba is discussed in Table 1. The focus is mainly on how an individual perceives information via a technological device, such as reading information from the corporate intranet.

Based on Table 1 it may be said that a variety of adaptations of the form of information are possible based on Psychological Customization. First, in creating cognitive effects, the system may automatically manipulate the form of the user interface, interaction modality or message received or sent in a manner that is optimal for the receiver based on the rule-database. For instance, typography and screen layout may be optimised for each receiver. Also, when possible, automatic translations from text to audio or vice versa may be sensible depending on environmental conditions. Second, one may create various attentional effects with for instance background music that may facilitate information processing. Third, one may use emotional responses and mood to facilitate information processing or make the use of information more pleasant. Fourth, one may vary the design of information in a way that may produce high presence states via engaging attention or bandwidth of information processing, for instance.

Next, the use of Psychological Customization for Systemizing ba are discussed. The dimension of systemizing mainly involves the use of socially oriented information technology aids, such as email, chat-groups, blogs, or dialogical groupware. In this case it is a bit more difficult to explain possible uses of the system. This is due to the fact that there are at least two discussants using a collaborative technology, rather than “only” one person processing ready-made information as in the case of Excercising Ba.

<table>
<thead>
<tr>
<th>Layer of Technology</th>
<th>Adaptations for Excercising Ba</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Physical -multimedia PC or mobile device</td>
<td>-Mobile device: user changeable covers in colors and shapes that facilitate emotion -Pc:s designs that make the information more visually ergonomical to perceive (backlit screens etc.)</td>
</tr>
<tr>
<td>2. Code -Windows-type user interface -Mouse, pen, speech,</td>
<td>-The user interface elements (background color, forms, shapes, directions of navigation buttons etc.) may be varied in real-time per page per user in which a certain information is located to create various emotions and ease of perceptual processing -audio channel may be used to create emotional and attentional effects (using audio input/output sound, varying pitch, tone, background music, audio effects etc.).</td>
</tr>
<tr>
<td>3. Content A. Substance - Fixed multimedia content</td>
<td>-Adding subliminal extra content to create emotion and mood towards the content, such as happy faces for priming effects</td>
</tr>
<tr>
<td>B. Form Modality -Multimedia</td>
<td>-Modality may be matched to cognitive style or pre-existing mood of the enable easier processing. -Background music, audio effects may be used as a separate modality to facilitate desired emotions and moods and ease of information processing -Animated text can be used to create more efficient processing of text facilitate some emotional and attentional effects.</td>
</tr>
<tr>
<td>Visual presentation</td>
<td>-Emotionally and cognitively evaluated and positioned layout designs and templates for information (colors, shapes and textures) may be utilized per type of user segment</td>
</tr>
<tr>
<td>Structure -temporal, other</td>
<td>-Offering perceptually and cognitively efficient structures for information presented</td>
</tr>
</tbody>
</table>

Table 1. Technological possibilities for Psychological Customization in excercising ba.

In order to a Psychological Customization system to function in Excercizing ba-type of settings, the users would need to create a user profile (personality, cognitive style, other relevant information) for the system to gain access to form factors that may create desired attentional, cognitive or emotional effects in the users. The users would also fill out a community profile that indicates which users have authority to send psychologically intelligent messages to them and vice versa. The system would need a database of design rules of probable psychological effects of each type of manipulation per type of user and some other functionality. Further, if the system could record the user’s
psychological state, this may make the system more reliable by making it possible for the system to more objectively verify the psychological states of the user. This may happen via the use of psychophysiological recordings, if feasible.

When crafting a message to another user, the system may automatically suggest for the sender of message a possibility to psychologically customize the message for a particular receiver. The user would select a desired effect, such as creating positive emotion in the receiver with a message in which the substance is written in text and the system would present the sender with ready-made and psychologically evaluated templates (consisting of graphics, animation, sounds, videos etc.) that with high probability may create the desired emotion for the receiver with a particular user profile. The sender would type in the text-message, record an audio message or shoot a video in the template, finalize the design and then send the message. The receiver would receive an emotionally optimised message and may then experience the desired emotion. Naturally, if the substance of the message and the form of the message communicate a different emotion, for instance the substance is hostile and the template is joyful, some effects may not be realized. Similarly, one may use the manipulation of form of information with other types of effects, such as high attention, presence and efficiency of information processing.

There are two issues to be dealt with here. First, in the case of real-time communication between two parties it may be difficult and uneconomical to manually edit the flow of messages as this creates extra work. Hence, an automatic system doing the adaptations based on user profiles and the setting of desired effects per person are a better alternative. Second, in the case of personal, non-social internet surfing or information perception via technology, one may also set a desired effect himself to “efficient processing” for instance and the system would try to adapt information accordingly.

It should be noted that there are technical complexities involved with the automatic real-time adaptation and profiling of information for psychological effects and they will not be dealt within this article [for a better technical description, see 31, 34]. In any case the possibilities of Psychological Customization in the examples discussed are very similar to the possibilities presented in Table 1 and hence are not repeated here.

4. Conclusion

The relevance of the presented framework to organizational knowledge-oriented research and communication research is evident as it may provide an approach to facilitate desired psychological states in individual users involved in knowledge-based teamwork.

Indirectly it may be said that by manipulating the conditions of technology-mediated perception and social interaction one may be able to create more “tacitness” into explicit information, hence easing the transfer of tacit knowledge into explicit knowledge and back, and perhaps helping in understanding how to build dialogical communication tools that would also operate at the “tacit” level of knowledge rather than at the level of explicit information only.

The key underlying assumption in this article has been that high presence and/or high social presence and consequent other psychological effects, such as emotion and efficiency of cognition may induce a high level of “tacitness” in communication, hence making it more “natural” and open to creation of new ideas and more in-depth understanding of the meanings involved in communication. Naturally, this is a somewhat naïve and philosophical view of communication, but it may be beneficial to hold as the wholistic big picture when thinking of how to enhance communication inside corporations from the point of view of innovation efficiency.

The increased “tacitness” brought about the use of Psychological Customization systems in corporations can be roughly divided into two parts. First, it may be thought of increasing the information processing bandwidth between ready-made and “non-dialogical” information, such as financial news and the perceiver of that information. In this case, various techniques may lessen processing load of such information, induce presence and perhaps facilitate sense making and grasping of the meaning of information also.

Second, in the case of more dialogical information exchanges between two or more people inside an organization the system could be used for both expressing individual’s explicit knowledge (i.e. externalized information sent to other people or to a community website, for instance) and tuning one’s way of receiving those messages. The dimension of dialogue and exchange of thoughts via this more social interaction oriented use of technology will bring about also information substance-related issues, such as the capacity to ask for more information from the other person or sender of the message in case of misunderstandings or vagueness in communication.

Unfortunately, clear and explicit conclusions and design guidelines for optimal use of Psychological Customization systems in organizational settings is beyond the scope of this single article. Hence, the contribution of the presented approach remains conceptual and requires further empirical study to legitimize the ideas presented. However, various empirical studies indirectly support the idea that new types of Mind-Based Technologies may be used in organizational learning facilitation via technology inside the knowledge-based view of organizations.

With these limitations in mind some concrete ideas on how to approach the use of Psychological Customization in organizations are presented.

A key driver of value in corporations is innovation – the capacity of the organization to launch new types of concepts and products that succeed in the marketplace. Innovations are brought about by creative knowledge workers, empowered teams and individuals that self-organize and produce new ideas for products. Such teams are often also using technology-mediated communication in the process of innovating.

The road from technological efficiency to innovation efficiency remains unknown, however. One strong
candidate for the future technology platform of knowledge work is ambient intelligence. This entails the use of smart mobile devices, wireless access and intelligent, context sensitive office spaces that interlink workers 24/7, providing a dialogical connection to other workers and information over time and space. This vision will take time to be realized.

One thing to do in the more near future would be to try to increase the information processing bandwidth of knowledge workers of corporate intranet information and social interaction related information. This simply means that people understand the matters described in the information better and faster, perhaps leading to new ideas also. However, turning bits from the screen into useful knowledge and further to fuel for innovation is difficult.

Let’s take the example of using media information, such as financial news flows, corporate intranet or corporate e-learning material as aids in knowledge work. First, one may use active exploration tools, such as auto summarizing of video and audio to enable quick grasping of the overall structure of information. Text-based information can be automatically put into context, providing semantic maps into related information as well as automatically generated key terminology and graphs. New functionalities enable sense making faster and more efficiently as they are adapted to the expertise level of the user.

Second, one may personalize the way of presenting information as discussed in this article. For instance, providing a person who is fluent in processing audiovisual information with just that modality may increase his information processing bandwidth, perhaps leading to new ideas also. However, the same person has to browse through endless pages of text. The solution here would be to briefly model the workers’ information processing styles and match information presentation modality to this.

Finally, it is clear that as the relationship of innovation and knowledge work and the mediating and facilitating roles for information technology are better understood, more developed ambient, dialogical and information processing oriented technologies can be designed to support desired tasks, processes, teams and individuals.

References


Auditory Presence, Individualized Head-Related Transfer Functions, and Illusory Ego-Motion in Virtual Environments

Aleksander Väljamäe, Pontus Larsson, Daniel Västfjäll, and Mendel Kleiner
1 Department of Signals and Systems, Chalmers University of Technology, Sweden
2 Department of Applied Acoustics, Chalmers University of Technology, Sweden
3 Department of Psychology, Göteborg University, Göteborg, Sweden
4 Program in Architectural Acoustics, Rensselaer Polytechnic Institute, Troy, NY, USA

{aleksander.valjamae@s2.chalmers.se, pontus.larsson@ta.chalmers.se, daniel.vastfjall@psy.gu.se, kleinm2@rpi.edu}

Abstract

It is likely that experiences of presence and self-motion elicited by binaurally simulated and reproduced rotating sound fields can be degraded by the artifacts caused by the use of generic Head-Related Transfer Functions (HRTFs). In this paper, an HRTF measurement system which allows for fast data collection is discussed. Furthermore, effects of generic vs. individualized HRTFs were investigated in an experiment. Results show a significant increase in presence ratings of individualized binaural stimuli compared to responses to stimuli processed with generic HRTFs. Additionally, differences in intensity and convincingness of illusory self-rotation ratings were found for sub-groups of subjects, formed on the basis of subjects’ localization performance with the given HRTFs catalogues.

Keywords--- Self-motion perception, Auditory presence, Binaural reproduction, Individualized Head-Related Transfer Functions.

1. Introduction

Creating a sense of presence in the end-user of a Virtual Environment (VE) is one of the main goals of Virtual Reality (VR) technology. The feeling of presence is often described as a sensation of “being there”, whereas several other definitions exist [1]. Being “spatially present” in a specific context provides a stable reference frame, which allows for a good spatial orientation and spatial updating. Illusory ego- or self-motion (vection) can be described as a sensation of actual movement relative to a stable surrounding environment and has been shown to be closely related to spatial presence. For example, positive correlation between presence ratings and on-set times for induced illusory self-motion has been recently shown in experiments with visual stimuli [2].

A large body of research is concentrated on the illusory self-motion elicited by visual stimuli. On the contrary, research on auditory illusory self-motion received little attention until recently [3-6]. In our first experiments conducted within the European project POEMS (Perceptually Oriented Ego-Motion Simulation) [7], we focused on illusory self-rotation induced by sound-fields [8]. In this study, rotating sound sources were presented to subjects via headphones and the sensation of ego-motion in the opposite direction was expected.

These previous experiments were based on the ideas of ecological acoustics and it was hypothesized that the type of the sound source is an important parameter when studying auditory-induced illusory self-motion. Opposite to “artificial” sounds (e.g. pink noise), ecological sound sources can be classified by a listener into spatially “still” (e.g. a church bell) or “moving” (e.g. footsteps) categories. A major finding was that both presence ratings and experience of self-motion were highest for the sound fields containing sound sources from the “still” category. Speed of sounds’ rotation and number of sources also positively affected the ratings [8].

The stimuli in [8] were created using binaural technology, which is a two-channel spatial sound rendering technology where headphones are typically used for playback [9]. Pre-measured catalogues of Head-Related Transfer Functions (HRTFs) are used for binaural sound synthesis, where a non-spatialized (“dry”) sound is convolved with transfer functions corresponding to the desired spatial position of the source. Larsson et. al. [8] used a catalogue of non-individualized HRTFs provided for the CATT Acoustics auralization software [10].

However, binaural sound synthesized with non-individualized HRTFs often can be perceived as distorted because of the mismatch between the listener’s own and generic transfer functions. When generic HRTFs are used the most common problem is in-head localization (IHL), where sound sources are not externalized but are rather perceived as being inside the listener’s head [11]. Another known artifact is a high rate of reversals in perception of
spatial positions of the virtual sources, where binaural localization cues are ambiguous (cone of confusion), e.g. front-back confusion [9]. Errors in elevation judgments can be also observed for stimuli processed with non-individualized HRTFs [12]. Applying individualized HRTFs for auditory VEs can significantly reduce the artifacts described above [11].

One of the goals of this study was to test the performance of the HRTFs measurement system designed by Chalmers Room Acoustic Group (CRAG). We decided to repeat some of the experiments on illusory self-rotation presented in [8], this time using individualized HRTFs. Responses from verbal probing in [8] showed that in some cases the participants experienced artifacts in presented sound scenes, such as non-circular trajectories and nonrealistic, very close distances to the moving sources. We believe that these artifacts may have been caused by the use of non-individualized HRTFs and that these artifacts in turn might have influenced ego-motion and presence ratings.

The main hypothesis for the current experiment stated that the higher presence ratings should be achieved with individualized HRTFs, since previous studies indicate that spatialization and localization may be linked to presence experiences [13-14]. A second hypothesis was that improved spatial quality of auditory scene might affect subjects’ experience of ego-motion.

Auditory localization performance can depend on several factors in complex spatial sound scenes. Langendijk et. al. [15] studied the effect of localization of target sounds in the presence of one or two distracter sounds, which were interleaved but not overlapped with the target sound in time. They found that the localization performance was degraded as the number of distracters increased. In auditory VE with multiple ecological sounds target-distracter pairs can easily occur.

We decided to investigate how distracters added to the auditory VE can influence presence ratings. Our third hypothesis was that adding the auditory distracters, which are irrelevant to the sound scene, would decrease the localization accuracy (divided attention). This in turn might decrease the influence of non-individualized HRTFs artifacts on the sound scene perception in VE.

2. HRTF measurement system

The procedure of measuring catalogues of individualized HRTFs is a cumbersome and time-consuming procedure (for reviews see [9, 12, 16-17]). In line with POEMS project requirements, it was decided to develop a HRTFs measurement system which would allow for fast data collection in non-anechoic and somewhat noisy environments such as offices. In the CRAG HRTF measurement system the transfer functions are recorded for a grid of spatial positions on a virtual sphere, which center is aligned with a subject’s head-related coordinate system.

2.1. Physical setup

In our laboratory setup we built an array of 32 loudspeakers as shown in Fig. 1, which can be seen as one sector or “vertical slice” of a virtual sphere with a radius of 1.25 meters (far-field acoustical mode measurements). Loudspeakers were non-uniformly placed at 16 elevation angles, which guarantee a resolution of less than 10 degrees in the vertical plane. A test person must be shifted 19 times (20 sectors) for the full HRTF catalogue measurement resulting in less than 8 degree resolution in the horizontal plane. If higher frontal resolution is required, one additional sector can be measured, which results in 63 measured azimuth positions in the horizontal planes with elevation angles of 16, 6, -4 and -14 degrees.
subject. This type of HRTF measurements gives results comparable to those measured with a microphone probe immediately inside the ear canal opening [18], but the procedure is faster and more convenient for the subject under measurement. An M-Audio Audiophile 2496 PCI-bus card is used for playback and recording of measurement signals because of its large dynamic range and the availability of driver routines for Linux. Specially written software for raw HRTFs data collection is used.

2.2. Post-processing of measured data

In our system we use frequency sweeps (quadratic chirp) as the deterministic stimulus for measurements of Head-Related Impulse Responses (HRIRs). This method was chosen due its immunity to harmonic distortions introduced by the measurement chain and the low crest factor of the signal (ratio between peak and root-mean-square in the voltage values) [19]. The latter property helps to ensure desired 20 dB signal-to-noise ratio for a desired frequency range. The chirp is band-limited from 0.1 to 15 kHz and its duration is 2048 samples (\(\approx 43 \text{ ms}, f_s = 48 \text{ kHz}\)).

After the measurement procedure, the raw HRIRs are processed using Matlab™ software. The raw HRIRs are first deconvolved with the chirp signal and a half-Hanning window is applied to keep first 256 samples of the response. At this stage, the Interaural Time Delay (ITD) is estimated using the difference between onset times of the left and right HRIRs. This allows for optimizing in further processing steps by working only with the HRFTs magnitude responses. The ITD is added only to the final HRIR dataset. This decision was motivated by the assumption that it is possible to use linear-phase HRFTs for binaural sound synthesis without adding perceptually significant artifacts [20].

Furthermore, the raw HRIRs collected by the system have to be compensated for the artifacts coming from transducers limitations. For this purpose “baseline” or free-field responses are measured with the microphones placed at the center of the virtual sphere. Free-field corrected HRTF magnitude responses can then be obtained by division of the raw HRTF by the corresponding baseline TF data in frequency domain.

Interpolation is needed to obtain a final HRTFs catalogue with uniformly distributed spherical grid. For the current experiment, a 5-degree resolution in the horizontal plane was required for smooth rendering of moving sound sources. This is comparable with an average localization blur, which is 3 degrees for frontal positions and up to 20 degrees for peripheral and rear positions [11]. Spherical spline interpolation in the frequency domain is known to give best results compared to other interpolation types [21]. Before interpolation, magnitude responses are smoothed using the procedure described in [22]. Instead of using a perceptually motivated reduction of magnitude response, smoothing is applied to all data points. After interpolation, HRTF magnitude responses are used for creating linear-phase FIR filters. A circular shift equal to the earlier estimated ITD is introduced within filter pairs, which represent HRTFs for certain spatial locations.

The processing steps presented above were used for creating the stimuli for the current experiment. One of the goals of this study was to test the performance of the CRAG HRTF measurement system and apply all necessary corrections at the post-processing stage if needed. At this stage we found good results for the HRTFs for the frequency range from 0.1 to 13 kHz. Deficiencies observed for frequencies above 13 kHz had no affect to the experiment as ecological sounds with frequencies from 0.1 up to 10 – 12 kHz were used.

3. Method

Twelve subjects (five male) with a mean age of 24 (SD 2.2) from the previous study described in [8] participated in the experiments. All subjects had normal hearing verified by a standard audiometric procedure [23]. After completing the experiment, subjects were debriefed, thanked and paid for their participation.

3.1. Measures

To assess auditory induced vection and subjective presence sensation, three direct measures were used in this experiment: presence, vection intensity and convincingness of vection.

Presence was defined in the questionnaire as “a sensation of being actually present in the virtual world”, which corresponded to the single perceptual dimension without any interaction with the VE. Vection intensity corresponded to the level of the subjective sensation when experiencing self-motion. On the Convincingness scale subjects had to report how convincing the sensation of self-motion was. It should be noted that the convincingness and intensity ratings often are highly correlated. Ratings of all three measures were given on a 0-100 scale.

Apart from the direct measures listed above, an indirect binary measure, reflecting the number of ego-motion experiences, was used (participants were asked to verbally indicate the direction of self-rotation). While on-set time for vection experience is often used in experiments with visual stimuli, the previous experiment [8] on auditory-induced vection indicated that this measure showed large inter-individual variance. The present study measured onset time, but since no systematic effects were found, results from this measure will not be presented.

3.2. Stimuli

In the current experiment, rendering of acoustic environment was not considered as being important; therefore all stimuli were simulations of an anechoic environment rendered off-line. Three parameters varied in the presentation of the rotating sound field:

- Rotation velocity (20 or 60 degrees/second)
- Number of concurrent sound sources (1 or 3)
- Type of HRTFs catalogue used for stimuli synthesis (individualized or generic catalogue)

Since results from the previous experiment showed that the “still” type sound sources are the most instrumental in
inducing ego-motion, only sounds from this category were used: “bus on idle”, “small fountain”, and “barking dog”. The stimuli duration was approximately 1 minute and consisted of the following parts: 3 seconds in stationary listening position, 4 seconds acceleration to maximum velocity, 60 seconds constant rotation speed and 4 seconds deceleration.

The stimuli were synthesized using one horizontal slice of HRTFs at –4 degree elevation. The stimuli synthesized with individualized HRTFs was contrasted with one processed with generic HRTFs catalogue, which resulted in 4 pairs of stimuli kept together for the verbal probing purposes. HRTFs measured from the KEMAR (Knowles Electronic Manikin for Acoustic Research) mannequin were used as the non-individual catalogue. Headphone equalization was applied to the final sound excerpts in order to prevent coloration artifacts.

For testing the effects of irrelevant auditory distracters, 20 clicks (6 kHz carrier, 4 ms duration) were added at random time moments to the two stimuli from the main set described above (1 and 3 sound sources, synthesis with generic HRTF, velocity of 60 degrees/second). Clicks were also convolved with KEMAR HRTFs and appeared at random positions in space. During the experiment stimulus with the clicks always followed the same stimulus without distracters hence creating two pairs for verbal probing.

Apart from the pair restrictions described above, all 10 stimuli were presented in the randomized order for proper statistical analysis.

3.3. Procedure

The experiment was conducted in a semi-anechoic room, where stimuli were played back with Beyerdynamic DT-990Pro circumaural headphones. Participants were asked to report verbally the direction of their rotation – i.e. left or right, if they felt self-motion during the particular stimulus playback. Stimulus was stopped after the positive ego-motion response and subjects were asked to rate presence, intensity and convincingness.

In the current experiment presence is studied from the ego-motion perspective and rapid interruption, which could certainly influence the presence ratings, is acceptable. If the ego-motion sensation was not reported during the stimulus playback, only the presence rating was asked. Apart from the verbal responses to the questionnaire, verbal probing was done by the experiment leader.

Taking into account results by Lackner [3], special measures were taken in order to achieve auditory ego-motion. Participants were seated on an ordinary office chair, which was mounted on an electrically controllable turntable with a wooden base plate. Subjects were instructed that turntable would be used during some of the experimental trials. However, the turntable was still throughout the experimental trials. This manipulation was foremost used to make participants believe that they could actually move during the experiment. In addition, the turntable height prevented the subjects from having their legs any contact with the floor. In order to make the experimental setup look more convincing, four loudspeakers, placed around the experimental chair, were visible to participant as he/she entered the test room. The loudspeakers were never in use during the experimental trials. Finally, during the experiment participants were blindfolded.

3.4. HRTFs quality-test

In order to evaluate the quality of the individualized HRTFs a short test was performed before or after the main experiment. The purpose of this quality-test was to justify the improved localization performance compared with localization when using the generic (KEMAR) HRTFs catalogue. Instead of common strategies of defining absolute localization accuracy (e.g. [24]), we decided to use a simplified procedure, where a level of the most usual binaural rendering artifacts acts as indirect quality measure (it was important to keep the quality-test short in duration, as it had to be conducted together with the main experiment). The major parameters then are the front-back confusions rate and the externalization of perceived virtual sources.

It was decided to evaluate 6 positions on the horizontal slice at –4 degrees elevation: three in front (315, 0 and 45 degrees) and their reverse positions on rear (225, 180 and 135 degrees). A fountain sound of frequency range from 2 to 12 kHz was used as a sound source, since it is well known that, apart from small head movements, spectral differences in HRTFs around 5 and 9 kHz help to resolve front-back confusion [9]. Figure 2 shows how spectral differences can vary between the subjects.

![Figure 2 Difference in spectra between two front-back source locations on a cone of confusion (45 and 135 degrees azimuth, -4 degrees elevation) for three representative subjects](image-url)

Four different measures were used for evaluating the HRTF quality: 1) the front-back confusion ratio, 2) the relative distance to the sound source, 3) errors in the elevation perception, and 4) responses from the verbal probing.

The front-back confusion ratio was based on the subjects’ estimates of the spatial position of the sound source. In order to simplify this task, participants were
asked to use a clock metaphor when verbally indicating one of the six positions listed above (e.g. 0 degrees azimuth corresponds to 12 o’clock).

For the distance evaluation, participants were asked to rate the distance to the source in meters. Later, answers were converted to the relative scale from 0 to 100, where 100 corresponded to the maximum perceived distance to the source from the stimuli pair processed with individualized or generic HRTFs.

For the elevation perception measure, subjects were asked to indicate the position of the source regarding the horizontal plane in the head-related coordinates system. Subjects were asked to indicate the height of the sound source relative to this plane.

As a last binary measure of the HRTF catalogues’ quality, results from the verbal probing in the main experiment were used. Based on the subjects’ comments on perceived trajectories of virtual sources, their distance and overall spatial scene consistency, decision was made regarding the preference between stimuli processed by individualized or generic HRTF catalogues.

The quality-test stimuli consisted of three pairs of subsets synthesized by individualized and generic HRTFs catalogues. Each subset contained a fountain sound sequentially presented for 6 seconds from each of 6 spatial positions in random order. One pair was used for the elevation perception measure and two other pairs for collecting data on the front-back confusion ratio and the perceived source distance.

4. RESULTS

4.1. Sub-groups based on the HRTFs quality-test

Results from the quality-test for all four measures highly varied among the subjects. However, when analyzing all 4 measures for each subject, these results could be combined into a final measure of preference between generic and individualized HRTFs used for the stimuli synthesis. Based on this binary measure, subjects were subdivided into two subgroups for further analysis: a G-group (better localization with individualized HRTFs) and a B-group (no clear preference between generic and own HRTFs). It has to be noted that localization performance could be degraded either by the HRTFs catalogues accuracy or due to individual localization abilities (see section 5 for a discussion).

For several subjects, the difference between the performance with individualized and generic HRTFs catalogues was not prominent but for a proper statistical analysis an equal number of participants was allocated to each group (i.e. median split). In this case the binary measure based on the verbal probing (see section 3.4) was used for the final decision.

Tables 1 and 2 show the average results for the three quality-test measures for all subjects and for the two subgroups formed. Table 1 shows average rates of front-back confusion, where responses from individualized and KEMAR HRTFs are compared. Means for all subjects showed an 11 % increase in the front-back confusion ratio for generic HRTFs. The sub-group analysis showed no effect for the B-group, but for the G-group 20 % improvement was found when individualized HRTFs were used.

<table>
<thead>
<tr>
<th>HRTFs</th>
<th>All</th>
<th>G-group</th>
<th>B-group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ind.</td>
<td>33</td>
<td>32</td>
<td>35</td>
</tr>
<tr>
<td>KEMAR</td>
<td>44</td>
<td>53</td>
<td>35</td>
</tr>
</tbody>
</table>

Table 1 Average rates (%) of front-back confusion for all subjects and the subgroups

Table 2 shows the responses for distance perception, where sub-group differences can be clearly seen. For the G-group, individualized HRTFs resulted in more distant percepts of the virtual source - 14 % improvement. For the B-group, the effect was reversed, resulting in small 8 % degradation for the stimuli processed with individualized HRTFs catalogues.

<table>
<thead>
<tr>
<th>HRTFs</th>
<th>All</th>
<th>G-group</th>
<th>B-group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ind.</td>
<td>61</td>
<td>72</td>
<td>50</td>
</tr>
<tr>
<td>KEMAR</td>
<td>58</td>
<td>58</td>
<td>58</td>
</tr>
</tbody>
</table>

Table 2 Average distance responses in relative scale (100 corresponds to the most distant percept)

The elevation perception measure was strongly biased by the high-frequency contents of the sound used for the quality-test. Sound with such characteristics is usually perceived as being located higher than its actual position [9]. This was clearly seen from the subject responses; more than 70 % of source positions were judged as being located above the horizontal plane. In general, both groups showed smaller deviations in the sound height judgments when individualized HRTFs were used. At the same time, the B-group showed roughly 3-times larger deviations in the answers to this measure compared to the G-group.

4.2. Main effects: Ego-motion and Presence

All dependent variables were submitted to separate 2(HRTF) x 2(Number of Sources) x 2(Velocity) ANOVAs. First, the analysis using the binary ego-motion measure yielded no statistically significant differences. The analyses of intensity and convincingness showed no significant main effects for HRTF’s or velocity, but in both instances a significant main effect of number of sources was found. The means for the intensity measure were 20.5 (1 source) vs. 31.0 (3 sources), F(1, 11) = 4.29, p<.05. Similarly the means for convincingness were 22.2 (1 source) vs. 31.3 (3 sources), marginally significant for a F(1,11) = 4.13, p = .06. No other effects were significant for these measures.

For the presence ratings a significant main effect of HRTF was found where the individualized HRTFs yielded higher presence (M = 61.8) than did the generic KEMAR HRTFs (M = 57.8), F(1,11) = 5.43, p<.05. No other main effects or interaction effects reached significance.
Stimuli with auditory distracters affected only the presence ratings. While almost no effect was found for the stimuli containing multiple sources: M = 58.3 (with distracters) vs. M = 59.2 (without distracters), presence ratings for the single rotating sound source were higher for the stimulus with distracters (M = 58.8) compared to non-distracter condition (M = 52.9). In addition verbal probing showed a clear difference in overall judgment of the presented sound scene and the trajectories of the virtual sound sources. In the presence of distracters less distorted trajectories were perceived.

4.3. Subgroup differences in ego-motion perception

HRTFs quality-test results presented in Tables 1 and 2 motivated a sub-group analysis of the data from the main experiment. Table 3 shows the results from the statistical analysis of 3 direct measures used in the main experiment, where 4 pairs of stimuli were used for non- and individualized HRTFs catalogues comparison. However, since the sample size was too small to allow for parametric statistical analyses, only trends are reported here.

As was shown in [8], multiple sources positively affect presence ratings and this can be seen for both subgroups in the table. The same trend continued for intensity and convincingness of ego-motion, but in the B-group the difference is negligible. On the contrary, the G-group showed clear discrimination in ratings when the number of sources presented in stimuli were increased.

<table>
<thead>
<tr>
<th>Measures</th>
<th>Source type</th>
<th>G-group</th>
<th>B-group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intensity</td>
<td>Single</td>
<td>10.4</td>
<td>30.6</td>
</tr>
<tr>
<td></td>
<td>Multiple</td>
<td>30.0</td>
<td>32.1</td>
</tr>
<tr>
<td>Convincingness</td>
<td>Single</td>
<td>12.9</td>
<td>31.2</td>
</tr>
<tr>
<td></td>
<td>Multiple</td>
<td>30.4</td>
<td>32.3</td>
</tr>
</tbody>
</table>

Table 3 Effects of concurrent sound sources number on intensity and convincingness average ratings in the sub-group analysis

5. Discussion

The major finding in the present study was that stimuli processed with individualized HRTFs catalogues resulted in a significant increase of presence ratings as compared to stimuli processed with generic HRTFs. Several other lines of research have independently showed that individualized HRTF increase spatial perception and spatial abilities [16, 24] or that more spatialized sound increase the sense of presence [8, 13-14]. However, to the best of our knowledge, this is the first study to show a direct link between individualized HRTFs and spatial auditory presence.

In addition, the results for ego-motion (intensity and convincingness ratings) showed consistency with the previous findings reported in [8] - a higher number of concurrently rotating sources and a higher rotational speed increased these ratings.

The sub-group differences shown for the HRTF quality-test (Tables 1 and 2) and the results of the main experiment (Table 3) can be explained either by the errors that occurred during the individualized HRTFs measurement procedure or by the subjects’ auditory localization abilities. Localization performance varies between the individuals and terms “poor” or “bad” localizers are used in the literature, e.g. [9, 25].

In general, the results from the HRTF quality-test were influenced by several factors. First, participants were not trained to perform localization tasks in previous experiments and the quality-test procedure did not include a training period. Second, utilization of an ecological sound as a stimulus might bias the judgments of the participants. More work has to be done for designing fast and reliable procedures for evaluation of HRTFs catalogues` quality.

Results presented in Table 1 showed that the front-back confusion ratio was significantly increased for G-group subjects when using non-individualized HRTFs. However, no such difference was found for B-group. It is known that the performance of good localizers degrades when using bad localizers’ HRTFs [9]. The opposite effect, when bad localizers improve their abilities using other person HRTFs catalogues, has not been fully evaluated [9]. Larger deviations in rated source heights found in the B-group for both individualized and generic HRTFs catalogues suggest the influence of individual performance. Table 3 presents the last evidence for difference in the sub-groups’ performance: for intensity and convincingness ratings no discrimination between stimuli containing single or multiple sound sources was done by the B-group subjects.

When re-examining the data from the previous experiment [8] for the same participant sub-groups, the trends presented in Table 3 were not found. Therefore it is more likely that sub-group difference is due to the errors occurred during the HRTFs measurements procedure. However since different stimuli synthesis procedure was used in [8] a direct comparison of the subject responses is not possible and further studies of this finding is needed.

Preliminary tests with 2 stimuli with added clicks supported the hypothesis that adding distracters to the auditory VE might influence overall perception of the sound scene and decrease influence of artifacts caused by non-individualized HRTFs utilization. Specially designed experiment should shed further light on the effects of divided attention on quality judgments of VEs.

6. Conclusions and future work

There are several conclusions from this initial investigation. First, it was found that individualized HRTFs increase presence ratings. Second, the results were consistent with the previous results reported in [8], where the number of sound sources influenced both presence ratings and ego-motion experiences. Third, inter-group differences were found within the subjects, which were more likely caused by the errors occurred during the fast measurement of individualized HRTFs catalogues. Participants from the group with poor localization performance showed no discrimination in intensity and convincingness ratings for the number of presented sound sources. Finally, it is important to note that stimuli
processed with generic (KEMAR) HRTFs also induced ego-motion regardless to the lowered rendering quality of spatial scene.

The authors are planning to test their findings in subgroup differences using a higher number of participants for reliable statistical analysis. Modification of existing methodology of both measuring and evaluating of individualized HRTFs catalogues is also planned for upcoming work. Influence of distracters on presence ratings and illusory ego-motion sensation is another topic for the follow up experiments.

Acknowledgements

This work was supported by the EU POEMS project (IST-2001-39223). The first author of this paper would like to thank for the support from Alfred Ots’ scholarship foundation.

References

[7] EU-Project POEMS (Perceptually Oriented Ego-Motion Simulation) URL: http://www.poems-project.info/
[10] CATT-Acoustics v 8.0 (Computer software), Gothenburg, Sweden. URL: http://www.catt.se
An Investigation of Presence Response across Variations in Visual Realism

Vinoba Vinayagamoorthy, Andrea Brogni, Marco Gillies, Mel Slater, Anthony Steed
Department of Computer Science, University College London, London WC1E 6BT, UK
†University College London, Ross Building pp1, Adastral Park, Ipswich IP5 3RE, UK
{V.Vinayagamoorthy | A.Brogni | M.Gillies | M.Slater | A.Steed}@cs.ucl.ac.uk

Abstract
This paper describes reports on the results of an experiment designed to study the impact of realism on the reported presence in an immersive virtual environment. An experiment was carried out with 40 participants who were asked to walk through a virtual street, which had virtual characters walking through it. Two factors were varied – texture quality (2 levels) and virtual character realism (2 levels). 10 participants were assigned to each cell, which was also balanced for gender. The results suggest that the lowest presence was achieved with the higher fidelity characters but the less varied textures. Other factors such as perceived realism of character behaviour were important. Finally a system called PIAVCA for virtual character control is introduced which was used in conjunction with DIVE [25, 29].

Keywords--- Virtual Environment, Textures, Virtual Characters, Visual Realism, Presence.

1. Introduction
Even with the current state-of-the-art technology available to us, it is difficult to design and implement a believable virtual environment. Traditionally it has been assumed that the more believable virtual experiences are constructed by making the environment as visually realistic as possible [9]. However a fundamental difficulty in achieving total visual realism is the complexity of the real world. The richness of the real world is observable in the surface textures, subtle colour gradations, shadows, reflections, and the slight irregularities in the surrounding objects [8].

On the other hand it is equally arguable that even simulations with a lower degree of realism can still contain the more important information necessary to give a believable experience such as those perceived in flight simulators or in stressful scenarios [18, 32]. We present a study that focuses on trying to explore various features that enhance reported presence in an immersive virtual environment (IVE). In particular we aimed to explore the impact of different texture quality and visual realism of virtual characters on the overall level of presence in a populated urban virtual environment (VE).

2. Objectives
Our main goal was to investigate whether there needs to be consistency between the levels of realism of the different elements within a scene. In this case this means whether the level of realism of the buildings needs to be consistent with the level of realism of characters populating the environment. Our interests in this study were three-fold.

- Firstly to investigate the relationship between reported presence and the variety of textures on the buildings in the scene.
- Second, the impact of the visual realism of the virtual characters on reported presence.
- Third, to investigate whether there needs to be consistency between the levels of realism of the different elements within the scene.

We hypothesised that less repetitive textures in the scene and more lifelike characters will enhance the participants’ presence experience in the virtual world. We vary visual realism of the IVE by altering the number of textures used on the buildings and by using two types of virtual characters in the study. Although both types of virtual characters were not visually realistic in terms of their human appearance, one was deliberately designed to be cartoon-like, and the other to have an appearance that was more realistic. For example, the second kind had a face that was texture-mapped from real human faces. We call the more realistic type ‘higher fidelity’ (HF) and the other form ‘cartoon form’ (CF). We will use these terms in the remainder of this paper.

We also wished to ascertain if there was an interaction effect between the visual realism of the character and the richness of texture variety of the environment. This was especially important, because earlier research [12] indicates that consistency is required between virtual character behaviour and appearance. Here we wished to see if this extended to consistency between environmental and virtual character realism.

A subsidiary goal was to investigate the correlation between three measures of presence: stress response measure by heart-rate variability, breaks in presence as measured by verbal report, and self-report questionnaires. We also used this study to test Platform Independent API
for Virtual Characters and Avatars (PIAVCA) in conjunction with DIVE\textsuperscript{1} [25, 29].

3. Background

It has been claimed that however visually appealing a virtual environment is, if it is static and empty of change and behaviour, the immersive experience is of limited interest to the participants [3]. In Neil Stephenson’s Snow Crash [27], surroundings in the Metaverse are highly photorealistic while characters are both visual realistic and expressive depending on the status of the owner. In fact both visual and behavioural realism perform seamlessly in real-time to produce a life-like experience.

Even though it is possible to generate visually realistic virtual environment and characters [1], one of the main constraints preventing the usage of these highly realistic models is the higher expectations of participants leading to increased sensitivities to inconsistencies in the VE [30]. It is not only the shape of virtual bodies that matter in the experience of virtual environments but also the level of detail with which they are represented [19]. Another major constraint is the tension between realism and real-time [11]. This constraint is of particular importance in larger environments due to the increased demands made on graphics cards.

It has been established that in order to enable participants to experience acceptable levels of presence in VEs, the virtual characters have to portray some level of non-verbal behaviour. This has been found in a series of studies involving interaction between three participants in a shared VE [22, 26, 30]. Tromp et al. concluded that higher visual realism in characters may lead to a heightened expectation for behavioural realism [30]. In particular in [28] there is anecdotal evidence that participants expect more photorealistic characters to behave in a manner, which portray greater human-like qualities. What seems to be important is that visually realistic worlds and characters have similar levels of behavioural realism [11, 12].

The issue of consistency within a scene can be traced back to works as early as Gibson [13, 14]. One insight recent that a recent study has provided is that pictorial or visual realism is not sufficient and sometimes not necessary to create a sense of presence or co-presence [12]. This is partially supported by work done by Longhurst et al [17], where studies showed artistically enhancing an image through the addition of dust, dirt and scratches increases the perceived realism of an image. In another study [6], carried out to explore the visual elements necessary to make an effective VE, it has been suggested that improving visual realism in a VE is not just about rendering flawless models.

The current study focuses on the perception of realism in urban virtual environments. It was designed to explore three aspects important to create a believable virtual environment: sense of space, realism of textures in VE and degree of visual realism of characters.

4. The Experiment

4.1. Factorial design

Our study was designed to assess some limits of visual realism in enhancing the believability of virtual environments.

We choose to study two factors:
- The level of detail employed in our VE with respect to the number of textures used;
- The visual realism of the inhabiting characters.

A gender-balanced between-group two-by-two factorial design was used in the experiment.

<table>
<thead>
<tr>
<th></th>
<th>Cartoon (CF) Characters</th>
<th>Form Characters</th>
<th>Higher Fidelity (HF) Characters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Repetitive Textures</td>
<td>5 males + 5 females</td>
<td>5 males + 5 females</td>
<td></td>
</tr>
<tr>
<td>Non-Repetitive Textures</td>
<td>5 males + 5 females</td>
<td>5 males + 5 females</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Factorial Design of the Experiment

Forty participants were assigned at random to one of the four conditions as illustrated in Table 1. Participants were recruited by poster campaigns across the university and paid £5 to take part in a one-hour study.

4.2. The virtual environment

A virtual world was created in 3D Studio MAX to resemble a typical street-like environment. One of the factors in the study was dependent on the number of different textures used in the scene. In the conditions with “Repetitive Textures”, 20 different textures were used to provide shop billboards and fronts whereas in the conditions involving “Non-Repetitive Textures” there was twice the number of textures (~40).

Figure 1: The Street.

The environment consisted of a high street lined with buildings on either side with a few secondary streets off the main street. The ends of the main and the secondary streets were sealed off in order to force the participants to stay within the limits of the designed world. A few tall buildings

\textsuperscript{1} Distributed Interactive Virtual Environment
were modelled in the horizon outside the peripherals of the street to maintain the illusion of a curtained off the high street. The same street model was used throughout the experiment with the only difference being the number of different textures used.

4.3. The characters

The second factor varied in the study was the visual realism of the characters. There were two types of characters used: cartoon form (CF) (Figure 2) and higher fidelity (HF) (Figure 3). All the characters were H-Anim [31] complaint humanoids. All the characters were animated to walk using the same motion file and given the same behaviour.

The characters were animated to walk out onto the street through one of four doors at the ends of the main street. The character walked to the opposite end of the street using a cyclic walk pattern defined by the motion file taking care to avoid each other and the participant undergoing the study. Each character started walking at slightly different times so that the animation of the characters did not appear synchronised. The only varying factor between conditions was in the visual appearance of the characters.

The street was populated with a total of 16 virtual characters walking up and down the main street however only 8 characters were on the street at any one time. The decision to activate only 8 characters in IVE at the same time was made due to a technical constraint in rendering all the characters in run-time at an acceptable frame rate.

4.4. Software implementation

The software used was implemented on a derivative of DIVE 3.3x [10]. This was recently ported to support spatially immersive systems [25]. DIVE (Distributed Interactive Virtual Environment) is an internet-based multi-user virtual reality system in which participants can navigate in a shared 3D space and interact with each other. Plugins make DIVE a modular system. DIVE also supports the import and export of VRML and several other 3D file formats. The VE used was created in 3D Max and exported to VRML.

The characters were animated using PIAVCA, the Platform Independent API for Virtual Character and Avatars. This is a character animation library that is designed to be independent of any underlying graphics engine, and which has been ported to DIVE. PIAVCA is able to animate characters using motion data stored in Biovision BVH format. BVH is a file format for representing human motion in a form that is independent of the character that performs this motion. This enables a single piece of motion to be applied to multiple characters. The motions data can either be acquired through motion capture or hand animation, for this study we used walking motion from a motion capture library. As well as simply playing animation data PIAVCA has a number of facilities for manipulating and sequencing motions. There are a number of methods of manipulating the pieces of motion, including smoothly sequencing motions into each other, interpolating between motions, and manipulations of individual motions such as turning a motion through an angle. A callback system is also available to sequence new motions at appropriate times.

A callback was used to animate the characters walking down the street and avoiding collisions with each other and the participant. A piece of motion representing a single footstep was used. At the end of each footstep the callback determined whether evasive action was required in the next footstep. Each character was aware of the position of each of the other characters. The position of the participant was sensed by DIVE and with this information the participant was treated as another character. If any of the other character or the participant were in the path of the character it would alter its path, using PIAVCA motion manipulations to smoothly step to one side. The same callback would also ensure that the character would remain on the street. If no evasive action was required a normal straight footstep was chosen. If a character reached the end of the street it would walk into a shop entrance and then become inactive.

4.5. Equipment

A CAVE™-like [7] Immersive Projection Technology (IPT) system, referred to as a ReaCTor, was used to generate our IVE. The Trimension ReaCTor used in the study, consisted of three 3m x 2.2m walls and a 3m x 3m floor. It was powered by a Silicon Graphics Onyx2 with 8 300MHz R12000 MIPS processors, 8GB RAM and 4 Infinite Reality2 graphics pipes. This machine processed all the graphics and audio input pertaining to the ReaCTor. The participants wore CrystalEyes stereo glasses, which were tracked by an Intersense IS900 system accurate to within 2mm with an end-to-end latency of 50ms. The ReaCTor runs at a maximum refresh rate of 45Hz in stereo. In the remainder of this paper we will use the term ReaCTor to refer to this system.
4.6. Procedure

Due to the complex procedure in the study, two experimenters guided the participants through the study. On arrival for the study the participants were asked to sign a consent form, which gave them information about the equipment, an outline of what the study involved and informed them of possible negative effects from using the system such as simulator sickness. They were told that they could withdraw from the study at any time without giving a reason, and that they agree not to drive or operate complex machinery for at least 3 hours after the conclusion of the study.

The participant was then given an online questionnaire asking them to provide information about their background for demographic purposes. They were also asked to fill in a memory questionnaire designed to assess how the participant recalled a physical surrounding they had been in earlier that day. The participant was then given a short training to help in understanding the concept of “breaks in presence (BIPs)” [4]. In this training they were asked to look at four Gestalt pictures and switch their focus from one perceived image to the other. The transitions they experience in the viewing exercise were equated to transitions they might feel to the real world of the laboratory while in the virtual environment. Through out the study, BIPs were referred to as “transitions to real”.

The participant was then invited to step into the ReaCTor and the experimenters fitted the participants with physiological sensing devices to collect the participants’ Electrocardiogram (ECG), Galvanic Skin Responses (GSR) and Respiration measures. The participant was then asked to stand still and quiet in the darkened ReaCTor for a full minute and a half. This was necessary to gather readings about each individual participant while they are inactive in order to have a baseline to compare physiological readings taken during the study. At the end of the minute and a half, a virtual training room containing three-dimensional numbers appeared on the walls of the ReaCTor. One of the experimenters showed the participants how to move through an environment moving from number to number. At the end of the training, the participant was told to exit through a door onto the street and to do as they pleased for a few minutes. We wanted the participant to notice any anomalies in the virtual environment and point it out in the post-session debriefing. At the same time, we felt it would not be wise to ask the participant to look for anything in particular since this might make the participants more interested in completing the task than getting involved in the experience. They were also asked to signal any transitions to real that they felt by pressing a button on the joystick they used for navigation. At this stage, the experimenters left the participant in the ReaCTor and opened the virtual door leading to the street (Figure 1).

While the study was in progression, the experimenters maintained silence and observed the behaviour of the participant in the VE. The participants were also videotaped during the study. At the end of 3 minutes, the global lights in the street were switched off and a target light was activated to point the participant in the direction of the training room.

The participant was then asked an immediate question asking them to access how present they felt overall in the street. The result from the immediate questionnaire is used in analysing the number of BIPs reported by the participant [23]. The participant was then asked to complete an online post-questionnaire designed to gather data on various matters including their subjective sense of presence (questions available in [23, 24]). They were also asked to fill in another memory questionnaire about the VE. An argument has been put forward that questionnaires are methodologically unsound for assessing presence when they are the sole instrument for assessment [20]. Hence in this study we have used questionnaires, post-experimental interviews, physiological measures, and BIPs. In this paper we concentrate largely on the questionnaire responses.

The session concluded with an audio taped semi-structured interview conducted by both experimenters with each participant.

5. Presence Measures

The sense of presence in the VE was evaluated by taking a number of measures, including questionnaires [16, 23], physiological measures [18], and reported breaks in presence (BIPs) during the experience [23].

5.1. Subjective

A variety of questionnaires were administered to assess the behaviour and views of the participants. Most of the questionnaires were administered online to each of the participant before and after their experience. We obtained responses on each of the following:

- Demographics: age, gender, occupation, language proficiency, experience in computer games, programming and virtual reality etc.
5.2. Physiological

Physiological measures were recorded during the participant’s experience, using non-invasive devices. In this paper we only consider ECG, which gives a representation of the heart’s electrical activity. Electrocardiograms can be used to obtain the HR (the number of times per minute the heart contracts). The oscillation in the interval between consecutive heart beats (RR intervals) as well as the oscillations between consecutive instantaneous heart rates is called Heart Rate Variability (HRV). HRV is a well documented marker in cardiology and physiological studies [2]. HRV analysis is based on measuring variability in heart rate; the distance in time from one heart contraction and the following (RR) is used to proceed in the time and frequency analysis of the HRV.

The RR intervals in ECG were used to construct a heart rate. There are several measures that can be used – in particular the mean heart rate, the HRV (standard deviation) and NN50 (the number of pairs of adjacent RR interval differing by more than 50 ms). These measures were collected during each study for three subsets of the duration of the experiment. These subsets were for the baseline period (1.5 minutes of silence and inactivity in the dark), the training (while training the participant in VR navigation) and the experimental period (3 minutes). Of greatest interest were the baseline and experimental periods.

Meehan, et al. [18] used physiological responses to measure how believable the experience of being at the edge of a pit was for the participants in his study. The hypothesis was that if the participants felt present in the room, the virtual pit would have been able to evoke physiological responses similar to those of a corresponding real environment. They note that heart rate and skin conductance measure the arousal of the individual, therefore might only be used when such arousal is intrinsic to the task, i.e. fear.

Other experiments have shown that GSR and HR could be used as objective measures in monitoring reaction in VEs [15]. Recent work [4, 21] showed that physiological measures can be used for event-correlated analysis, in particular in relation to BIPs. The detection of different properties in the physiological responses in the vicinity of an interesting event is a new strategy employed in objective presence measurements.

6. Results

6.1. Data variables

The independent variables in this study are variety of textures (repetitive textures and non-repetitive textures) and visual realism of characters (CF and HF). The explanatory variables collected a range of demographic data including habits such as experience in computer games play and physiological measures. The response variables collected using an online questionnaire was as follows.

- Presence as measured by the SUS questionnaire [23, 24] which is the count of the number of “high responses” out of 5 questions.
- Presence based on the 6 questions derived from the ITC-SOPI questionnaire [16] but using the high score count criterion.
- BIPs which is the number of reported breaks in presence.

In this study a 7-point Likert-type scale was used to gather responses to all the questions where 1 corresponded to strong disagreement and 7 was strong agreement. A score of 6 or 7 was counted as a high response.

6.2. Methods of analysis

In this regression model the deviance is the appropriate goodness of fit measure, and has an approximate Chi-squared distribution with degrees of freedom depending on the number of fitted parameters. A rule-of-thumb is that if the deviance is less than twice the degrees of freedom then the model overall is a good fit to the data (at the 5% significance level). When covariates are included, then this is equivalent to Two-Way Analysis of Covariance (again using the binomial logistic model).

In this regression model the deviance is the appropriate goodness of fit measure, and has an approximate Chi-squared distribution with degrees of freedom depending on the number of fitted parameters. A rule-of-thumb is that if the deviance is less than twice the degrees of freedom then the model overall is a good fit to the data (at the 5% significance level). More important, the change in deviance as variables are deleted from or added to the current fitted model is especially useful, since this indicates the significance of that variable in the model. Here a large change of deviance relative to the Chi-Squared distribution indicates a significant contribution of the variable to the overall fit of the regression model.
6.3. Basic findings

Table 2 and Table 3 give the mean presence scores using the high responses count method as collected using the SUS and the ITC-SOPI questionnaires respectively.

<table>
<thead>
<tr>
<th></th>
<th>CF Characters</th>
<th>HF Characters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Repetitive Textures</td>
<td>2.2±0.04</td>
<td>0.4±0.70</td>
</tr>
<tr>
<td>Non-Repetitive</td>
<td>2.3±0.26</td>
<td>2.3±1.77</td>
</tr>
</tbody>
</table>

Table 2: Mean of Presence (SUS) (maximum score = 5).

<table>
<thead>
<tr>
<th></th>
<th>CF Characters</th>
<th>HF Characters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Repetitive Textures</td>
<td>3.9±2.28</td>
<td>2.4±0.97</td>
</tr>
<tr>
<td>Non-Repetitive</td>
<td>3.4±2.27</td>
<td>3.6±1.71</td>
</tr>
</tbody>
</table>

Table 3: Mean of Presence (ITC-SOPI) (maximum score = 6).

Note the similarity between the two sets of results. These scores are not inconsistent with the results of the previous studies [12]. The lowest reported presence occurs in the virtual environment with repetitive textures and higher fidelity characters. Previous results indicate that what is required is consistency [12, 30].

6.4. Analysis of covariance on the presence scores

In this section we give the formal analysis of the binomial logistic model described in section 6.2 for the reported presence scores. Each of the independent variables and the interaction effect is significant, as is obvious from Table 2.

The overall model has deviance = 132 on 36 d.f. (i.e., the overall model is not a good fit). This model includes the main factors (variety of textures and visual realism of the characters) and the interaction term Texture*Character. Deletion of the interaction term increases the deviance by 10.72 on 1 d.f. is the tabulated Chi-Squared value on 1 d.f. at 5% is 3.841 and therefore this term highly significant. This strengthens the impression given by Table 2, which is that HF characters with repetitive textures give the lowest reported presence.

When additional explanatory variables are factored into the analysis, the following variables become significant.

- **Game-time**: The number of hours a participant spent in playing computer games is positively associated with reported presence (change in deviance on 1 d.f. = 7.73).
- **Real-behaviour**: This is the perceived behaviour realism of the characters. It is very highly significant (deviance = 27.12 on 1 d.f.) and positively associated with presence. In other words the more that behaviour was perceived as being real, the higher the reported presence. This is the most significant factor of all, contributing the greatest to the overall fit. It is important to note that the behaviour was the same for all the conditions in the study.

When all of these factors and variables are included in the model for reported presence, the overall deviance is 97.3 on 34 d.f., still not a very good overall fit, but much better than only using the independent variables.

6.5. BIPs

In the original paper [23], Slater et al. did not use the number of BIPs directly but rather had to take into account whether overall ‘presence’ was high or low. For example, if a person reports 0 BIPs this may be because they were never present or present all the time. It is important to know which side of the divide the participant falls on. As suggested in the original paper the participant was asked a question immediately after their experience in the street to rate whether their overall presence was in the street, the real world of the laboratory or about 50/50.

So as a rule if the participant felt less than 50% of the time in the street then the adjusted number of BIPs determined as the maximum number of BIPs minus the reported number of BIPs. In this case, highest BIP score was 13. This is a much simpler procedure for correcting for ‘high/low’ presence than the one used in the original paper [23], but does exactly the same job of reversing likely low presence BIPs, without the need for introducing an artificial time interval for a BIP as in the original paper.

A regression of the presence score from the logistic regression model on the adjusted number of BIPs was significantly and negatively correlated with reported presence (deviance = 14.53 on 1 d.f). However, unlike in [4] there was no correlation between the actual number of BIPs and the reported presence. If we take the actual number of BIPs as the response variable then it is significantly and negatively associated with the difference between the experimental NN50 score and the baseline NN50 score (deviance = 3.857 on 1 d.f.). This lends support to the proposition in [21] that breaks in presence are associated with physiological changes. There was no association between the texture and character factors and the number of BIPs. There were also no variations found in the heart rate measures across the varying conditions.

7. Discussion

The main result of the experiment was that the condition with repetitive textures and higher fidelity character representations produced a lower reported presence response from the other three conditions. This means that our original hypothesis, that the level of visual realism needs to be consistent, independent of the level, is not fully supported. We had sought to expand previous results that suggest that behaviour and visual representation should be consistent. It had been suggested in [12, 30] that participants represented by HF characters were treated differently than CF characters. This led us to speculate that characters’ being notably different in representation than the environment they inhabit, might lead the participant to
think the world inconsistent and thus might lead the participant to feel less present. This hypothesis can also be seen as an extension of the “uncanny valley” hypothesis [5]. That is, that as character realism (or robot anthropomorphism in the original paper) increases, believability increases, until realism reaches a point at which inconsistencies start to disturb believability. With multiple elements in the environment, there might be an “uncanny valley” for each of them, in which case designers have to be careful not to fall into any of the “valleys” because this might disturb the overall believability. The uncanny valley hypothesis would suggest that, when presented with more visually realistic characters, participants would find any inconsistencies in character and world more noticeable and jarring, whereas with more CF characters these inconsistencies would be overlooked. This prediction is supported by our study. What we can say is that there appears to be an interaction between the levels of realism. We would suggest that our understanding of what constitutes a “level of realism” is not well developed. There are many different types of visual properties that could be changed. It may be very hard to make the judgement about whether the visual appearance of different elements is consistent. In our environment, the levels of realism in the world were different only in the repeating nature of textures. In designing the experiment, we had to be careful not to vary the visual appearance of the world too much because, for example, simply removing the textures would have meant that one world was lacking texture gradient which is an important visual cue for surface normals.

A final point to make is the highest significant subjective measure in the study was the participants’ perceived sense of realism with respect to the behaviour of the characters. The more a participant perceived the characters behaviour to be real, the higher the reported sense of presence. This strengthens the views reported in previous studies [12, 30].

8. Conclusion and future work

In this paper we have presented a study that started to examine whether different elements of a virtual environment need to be presented with consistent levels of detail. Several authors have suggested or shown that there needs to be a match between visual appearance and behaviour of characters. However important characters are for the scene, they are placed with a space that also has visual appearance and behaviour. It is therefore logical to suggest that different elements need to be presented at similar levels of detail in order that no-one element is notably inconsistent from the others.

In a case-based experiment we started to study this for populated urban environments. We had two conditions of texture detail on the buildings, and two of levels of visual realism of characters populating that space. We found that within one of these conditions, “low texture detail” and “high visual realism of characters”, the participant’s reported presence was notably different from the others. This doesn’t confirm the original hypothesis that visual appearance between elements needs to be consistent, but we have discussed that it is very hard to determine visual consistency, and there may be other interactions other than consistency between the different conditions.

We suggest that this area deserves further work, because it could indicate useful criteria by which to start discriminating between contributions of different elements in a scene to the overall sense of presence. Previous work has treated presence as a single response to an experience, whereas that experience is composed of interactions with many different elements with widely varying attributes. Work on BIPs has indicated that participants can be very sensitive to small inconsistencies, so it is worth developing an understanding of how those inconsistencies might arise. Furthermore, since an increasing proportion of the effort in building VEs today goes to modelling elements in the scene, it will be important for designers and engineers to have guidelines about consistent levels of detail, so that, the appropriate design choices can be made. Specifically, if consistency is really an issue, we would like to avoid investing too much effort in the construction of particular element in a scene if it would then be inconsistent with the other elements.

9. Acknowledgements

This research is supported by the EPSRC funded Interdisciplinary Research Collaboration (IRC) project: EQUATOR (EPSRC Grant GR/N15986/01). The original higher fidelity characters were created by David-Paul Pertaub. We like to thank David Swapp, the manager of the ReaCTor system for his help throughout the project. We would also like to extend our gratitude to Amanda Oldroyd of BT plc for supplying the cartoon form characters and Joel Jordan for his invaluable help in creating the street. We would also like to thank Ashwin Beeharee and Angus Antley for their help in running the study.

10. References

Sense of Presence in Emotional Virtual Environments

Rosa Baños (1), Cristina Botella (2), Víctor Liaño (2), Belén Guerrero (2), Beatriz Rey (3), Mariano Alcañiz (3)

(1) Universidad de Valencia (Spain)
(2) Universitat Jaume I de Castellón (Spain)
(3) Medical Image Computing Laboratory, Universidad Politécnica de Valencia (Spain)
rosa.banos@uv.es, botella@psb.uji.es, {beareyso | malcaniz}@mediclab.upv.es,

Abstract

Most definitions on presence have been based on cognitive or environmental aspects. However, we think that presence, like all human experiences, is influenced by emotions. EMMA project (IST-2001-39192) is aimed to study the nature of this relationship between emotions and presence. One of the main hypotheses proposed by EMMA is that emotions may enhance presence. In this line, the main objective of the present paper is to study the differences in presence between “emotional” environments and “neutral” environments. In order to achieve this objective, we have designed a Mood Induction Procedure using VR (VR-MIP, to induce different moods (sadness, joy, anxiety and relax) in experimental subjects. Our results point out that VR-MIPs are able to induce different moods in the users. Regarding the role of mood on the sense of presence, our results show differences between emotional and neutral environments in some presence measurements.

Keywords--- sense of presence, mood induction, emotions.

1. Introduction

One of the main characteristics of Virtual Reality (VR) is that allows the users to explore and to interact with an artificially created environment. Thus, the users “have the feeling” of being in a different place from where they physically are, and this feeling has been named as “sense of presence”. Up to now, definitions on presence have been based on cognitive or environmental aspects. However, we think that presence, like all human experiences, is influenced by emotions. Emotions play an important role in the way we make our subjective judgments, we react to the world and we learn things about it. It has been demonstrated that this variable is especially important in order to generate and enhance presence in Mental Health applications of VR, e.g., [1] [2] [3] [4] [5] [6]. In this sense, Hoorn, Konijn & Vand de Veer [7], in a paper entitled “Virtual Reality: Do not augment realism, augment relevance” argue that VR experience gains more from increased emotional relevance than from higher realistic solutions. These authors claim that to design VR, experience instead of technology is the key word, and they recommend that VR designer focuses on developing features that sustain relevance to the goals and concerns of the user. According to them, “The sophisticated technology of VR may be powerful but it is not enough to initiate a reality-experience that is true-to-life. Basic to reality-experiences that are true-to life is that the experience is emotionally loaded (…). The basis of emotion psychology is personal meaning: without relevance no emotion occurs. Thus, VR needs personal relevance for the user to arrive at the intended (total) involvement as manifested in the experiences of immersion and presence”.

Taking into account this line of thinking, we think that emotions may play a role on the sense of presence. From this point of view, the focus would lay on designing affectively significant environments. In order to achieve this, it would be necessary to include elements with the potential of activating emotions. This is especially true for Clinical Psychology, because the goal is to achieve important changes in the users. But this relationship between emotion and presence is not only of interest for Clinical Psychology, but also for the whole research on presence. So, many authors point out that a central question in this research is the role of presence in the satisfaction with, and the enjoyment of mediate experiences. Furthermore, this relationship could help researchers to define the parameters for the minimum requirements-relevance of presence and to generate alternative methods to define and measure the sense of presence.

In short, presence is determined by a complex variety of characteristics of the medium, the user and the context. Regarding users, emotions can play an important role in the sense of presence, and EMMA project (IST-2001-39192) is aim is to study the nature of this relationship between
emotions and presence. One of the main hypotheses proposed by EMMA is that emotions may enhance presence. It is more likely that an environment able to elicit anxiety, sadness, joy, etc. could provoke the user to feel more present in that environment. Therefore, it could even be possible to consider emotions as an indicator of the degree of presence. In this line, the main objective of the present paper is to study the differences in presence between "emotional" environments and "neutral" environments. In order to achieve this objective, we have designed a Mood Induction Procedure using VR (VR-MIP), to induce different moods (sadness, joy, anxiety and relax) in experimental subjects.

The specific hypotheses that have been tested are: 1) VR-MIPs will be able to induce moods in the users; and 2) Sense of presence will be greater in the "emotional" environments than in the "neutral" environment

2. Method

2.1. Participants

Sample was composed of 80 university students volunteers (30 men and 50 women). Their age range was from 18 to 49. Before the experiments, all participants were screened, using an interview and several questionnaires, in order to detect any of the following exclusion criteria: history of neurological disease, head injury, learning disability or mental disorder; history of psychological disorders; use of any medication for psychological or emotional problem; scoring 18 or higher in BDI (Beck Inventory Depression, [8]).

2.2. Measures

- Visual Analogue Scale (VAS): A variation of the original measure [9] has been used. Participants were asked to rate in a 1-7 points Likert Scale (1 = Not feeling the emotion at all, 7= Feeling the emotion extremely), how they felt at that moment in every one of the following emotions: sadness, joy, anxiety, and relaxation.

- The ITC-Sense of Presence Inventory (ITC-SOPI) [10]. This subjective measure is divided in two parts. Part A is composed by 6 items and it is referred to the impressions of the user after the virtual experience has finished. Part B is composed by 38 items and it is referred to the impressions of the user during the virtual experience. A 1-5 points Likert scale is used in both parts. Factor analysis of this 44 item questionnaire showed it measured the following components/dimensions of presence: physical space; engagement; ecological Validity, and negative effects

- Reality Judgement and Presence Questionnaire (RJPJ) [1] This questionnaire is composed by 57 items. A 1-10 Likert scale is used to answer all items. The following factors were taken into account: Quality/Realism, Reality Judgment; Presence: Positive issues; Presence: Negative issues; Interaction/Navigation, Emotional engagement; Emotional Indifference.

2.3. Experimental conditions.

Participants were randomly assigned to one of the five virtual environments (VR-MIPs): sad, happy, anxious, relax and neutral. “Neutral” condition is included as a control condition where no mood changes are expected.

2.3. Virtual Environment (VR-Mood Induction Procedures)

Virtual environment consists of a neutral environment that will progressively change depending on the mood state that we need to evoke on the user. The chosen scenery was a park, because it is an environment where nature elements are present (trees, flowers, water, etc...), and by changing some light parameters (tone, direction, brightness) it is possible to modify the aspect of these elements achieving a set of different moods on the user.

The virtual experience starts with a narrative. The user listens to a woman voice (representation of the EMMA avatar) that gives an introduction of the virtual environment. Then, the user starts walking through a park. The initial appearance of the environment is the same for all users. However, the aspect changes soon depending on the emotion. For example, in the case of sadness, the park is grey, it is a cloudy day, the trees have no leaves, there are no people in the park and the music that is heard is a very sad piece. After two minutes walking through the park, the user finds a band stand. He/she can find five statements in the lower side of five of the sides of the stand (it is an eight-faced polyhedron). They are statements from the Velten technique [11], and they change depending on the emotion condition. EMMA asks the user to think about personal meaning of the state for some time in every statement. At the same time, an image is shown over the sentence. After that, the user returns to the park and continues walking, until he/she finds a summer cinema. EMMA voice invites the user to see a movie. These movies are film scenes of short duration with sad, happy, anxious, neutral and relaxing contents. Finally, EMMA asks the user to think about a similar personal situation (autobiographical recall) and to explain the experience/situation in loud voice (if he/she agrees to). Then, EMMA congratulates the user for his/her experience.

2.4. Hardware devices

As open immersive display we have used a metacrilate retro-projected screen of 400x150cm. The retro-projection option allowed the user to walk near the screen with-out blocking the image and projecting shadows on the screen. Projectors had a resolution of 1024x768 pixels and a power of 2000 lumens. However, we have regulated them for a power of 1000 lumens in order to not be molest for the user.

Regarding interaction device, a Wireless Pad was used: This device has two special features; it has no wires so the
was the fact that the higher order interactions of group x other main effects of group were found. More important groups who scored lower compared to the other groups. No followed by the neutral group, being the sad and anxiety happy and relax groups scored higher on joy mood, moment were statistically significant for joy mood (F (1,4) 6,212 p< .001), sad mood (F (1,4) 5,854 p< .001) and relax mood (F (1,4) 3,849 p< .009). Post hoc analyses results showed that the “sad”, “happy” and “relaxation” emotional environment induced more relax. Furthermore, there were not differences in the neutral condition. Regarding the “anxiety” environment, it was the only condition where participants scored higher on anxiety mood at the end of the trial, but this result did not reach statistical significance.

2.5. Procedure

The experimental session started with participants filling in the pre-induction VAS measure. Then, participants entered into a room where the virtual system was placed. The user practised, with the help of an experimenter, in a brief training environment how to move and how to interact with virtual objects. After that, they were stayed alone in the room and the virtual session started. It took 30 minutes to complete the virtual walk. After that subjects filled in emotion (VAS) and presence (ITC-SOPI and PRJQ) measures

2.6. Results

2.6.1. Mood Induction: Means and standard deviations can be found in Table 1. Analysis of variance was conducted on the mood measures, with emotional groups (sad, happy, anxious, relax and neutral) as between-group factor, and time of testing as the within-group factor (before vs. after). The dependent variables were the mood state measures (VAS).

A main effect of time of testing was found for sadness mood (F (1,45) 14,545 p< .001), No other main effects of time were found. A main effect of group was found for joy mood (F (4,45) 2,742 p< .040). Post-hoc tests revealed that both happy and relax groups scored higher on joy mood, followed by the neutral group, being the sad and anxiety groups who scored lower compared to the other groups. No other main effects of group were found. More important was the fact that the higher order interactions of group x moment were statistically significant for joy mood (F (1,4) 6,212 p< .001), sad mood (F (1,4) 5,854 p< .001) and relax mood (F (1,4) 3,849 p< .009). Post hoc analyses results showed that the “sad”, “happy” and “relaxation” emotional environments had induced the mood in the predicted directions for every single environment. That means that the sad environment induced more sadness, the happy environment induced more joy, and the relaxation environment induced more relax. Furthermore, there were not differences in the neutral condition. Regarding the “anxiety” environment, it was the only condition where participants scored higher on anxiety mood at the end of the trial, but this result did not reach statistical significance.

2.6.1. Sense of Presence: Means and standard deviations can be found in Table 2. Analysis of variance was conducted on the presence measures, with groups (emotional versus neutral) as between-groups factor. The dependent variables were the presence measures (ITC-SOPI, RJPO). As it can be seen in Table 2, ANOVAs showed a group effect for Engagement and Ecological Validity from ITC-SOPI, and most of PRJQ factors.

Table 1.: VAS scores (before and after the trial): Means and (standard deviations) ratings

<table>
<thead>
<tr>
<th>Group</th>
<th>Joy before</th>
<th>Joy after</th>
<th>Sadness before</th>
<th>Sadness after</th>
<th>Anxiety before</th>
<th>Anxiety after</th>
<th>Relaxation before</th>
<th>Relaxation after</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sad group</td>
<td>(0.7)</td>
<td>(1.2)</td>
<td>(0.5)</td>
<td>(1.6)</td>
<td>(2.3)</td>
<td>(1.8)</td>
<td>(1.2)</td>
<td>(1.9)</td>
</tr>
<tr>
<td>Happy group</td>
<td>(4.7)</td>
<td>(5.5)</td>
<td>(2.4)</td>
<td>(2.2)</td>
<td>(2.8)</td>
<td>(2.1)</td>
<td>(4.5)</td>
<td>(5.0)</td>
</tr>
<tr>
<td>Anxious group</td>
<td>(0.9)</td>
<td>(1.7)</td>
<td>(1.0)</td>
<td>(1.8)</td>
<td>(0.5)</td>
<td>(1.9)</td>
<td>(0.9)</td>
<td>(0.9)</td>
</tr>
<tr>
<td>Relax group</td>
<td>(4.7)</td>
<td>(4.7)</td>
<td>(2.2)</td>
<td>(3.0)</td>
<td>(2.2)</td>
<td>(1.5)</td>
<td>(0.9)</td>
<td>(0.9)</td>
</tr>
<tr>
<td>Neutral group</td>
<td>(0.7)</td>
<td>(0.9)</td>
<td>(1.5)</td>
<td>(1.2)</td>
<td>(0.9)</td>
<td>(1.5)</td>
<td>(0.5)</td>
<td>(0.9)</td>
</tr>
</tbody>
</table>

Table 2: ITC-SOPI and RJPO factors scores: Means (and standard deviations) and significance ratings

<table>
<thead>
<tr>
<th>Group</th>
<th>Emotional group</th>
<th>Neutral group</th>
<th>Group differences</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITC-SOPI</td>
<td>Spatial presence</td>
<td>3,12 (0,69)</td>
<td>2,90 (0,96)</td>
</tr>
<tr>
<td></td>
<td>Engagement</td>
<td>3,47 (0,57)</td>
<td>3,19 (0,45)</td>
</tr>
<tr>
<td></td>
<td>Ecological validity</td>
<td>3,41 (0,74)</td>
<td>2,99 (0,77)</td>
</tr>
<tr>
<td></td>
<td>Negative effects</td>
<td>1,73 (0,76)</td>
<td>1,67 (0,73)</td>
</tr>
<tr>
<td>RJPO</td>
<td>Quality /realism</td>
<td>7,48 (1,19)</td>
<td>6,73 (1,46)</td>
</tr>
<tr>
<td></td>
<td>Reality /judgment</td>
<td>6,12 (1,96)</td>
<td>5,06 (2,14)</td>
</tr>
<tr>
<td></td>
<td>Presence (positive )</td>
<td>6,64 (1,68)</td>
<td>5,86 (1,36)</td>
</tr>
<tr>
<td></td>
<td>Presence (negative)</td>
<td>3,43 (1,96)</td>
<td>3,63 (1,83)</td>
</tr>
<tr>
<td></td>
<td>Interaction/ navigation</td>
<td>7,28 (1,48)</td>
<td>6,91 (1,32)</td>
</tr>
<tr>
<td></td>
<td>Emotional engagement</td>
<td>8,13 (1,87)</td>
<td>5,85 (2,23)</td>
</tr>
<tr>
<td></td>
<td>Emotional indifference</td>
<td>1,47 (1,96)</td>
<td>3,60 (1,45)</td>
</tr>
</tbody>
</table>

158
2.7 Discussion

Regarding the hypothesis, our results point out that VR-MIPs are able to induce different moods in the users. Data from subjective mood state measurements show that the 4 emotional environments (sad, joy, anxiety, and relax) are able to produce mood changes in the users. The changes in mood were produced in the predicted directions for every single environment. As for the neutral environment no changes are observed. Results from VAS measurements are robust for sadness, happy, and relaxation inductions, but are less conclusive for anxious induction. The anxiety environment was not designed to induce fear, but anxiety. Fear is an emotional response provoked by a “present” and specific threat, while anxiety is a more diffuse emotion. In general, the core element that is called on to distinguish between fear and anxiety is the presence of identifiable cues. Anxiety usually is defined as an emotional distress characterised by worry and tension, whereas threat is not so easy to define. Using Barlow’s [12] words, anxiety is “a diffuse, objectless apprehension” (page, 7). In this sense, our “anxiety environment” include elements about a “possible” danger, trying to provoke an apprehensive and negative expectation. Results indicate that it is more difficult to induce this mood than other negative (sad) or positive (joy, relax) moods. So this environment should be improved, and perhaps it might be useful to design a “fearful” environment in order to test possible differences in the capability to induce “fear” and “anxiety”

These results are important since they reveal the utility of the VR-MIPs as mood devices and, therefore, the possibility of using them in the future from both an applied and experimental perspective. For example, they could be used as a therapeutic tool to induce specific moods (relaxation or joy) in people who need it, or as mood induction procedures in experimental psychopathology. Most important is the fact of having achieved a “neutral” mood induction procedure that permits to make comparisons with the motional mood devices in latter works.

Regarding the role of mood on the sense of presence, our results show differences between emotional and neutral environments in some presence measurements. Regarding ITC-SOPI, there were differences between groups in “Engagement” and “Ecological validity” scales, indicating that emotional environments seemed more natural, believable and real to subjects than the neutral environment. Furthermore, regarding PRJQ, statistically significant differences are observed between the emotional and the neutral environment in most of the factors. The participants report a higher degree of realism, they also judge the experience as more real, they feel more present, and they experience a higher emotional involvement.

In conclusion, our results point out that the “new technologies” are useful to create mood devices that are able to induce emotions in the users; it is possible to induce “specific emotions” since the different mood devices that have been created allow the induction of an specific “target” emotion. Furthermore, the sense of presence is related to the emotion. From these results it is possible to argue that emotions may enhance presence. It is more likely that an environment able to elicit anxiety, sadness, joy, etc. could provoke the user to feel more present in that environment. Therefore, it could even be possible to consider emotions as an indicator of the degree of presence. It seems that sense of presence is determined by the emotions that a virtual environment is able to provoke in the user, and emotional issues are important variables in order to enhance presence. Further investigation is however needed, especially as far as objective measures are concerned.

References

Navigating Virtual Reality by Thought: First Steps

Doron Friedman¹, Robert Leeb², Angus Antley¹, Maia Garau¹, Christoph Guger³, Claudia Keinrath², Anthony Steed¹, Gert Pfurtscheller², Mel Slater¹

Virtual Environments and Computer Graphics Lab, Department of Computer Science, University College London, UK¹
Institute of Human-Computer Interfaces, Graz University of Technology, Austria²
g.tec - Guger Technologies OEG, Graz, Austria³

{d.friedman@cs.ucl.ac.uk, robert.leeb@tugraz.at, a.antley@cs.ucl.ac.uk, m.garau@cs.ucl.ac.uk, guger@gtec.at, , claudia.keinrath@tugraz.at, A.Steed@cs.ucl.ac.uk, pfurtscheller@tugraz.at, m.slater@cs.ucl.ac.uk }
2. Background

Previous research has established that a BCI may be used to control events within a VE, and some research has also been done in immersive systems. Nelson et al. [8] were interested in BCI as a potential means for increasing the effectiveness of future tactical airborne crew stations. They have investigated the usage of CyberLink™, an interface that uses a combination of EEG and electromyography (EMG) biopotentials as control inputs. They were interested in single-axis continuous control. The participants used the CyberLink interface to navigate along a predetermined flight course that was projected onto a 40-foot diameter dome display. Continuous feedback was provided by a graphical heads-up display (HUD). Participants were not given any BCI instructions or training. Scores of effective task performance gradually increased with training and reached an average of 80% success.

Middendorf et al. [7] harnessed the steady-state visual-evoked response (SSVER) as a communication medium for the BCI. Two methods were employed; one of them was tested with a flight simulator. In this method operators were trained to exert voluntary control over the strength of their SSVER. One of the conditions involved controlling a flight simulator, where the roll position of the flight simulator was controlled with BCI. The simulator rolled right if 75% or more of the SSVER amplitude samples over a half second period were higher than some threshold, and left if most of the samples were lower than another threshold. Most operators were able to reach 85-90% of success after 30 minutes of training.

Bayliss and Ballard [1] used the P3 evoked potential (EP), a positive waveform occurring approximately 300-450 ms after an infrequent task-relevant stimulus. They used a head-mounted display (HMD)-based VR system. Subjects were instructed to drive within a virtual town and stop at red lights while ignoring both green and yellow lights. The red lights were made to be rare enough to make the P3 EP usable. The subjects were driving a modified go-cart. Whenever a red light was displayed, data was recorded continuously from -100 to 1000 ms. Results show that a P3 EP indeed occurs at red and not yellow lights, with recognition rates that make it a candidate BCI communication medium.

In further research Bayliss [2] continued exploring the usage of the P3 EP in VR. Subjects were asked to control several objects or commands in a virtual apartment: a lamp, a stereo system, a television set, a Hi command, and a Bye command, in several conditions, including an HMD. Using BCI, subjects could switch the objects on and off or cause the animated character to appear or disappear. The BCI worked as follows: approximately once per second a semi-transparent sphere would appear on a randomly selected object, for 250ms. Subjects were asked to count the flashes on a specific object (to make the stimulus task-related, as P3 requires). An epoch size from -100ms (before the stimulus) to 1500ms was specified. Text instructions in the bottom of the screen indicated the goal object. The subject had to count the flashes for that object only. The subject was given a visual feedback when a goal was achieved, i.e., when a P3 event was recorded when the target object was flashing. Subjects were able to achieve approximately 3 goals per minute. Bayliss found no significant difference in BCI performance between VR and a computer monitor. Most subjects preferred the VR environment; all of them liked the fixed-display condition (looking through a fixed HMD) the least.

Lalor et al. [11] used the SSVER as a control mechanism for a 3D game; players had to intervene when a character walking on a thin rope lost balance, by looking at checkerboard images on two sides of the animated image. The setting did not include VR, but the game was played in front of a large screen rather than a monitor, so the experience was immersive to some degree. They report robust BCI control, with best accuracy of 89%, and attribute relative success to motivation.

Leeb et al. [6] used the Graz-BCI paradigm based on motor imagery, in a VR rotation task. They have used a fixed-display HMD setting, which means the subjects actually experienced a limited form of VR. They report BCI performance success rates of 77-100%. This was the first step in the research reported here.

3. The Experiments

3.1 The System

In order to carry out the “navigation by thought” experiments we had to integrate two complex software and hardware systems: the BCI and the Cave-like VR system. A system diagram appears in Figure 1 and is explained below.

The experiments were carried out in UCL’s four-sided Reactor system, which is driven by an Onyx IR2 with four graphics pipes. Users were head tracked using a wireless tracker. The applications were implemented on top of the DIVE software\footnote{1 http://www.sics.se/dive} [4,15].

The Graz-BCI system consists of a biosignal amplifier (g.tec, Graz, Austria), a data acquisition card (National Instruments, Austin, USA) and a standard PC running Windows 2000. The signal processing is based on MATLAB and Simulink (Mathworks, Inc., Natick, USA).

The communication between the PC running the BCI and the VR host is done using a communication system called Virtual Reality Peripheral Network (VRPN).\footnote{2 http://www.cs.unc.edu/Research/vrpn/} VRPN provides synchronization and logging of multiple data channels, and has built-in support for many VR devices. On the PC it communicated with the BCI Matlab-based software via a Dynamic-Link Library (DLL). On the Onyx machine controlling the VR, a VRPN plug-in for DIVE was implemented. VRPN uses UDP to establish connection and TCP for sending messages over the network. The main challenge was to balance the CPU time between the communication and the rest of the system: the BCI on the

1 http://www.sics.se/dive
2 http://www.cs.unc.edu/Research/vrpn/
PC, and DIVE on the Unix machine. We were interested in receiving 20 updates per second with a minimum delay time, so that the feedback for the BCI decisions would be as smooth as possible.

Figure 1: A diagram of the BCI-Cave integrated system.

3.2 Brain-Computer Interface

The experiment included three subjects: one female and two males. All subjects were familiar with the Graz-BCI [10] over a period between four months and two years. In addition, they were specifically trained for this experiment by performing identical tasks in similar VEs with an HMD.1 Two different experiments had been performed. One was rotating to the left or right inside a virtual bar by imaging a right or left hand movement and the second experiment was walking forward along a virtual street by imaging a foot movement.

In all experiments the subject was sitting on a comfortable chair in the middle of the Cave (see figure 2). Three bipolar EEG channels (electrodes located 2.5 cm anterior and posterior to C3, Cz and C4, respectively, according to the international 10-20 system) were recorded with a sampling frequency of 250 Hz. The logarithmic band power was calculated sample-by-sample for 1-sec epochs in the alpha (10-12 Hz) and beta (16-24 Hz) bands of the ongoing EEG and classified by a linear discriminant analysis (LDA). The LDA classification result was used as a binary control signal and sent via the VRPN to the Cave system to modify the position inside the VE. The LDA classifier used in these experiments were calculated offline from data recorded previously in Graz using similar VEs in an HMD experiment [6].

Each subject participated in two sessions on two consecutive days and each session included four feedback runs. Each run consisted of 40 trials (20 left and 20 right cues, in the case of the rotation experiment and 20 foot and 20 right-hand cues for the walking experiment, respectively) and lasted about seven minutes. The sequence of right/left or foot/right cues was randomized through each run. Depending on the affiliation of the acoustic cue, the subject was instructed to imagine a left or right hand movement in the bar-rotation experiment or to imagine a foot or right-hand movement in the street-walking experiment.

3.3 Experimental Setup

BCI can be realized in an externally-paced mode (synchronous BCI) or in an internally-paced mode (asynchronous BCI). In the former case, specific mental patterns have to be generated in response to an external event, i.e., changes in brain activity are tracked over a predefined time window. In the later one the EEG has to be analyzed continuously. We have used a synchronous BCI which is more limited, but more reliable. This has a great disadvantage compared with traditional VR navigation devices.

Classification of the signal can start immediately after the trigger, but the optimal classification point varies between individuals, and is typically at least two seconds after the trigger [5]. This delay is too long for a user interface feedback, thus we prefer to provide visual feedback immediately after the BCI decision. Thus, the system sent the classification result every 52 milliseconds approximately, over a period of 4.16 seconds after the trigger, and continuous feedback was provided.

The course of events is as follows: the application decides that a navigation decision is required, and DIVE sends a request over the VRPN network. This request, along with all other network events, is logged with an accurate timestamp for post-hoc analysis. The request is intercepted by the VRPN component running on the BCI PC. It communicates with the DLL, which then sends a request to the Matlab BCI software. The BCI software makes a random decision about the navigation decision required by the user, and initiates an auditory trigger accordingly. The BCI software immediately starts analyzing the EEG signals, and makes a classification decision approximately every 52 milliseconds. A binary value is passed back to the DLL, and then over the VRPN network. The DIVE VRPN plug-in, on the Onyx host, intercepts this event and feeds it into the application. The application changes the VE to reflect the users’ rotation or translation.

We selected two VEs for simple navigation tasks. The first task was rotation: the VE depicts a virtual bar (see figure 2). It is populated by four virtual characters and a virtual barman. Originally, the virtual characters talked to

---

1 The VEs were not identical: in the training phase the subjects had to rotate an auditorium, whereas in the CAVE experiment the subjects had to rotate a virtual bar. In the walking task we used two different models of streets.
the users when their head gaze and the user’s gaze crossed (this is possible since the user is head tracked), and the bar included background music and chatter. Eventually we decided to remove the background music and the speech audio, because the BCI triggers were auditory, and we wanted the subjects to hear them clearly. We left the background chatter, and that was the first time these BCI subjects were exposed to an experiment with an audio track, other than the trigger signals.

The size of the virtual bar is not much larger than the size of the physical projection room, and there is no need for the user to navigate around the VE; it is enough for them to look around. This environment is, thus, suitable for a simple first experiment, in which users operate in the VE by rotation only.

The BCI sent 80 classification results over a period of 4.16 seconds, in fixed intervals. These results were used by the VR to provide continuous feedback; the total rotation per trigger could be up to 45 degrees, and for each BCI update the VE rotated by 0.56 degrees.

The BCI control was as follows: if the trigger indicated that the user could walk, the user had to imagine foot movement to move forward. If the trigger indicated that the user should stop, they had to imagine right hand movement. If the classification indicated hand imagery when ordered to walk, the user would stay in place. If the classification indicated foot imagery when ordered to stop, the user would go backwards. This punishment was introduced so that the subjects will not be able to adopt a strategy of always imagining their foot. The same timing as in the rotation experiment was used, but instead of rotating, the BCI updates moved the subject by 0.05 distance units, so the total distance per trigger could be up to four units.

4. Results

In this paper we will only report the overall BCI performance results; details will be reported in a separate paper. Here we are interested in evaluating the participants’ experiences and the implications for presence.

The three subjects (L1, O3, S1) participated in this study were able to rotate the bar with a relatively high level of success: the average performed rotation of one run to the left was -21° and to the right 22°; resulting in a BCI performance from 80 to 100%. Two of the three were also able to navigate the street. One of the three subjects (S1) was not able to stop properly, using hand motor imagery, and was thus replaced with subject J8.

The cumulative distance travelled by the subjects in the street scenario can be used as a performance measurement of the experiment. The maximum achievable mileage would be 80 distance units and the result of a random session would be 0 units. This cumulative performance measure is different than the BCI performance: BCI performance can be assessed taking into account only the optimal classification points in time, whereas mileage takes into account classification throughout the whole 4.16 seconds epoch.
Figure 4 compares the performance for each session of the three different conditions for an arbitrary subject (O3). The results performed in the Cave are displayed in the right columns, the performance achieved with the HMD in Graz and the performance of the standard BCI without using the VR as a feedback medium are displayed to the left. The mileage for the standard BCI (left columns), were simulated offline to be able to compare the results. The total mileage travelled over all Cave sessions was 55.4 units, in the HMD it was 43.3 units, and using the standard BCI 34.4 units. The BCI performance in the Cave was between 68 and 100% classification accuracy. We can see that the Cave performance was statistically significantly better (F2,21=22.08; p<0.001), compared to the other conditions.

After completing a session, the subjects were asked to fill in the Slater-Usoh-Steed presence questionnaire [12], and then a non-structured interview was conducted. This was carried out with three subjects after the bar VE and with two subjects after the street VE. Based on questionnaire and interview data, we can evaluate several aspects of presence: overall sense of presence, body representation, and social presence. In addition, we can look at other interactions between BCI and VR.

Constraints posed by the BCI make it difficult to evaluate the sense of presence, as well as the overall experience: subjects are BCI trained for long durations (of four months up to a two years), and they were specifically trained for the same task. The subjects were very limited in their actions: they are trained not to move or blink during the BCI feedback epochs. Since our experiment included four seconds of BCI feedback followed by approximately five seconds of rest until the next trigger, they were asked not to move throughout the whole experience.

4.1 Overall Presence

Appendix A provides the details for five different questions from the questionnaire that relate to overall presence, and that were applicable to the BCI experiments. The number of subjects is too small to allow statistical analysis, but the results are consistent: there is a difference among subjects: L1 reported high presence, S1 medium presence and O3 low presence. There seems to be no difference between tasks.

One subject commented that in the beginning he was very focused on the BCI task, so he could not pay enough attention to the VE to feel present in it. Since gradually the BCI control became more automatic, he could be absorbed by the VR and feel more present. This description is different from what most subjects who experience the same VEs without BCI report: initially they feel a high sense of presence, but this gradually drops as they realize the limitations of the VR.

All subjects report that the Cave is more comfortable than the HMD. One subject reported that the wide field-of-view made him feel as if the landmarks were all around him, more like in the real world than in a typical BCI session; this may imply that spatial presence may facilitate BCI. The subject mentioned he did not experience this with the HMD setting.

4.2 Body Representation

Previous work has suggested the very important role of perception of the body within virtual reality [14]. The more that the body is used naturally, and the more that it is anchored into the VE the greater the chance for presence. This research referred to a virtual body representation. However, BCI may be considered a (very unusual) extension of the body; it is thus interesting to learn about the subjects’ sensation of their own body in the experience. Subjects reported that they felt natural compared to other BCI experiments. One subject noted that she felt as if her whole body was rotating. Another one, asked whether he felt as if his body was actually rotating, answered: “No, it was more like in a dream – you move but you do not feel your body physically move. And just like in a dream – at that moment it seems real”. The interaction seemed more natural than traditional BCI to all subjects, even if the mapping between the motor imagery and the application functionality was not perfect; all of them reported it seemed quite natural to use their hand and feet imagery as means of rotating or moving, and that this is very different from controlling a bar on a monitor, which is the typical BCI training setting.

4.3 Social Presence

The two VEs included animated characters. It was thus possible to evaluate the subject’s sense of social presence, or to what extent the subjects felt as if they were in a socially populated VE. Appendix B details a subset of the questionnaire questions related to social presence; however, unlike the overall presence questions, the results do not seem to be consistent even within subjects.

Subjects did not generally report a high level of social presence, though they certainly paid attention to the virtual characters. One subject commented that the characters in the bar were used as landmarks for the rotation, which made them be treated as inanimate objects. Another one, when asked if he felt the characters were real, replied: “Yes, but not exactly. It was as if I am a space explorer who...
just met some aliens. They look humanoid, but they behave different; as if they were some other life forms”.

4.4 Other Interactions of BCI and VR

The highly immersive (Cave) condition achieved the highest level of BCI accuracy, compared to less immersive conditions. Further research is needed in order to understand why this happens: is this because participants are more motivated in highly immersive VR, or are there additional factors? Does presence play a role here?

Subjects report a conflict between what they wanted to do during the experience and the BCI task for which they were trained. One subject wanted to reach his hands for the beer or talk to the characters in the room; he said: “It was like a little voice in my head saying “try this, try this,...”, but I know I am not allowed to.”

Subjects noted that the bar room had two areas: the virtual characters concentrated in one area, whereas the other side of the room was empty, and didn’t even contain furniture (only a disco-style chandelier). One subject said BCI control was easier for her in that area, because it was less distracting. For another subject BCI control seemed more difficult in the empty space, because there was no clear spatial information.

The audio chatter was difficult for one subject. It was distracting, but also annoying because of repetition. He noted that the visuals repeated too, but this was not annoying or distracting. The difference could be due to the fact that the audio was a repeated loop (its duration was one minute), whereas the character’s gestures are pseudo random. Also, the subjects reported that part of the audio track included a very distinct laughter, which stood out when repeating itself.

5. Discussion and Future Work

Subjects performed best in the Cave. Moreover, all subjects liked the Cave setting more than the HMD, and both were very much preferred over BCI training on a monitor. The main reason given for preferring the VR was that they provided motivation. Specifically, the street VE was treated as a sort of race course: subjects wanted to get as further away in the street, and further than other subjects in previous sessions. An interesting comment was made by one of the subjects: motivation seems to greatly improve BCI performance, but too much excitement might be distracting. It may be interesting to explore this in follow-up research.

Previous research has demonstrated that BCI-based navigation of VR is possible, and we were able to repeat this result in the Cave-like setting. In addition to the feasibility of BCI control in the Cave, the main lesson learned is that typical BCI procedures are, at the moment, too prohibitive to become a natural user interface. Given that VR seems to be a promising counterpart to BCI, we may ask: what needs to be done in order to make BCI control of VE or VR a positive experience for a wide range of participants?

Thus far we have not been able to demonstrate a realistic scenario of navigating a VR by thought; only subsets of the navigation task were demonstrated. Even in these simplified tasks, we had to remove substantial parts of the VE functionality to make it easier for subjects to concentrate. Both VEs did include significant visual input, and the bar included a background audio track as well. Future research will need to go beyond adapting the task to the constraints posed by the BCI, and assess BCI performance in a wide range of rich VEs with various types of tasks and interactions.

Even in the scope of these limited tasks, there are several issues that come up. First, we note that for the classification to reach a high level of accuracy, the subjects need to be trained over many sessions, typically over a few days or weeks. Even then, the accuracy is seldom 100% [9], as is expected from traditional devices. In our case, note that subjects were not only highly trained to use BCI in general, but were also extensively trained for the specific task!

BCI poses physical constraints on the subject that might make it unacceptable for some applications. Subjects need to sit still, and during the BCI control epochs are trained to stop blinking, or moving their muscles.

Another problem with BCI is that the reliable methods are typically trigger-based (synchronous). This has a great disadvantage compared with traditional navigation devices. The best we can do in synchronous paradigms is to seamlessly incorporate the BCI cues into the environment.

Further difficulties are specific to the motor imagery paradigm, and may be avoided using other BCI techniques: using the motor imagery paradigm, the classification is optimal only a few seconds after the cue. As previously explained, such delay is not acceptable for a UI, so we introduce continuous feedback. This has a penalty: the VE functionality does not depend on the optimal BCI decision. A method taking both feedback and accuracy may be possible.

As a conclusion from this experiment, our aim is to continue this research towards free-form BCI-controlled navigation, in which the participants will be free to traverse the VE rather than required to follow cues, based on a synchronous grid. Interestingly, such free-form control of BCI in VR was never attempted, since researchers are typically only interested in measuring BCI accuracy. In addition, we want to try alleviating some of the constraints posed by the BCI, and try to make BCI a more natural interface for navigation. We could still be able to evaluate BCI performance by looking at overall task performance, where the task depends on navigating the VE. It is possible to compare the results of navigating with BCI to navigation with a traditional VR input device, such as a wand. This could be interesting especially for those people who have difficulty in navigation using a wand. However, we are not only interested in the effectiveness of the interface; it would be very interesting to compare the participant experience in these two conditions as well.
The mapping between the recognized thought–related EEG patterns and the VR functionality needs to be further explored. We hope to learn whether a more natural mapping improves BCI accuracy or learning rates, and to what extent can users adapt to counter-intuitive mappings (e.g., moving to the right when thinking about the left hand).

We conclude that while we have shown that BCI can be used as an interface for navigating VR in a Cave-like setting, we are still far from being able to use the brain as a natural interface for this, which will be able to compete with traditional interface devices in accuracy and level of comfort. In order to reach such a goal, more research is needed that concentrates not only on BCI accuracy, but also on the overall participant experience.

Acknowledgements

This project is funded by the European Union project PRESENCIA, IST-2001-37927. We would like to thank Angus Antley, Reinhold Scherer, Anthony Steed, and David Swapp. We would like to thanks Joel Jordan and Vinoba Vinayagamoorthi for the virtual street VE. We would also like to thank the participants (Jörg Biedermann, Gernot Supp and Doris Zimmermann) for their cooperation and patience throughout these long and difficult experiments.

References


Appendix A: Subset of Presence Questionnaire: Overall Presence

A subset of questions related to overall presence from the SUT presence questionnaire [12]. For all questions below the subjects were given choices from 1 to 7, and the answers were normalized such that higher scores indicate a higher sense of presence.

<table>
<thead>
<tr>
<th>Question</th>
<th>L1 bar</th>
<th>L1 street</th>
<th>o3 bar</th>
<th>o3 street</th>
<th>S1 bar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Please rate your sense of being in the room, on the following scale</td>
<td>5</td>
<td>6</td>
<td>3</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>from 1 to 7, where 7 represents your normal experience of being in a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>place</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>To what extent were there times during the experience when the room</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>was the reality for you and you almost forgot about the real world</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>of the laboratory where the experience was really taking place?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>When you think back about your experience, do you think of the room</td>
<td>3</td>
<td>6</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>more as images you saw, or more as somewhere that you visited?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(visited = 7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>During the course of the experience, which was strongest on the whole,</td>
<td>6</td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>your sense of being in the bar, or of being in the real world of the</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>laboratory? (bar = 7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>During the time of the experience, did you often think to yourself</td>
<td>5</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>that you were just standing in a laboratory or did the bar overwhelm</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>you? (bar = 7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Appendix B: Subset of Presence Questionnaire: Social Presence

The table below is an excerpt related to social presence from the SUT presence questionnaire [12] for three subjects on two tasks. For all questions below the subjects were given choices from 1 to 7, and the answers were normalized such that higher scores indicate a higher sense of presence.

<table>
<thead>
<tr>
<th>Question</th>
<th>L1 bar</th>
<th>L1 street</th>
<th>o3 bar</th>
<th>o3 street</th>
<th>S1 bar</th>
</tr>
</thead>
<tbody>
<tr>
<td>During the course of the experience, did you have a sense that you were</td>
<td>5</td>
<td>4</td>
<td>7</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>in the room with other people or did you have a sense of being alone?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>How aware were you of the characters in the room?</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>How curious were you about the characters?</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>When you first saw the characters, to what extent did you respond to</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>them as if they were real people?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Now consider your response over the course of the whole experience.</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>To what extent did you have a sense of being in the same space as the</td>
<td>6</td>
<td>7</td>
<td>1</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>characters?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>To what extent did the presence of the characters affect the way you</td>
<td>6</td>
<td>5</td>
<td>3</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>explored the space?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>How much do you think you disturbed the characters in the room?</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>When you first saw them, did you respond to the characters more the way</td>
<td>5</td>
<td>4</td>
<td>7</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>you would respond to people, or the way you would respond to a computer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>interface?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Now consider your response over the course of the whole experience.</td>
<td>5</td>
<td>3</td>
<td>7</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>Did you respond to the characters more the way you would respond to</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>people, or the way you would respond to a computer interface?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Eliminating Design and Execute Modes from Virtual Environment Authoring Systems

Gary Marsden & Shih-min Yang
Department of Computer Science, University of Cape Town, Cape Town, South Africa
Email: gaz@cs.uct.ac.za, Tel: +27 21 650 2666, Fax: +27 21 689 9465

Abstract
In this paper we report on efforts to create a virtual environment authoring tool for novices. In particular we set out to eliminate separate design and execute behaviors from these tools. We present two alternative prototypes for achieving this and report on the results of a usability experiments comparing each environment.

Keywords- Virtual reality authoring, modes, programming environments.

1. Introduction
Within the Computer Science Department at the University of Cape Town we are embarked on building a software that supports novice users in creating desktop virtual environments. The most comparable system to our own is Alice[1], but whereas Alice is designed explicitly for children, our system is designed for domain experts (e.g. architect, teacher etc.) who do not have the necessary computer science skills to use existing systems such as DIVE[2] or Genesis[3].

The work for creating this tool is divided across various groups of researchers and programmers; in this paper we will concern ourselves only with the specific portion of the research conducted into the high level interface design.

In creating an interface to the authoring system, our intuition was that the biggest barrier to novices would be the programming interface. From our initial observations, however, we discovered that even before they reached the scripting level, users were confused by the whole notion of having separate design and execute modes. Users would click in frustration at objects, expecting them to react, only to be greeted with an attribute browser window.

It has long been known that modes in general are undesirable in a user interface, but this problem of separate design and execute modes has also been well reported in research relating to 2-dimensional interface authoring tools [4]. It seems that this duality, which is so obvious to programmers, confuses those not used to programming environments. This manifests in the form of mode confusion behaviors, like the appearance of the attribute browser, mentioned above.

In virtual environment authoring, we speculated that the impact could be even greater than in two dimensional interface builders.

Virtual environments are often touted as the last word in direct manipulation environments [5]. In a sense, there is no interface as the environment is a complete visualization of the entire system. In a really good virtual environment, the user should have a sense of ‘being there’[7], focusing more on the virtual environment than their physical environment. This notion of ‘being there’ is formally called ‘presence’ and is a very active research area in the virtual reality community. Environments which generate a strong sense of presence are obviously more effective than environments which do not (if the user does not relate and engage with the VE, then there is little point in creating it in the first place). One of the ways in which environments may inadvertently reduce presence is through the intervention of external events, or breaks-in-presence (BIPs) [13]. If we return to our user who is engaged in the activity of creating an environment, the need to switch out of the execute environment, to the design environment, to effect a change must surely introduce a break in presence. (The break would be further exacerbated if the user was wearing data gloves and a head mounted display). It would seem logical, therefore, to create an interface which allowed the user to alter the environment from within the environment itself. This will necessitate the removal of the “design” environment completely and somehow integrate that functionality into the “execute” environment.

2. Previous Research
By removing the design environment, we do not remove the need for separate modes of interaction. In the parlance of [4], we will always want to use things and
other times want to mention them, regardless of being in a real world or a virtual one. How then should one add these two capabilities to a virtual world?

The goal of supporting in-environment use and editing is not entirely new. For example, the Worlds in Miniature (WIM) system allows the user to manipulate a facsimile of an environment whilst in that environment [6]. However, it forces splitting the display real estate between the original environment and the miniature copy. In addition, it is difficult to select, and manipulate objects especially for fine-grained manipulations as the entire environment is scaled down into a hand-held size [8]. Perhaps most crucially, the metaphor cannot maintain the sense of presence because working with two different worlds emphasizes that the world is artificial.

The “Voodoo Dolls system” was proposed and implemented by Pierce et al [9]. This technique allows the users to interact with objects that are beyond one’s physical reach. It supports direct manipulation of an object by creating miniature copies of the object. This method gives the user an illusion of interacting as if in real life. However, the miniature copies of objects have different properties when they are held on the user’s right or left hand. This feature might confuse novice users since they might not fully understand the concept of the metaphor. Furthermore, presence may be reduced as users are forced to interact with the ‘doll’ of an object, rather than the object itself.

We are looking for a more general solution which tries to mirror the user’s real world experience as closely as the virtual environment will allow.

3. Migrating Functionality

Returning again to the two dimensional interface world, the “tool” approach is one of the proposed solutions for the problem of “mention and use” in a single window[4]. The tool approach is derived from everyday experience by adopting the interaction methods used in real life. It uses the idea of direct manipulation mediated through some tool. Every interaction requires some kind of tool, e.g. a “paintbrush” is used to paint objects, and a “hand” tool is for grabbing objects. Different tools can be seen as different modes so the interaction takes on the form of a global mode.

An alternative to the tool approach is adopting a mode-per-object. This approach allows different objects to be in different modes, meaning that there will be multiple active modes in one environment. An example of using the “mode per object” approach is presented in our earlier work[10]. A screw is attached to every user interface component. To edit a particular component, one clicks on the screw which reveals an attribute browser. To use a component, one simply clicks on any part of its surface (other than the screw) as in a normal system – see Figure 1 for a typical sequence if interactions. Thus in one environment, some interface components are in design mode, while some are in execution mode. In this particular interface, the screw is shown screwed in to reflect an object in execute mode; or shown screwed out to reflect edit mode.

Certainly this approach more accurately reflects the real world where different objects can be in both modes simultaneously. The difficulty of this approach are visual cues to indicate state, as they consume space on the virtual objects. Additionally, these visual cues would diminish the sense of the presence as users are aware of the artificial visual cues which appear on the virtual objects – we cannot simply place screws on every object in a virtual environment.

Whilst the mode-per-object approach more closely mirrors real world behavior, the tools approach is more common in current software. Therefore, before committing to a particular design we set out to investigate if different ways of interacting with objects will cause users to behave differently.

4. Investigating Behavior

We hypothesize that the users will change their ways of working depending on the system they are using. For instance, with the tool approach, we expect the users to work in the fashion of “tool by tool” and users might work “object by object” in the mode-per-object system.

4.1 Tool system

In real life, workers normally carry a toolbox to the working site. When they want to modify or fix an object, they take out an appropriate tool from the toolbox. Once
the job is done, they put the tool into the toolbox and walk to another place with the toolbox.

Most virtual environments are quite similar to real life in this respect, and therefore we felt that using the toolbox idea in virtual environments is appropriate. To implement the idea, a toolbox is provided in the virtual environment as a 3D object that can be opened and closed. In real life, the toolbox is not seen until it is actively sought out. It is difficult to implement this in the virtual environment because objects outside the users’ view frustum are difficult to access. Instead, we place the toolbox in a fixed location related to the user’s viewpoint. Thus the users know where to find the toolbox when it is needed, and time is reduced in searching the entire virtual environment for the toolbox if it is located at a fixed position. Figure 2 shows the tool environment.

![Figure 2 shows the toolbox, circled in the bottom left corner](image)

While the toolbox is open, the tools appear for selection. These virtual tools are represented as buttons and arranged in a virtual menu, shown in Figure 3.

![Figure 3 Shows the tools available once the box is opened](image)

In Houde [11], the different shapes of the mouse cursor are used as an indication of the action performed. In our work, we follow this method – once the desired tool is selected, the shape of the mouse cursor will change accordingly. For instance, while the “paintbrush” is selected, the mouse cursor will change to a “brush” shape. While the toolbox is closed, the virtual menu disappears automatically and the shape of the mouse cursor changes to default (i.e. the arrow shape).

4.2 Pin System

As a comparison, we built an identical environment to that described above, the difference being that the interaction was conducted on a mode-per-object basis. This approach allows the users to “use” an object while the object is not in editing mode. For this reason, we have created objects to populate the scene, each of which has a default “use” function. For instance, the users can open and close the door or turn on the television set. As we mentioned earlier, we need a way to indicate the mode status of each object and an easy-to-understand metaphor to edit objects.

We have adapted our earlier idea of the screw indicating mode and have drawn our metaphor from the way that an artist draws a picture. In real life, an artist would put drawing paper on the drawing board and pin the paper on. The pin is used to fix the paper on the drawing board. It can be also seen as an indication that the drawing is in process. Therefore, to pin an object and then edit it is the metaphor we use in the virtual environment, eliminating the need to place a screw equivalent on every environment object.

As the metaphor required, a drawing pin is provided in the virtual environment. The drawing pin, similar to the toolbox, is placed at the left-bottom corner of the screen and it is always in this fixed location. The drawing pin will not block the users’ view and it is always available despite of the users’ position (the system re-places the pin in the object so that it is always visible). See Figure 4.

![Figure 4. The drawing pin is at the corner of the screen (circled). The chair is pinned and the pin attached to the chair is in a different color from the drawing pin at the corner. While the chair is pinned, the users can still invoke the basic function of any unpinned object in the virtual environment.](image)

To pin an object, the user simply drags the drawing pin and drops it on the object. To show that it is being
edited, a new drawing pin will appear on that object. A tool list then appears from which the user can choose the desired tool and apply it to only that object. The users have to unpin the object in order to use the object, even when there is no tool mode set for that object. In order to avoid confusion between the drawing pin, which is always at the left-bottom corner of the screen, and the pin attached on object, we use different colors to differentiate class from instantiation.

We have provided visual feedback to indicate the currently active mode of the pinned objects. While the mouse is moving over the pinned object, the mouse cursor will change the shape according to the status of the object. Further feedback is provided via the texture on the pins, which are attached to the objects. The texture on the knob of the drawing pin will be the same as the active tools.

5. Study Design

For our study, we were able to find nine subjects who fitted the target user population we were interested in. These participants were paid volunteers and were students from various faculties in our university. We expect the end-users of these prototype systems to be non-experts in computer programming and computer graphics – they are interested in creating and editing their own virtual environments. However, they must be familiar with, and know how to use, standard computer input and output devices.

The experiment has a between-groups design. The participants were divided randomly into two groups and each was assigned to one of the prototypes. Four participants used the tool approach VE and five participants used the pin approach VE.

5.1 Tasks

There are two virtual rooms in the two prototypes. One room is a storage room, which contains all the furniture at the beginning of the experiment. The other room is a living room with a lamp inside. The users can walk freely in the virtual environment. However, they can only walk from storage room to living room through the door, and vice versa.

The task is to arrange the virtual living room according to the image in one of the virtual books. The users need to move all the furniture and objects to the other room (the living room), through the door. There are three books in the virtual environment. The books contain the images of three different arrangements of the room. There are three channels on the TV. On each channel, there are four images of the particular arrangements from different viewpoints. (One view is shown in Figure 5). The three books are marked differently at the back. One book is marked a “L”, one is marked a “I” and the last one does not have a mark. The users are asked to find the book with “I” at the back, and arrange the room accordingly.

We use the living room and storage room scenario as this is a real-world task familiar to users. Additionally, compared to manipulating some boxes, manipulating virtual furniture is more realistic. By placing instructions in books and on the television, we ensure that users are required to perform “use” actions in the environment – if users were give the plan on physical paper, they would not need to “use” the objects and hence comparison of the two interaction techniques would be pointless.

Figure 5. Pictures on the television set which shows users how to lay out the environment.

6. Measurement of Usability

We used two ways to measure the usability of both prototype systems, namely observation and questionnaires. The questionnaire we used is the Computer System Usability Questionnaire from AMC [12]. This was developed by J. R. Lewis of IBM and measures on a 7-point Likert scale. There are nineteen questions, which are based around three themes: system usefulness, information quality and interface quality.

Users were observed unobtrusively by splitting the monitor output lead and recording everything that happened on the user’s screen.

6.1 Tool Usage Observation Results

From post experiment video analysis, we have identified some behavioral patterns among the participants within each system and across both systems. With the tool approach prototype, we have found some patterns that the participants do while performing the task. These are summarized as following:
• Of all four participants, only one participant did not manage to complete the task. The other three managed to move all the furniture and books across the room and painted some objects.
• The users tried all tools on one object at the beginning.
• The users seemed to be familiar with the initially, but rapidly learnt how to use them efficiently.
• The participants moved objects to the other room in any order and put the objects in any position. Once all objects were in the living room, they then put the objects in the correct location. The last thing they did was to paint them.
• Some participants would put the book in the same orientation as the living room. That is, if they are standing in the storage and facing the living room, then they will place the book vertically and rotate it in such a way that the door is at the left bottom of the book.
• After using the system for a short while, the users understood that tools were applied to all objects. In other words, if they wanted to use the currently used-tool on other object, they did not have to click on the tool again.

6.2 Pin Usage Observation Results

With the pin approach prototype, we have identified some patterns of performing the tasks.
• Of all five participants, two of them did not complete the task of placing the objects correctly. The other three managed to move all the furniture and books across the room and only one of them did not start painting objects.
• When editing an object, no matter if the object was pinned or not, the participant would pin the object before starting.
• The participants would move objects to the correct position one by one. Once all objects are in the correct position, then they would paint the objects according the image in the book.
• Most participants did not fully understand the function of the drawing pin and would tend to invoke the functionality of the object while the object was pinned. Some participants only took a few mouse clicks to learn how to invoke the functionality of the objects; some took longer (about ten minutes). Eventually all participants learned before the end of the experiment. Once they learned it, they made fewer mistakes, and took fewer tries.
• The participants would move objects away from each other if they collided.
• Most participants were confused with moving the object up and down (along Y-axis) and push and pull (along Z-axis). Even with exploring in the virtual environment for a while, some subjects were still confused with these two operations.
• To put objects one on top of the other, the users used all manipulation tools (rotation, movement, and push and pull) to make sure that the objects were lying flat on each other, even though this was not necessary (there was no physics modeling).
• The way of choosing color – using red, green and blue sliders – is not a good idea. Most subjects took a long time to find the desired colors.

6.4 Questionnaire Results

We asked the participants to answer the usability questionnaire (CSUQ). We present a summary findings.

5.4.1 Positive Comments

The subjects feel that using the mouse to navigate, and to interact in the virtual environment is easy to understand.
• Collision detection on the walls was good because it let them know if the object is against the walls.
• The manipulation method is intuitive, as if the users were in the real world.
• They feel that they have control of the environment.
• Negative Comments
• There is no Undo function.
• There is no Zoom function.
• Pieces of furniture can pass through one another and people can walk into objects.
• It is difficult to find the correct colors.

7. Discussion

From an analysis of the questionnaire results, it seems that the tool approach prototype is preferred to the pin approach prototype in terms of usability. This is also confirmed by observations, most probably because the users are more familiar with the tool interaction metaphor. Apart from that, the questionnaire picked up shortcomings in the prototype (e.g. no undo feature) rather than provide meaningful insight into the difference in interaction techniques.

We had expected that the users work “tool by tool” in the tool approach prototype and “object by object” in the pin approach prototype. However, there were no observed behavioral differences in working technique.

One interesting observation was that the participants who used the pin approach prototype would re-pin the objects, even if they could see the pin. It would seem that, rather than reflect on the state of an object, it was much faster just to pin it regardless. However, the participants who used the pin approach prototype more
frequently invoked the function (“use” behavior) of objects, than those who used the tool approach prototype. In both systems, the subjects avoid walking into objects and when one object collided into another, they moved one away from the other. This would indicate a high level of presence amongst the subjects, as they could have walked through objects put instead projected the attributes of the physical objects onto those in the environment. Many researchers suggest that high fidelity graphics are required to induce high levels of presence. However, our virtual environments are not photo-realistic and some physical laws, such as gravity, and collisions between users and objects are ignored. Nevertheless, the participants attempt to model the virtual environment as the real world, and obey the laws of reality without prompting.

8. Conclusions and Recommendations

Initially we thought that the pin approach prototype would be more useful than the tool approach prototype. One reason is that the ability to work on different objects with different modes is more efficient because users do not need to change modes constantly. Another advantage of the pin approach prototype over the tool approach prototype is that the pin approach is somehow more intuitive for the “use” functionality than the tool approach – the users do not need to use the “hand tool” to use an object in the pin approach.

That said, our observation of the users’ actions shows that the users are more familiar with the tool approach than the pin approach. We suspect the reason it that global mode is more common in interactive software. As our subjects were all computer literate, this style of interaction may be more intuitive for them.

Although less suited to this task, it could be the case that the pin system would be more appropriate in an environment where the focus was primarily on use with only occasional editing required. This approach may also be suitable for collaborative virtual environments. The drawing pin can serve as a lock – while the drawing pin is on an object, others can see that the object is currently being edited. As the texture of the pin would change depending on the tools applied on that object, others in the environment can also know which operation is used on that object as well. These applications include architecture building, or design discussion meetings.

Acknowledgements

This work was made possible via NRF funding and Innovation Fund grants from the South African government.

References

Emotionally controlled Virtual Environments: A New Tool for Enhancing Presence through Media Content Manipulation

Beatriz Rey (1), Mariano Alcañiz (1), José Antonio Lozano (1), Rosa Baños (2), Cristina Botella (3), Nuria Lasso (3), Jorge Osma (3)

(1) Medical Image Computing Laboratory, Universidad Politécnica de Valencia (Spain)
(2) Universidad de Valencia (Spain)
(3) Universitat Jaume I de Castellón (Spain)

{beareyso | malcaniz | jlozano}@mediclab.upv.es, rosa.banos@uv.es, botella@psb.uji.es

Abstract

A participants’ sense of "being there" in a mediated experience is determined by a variety of characteristics or components. The EMMA Project (IST-2001-39192) is interested in analyzing the relationships between presence and emotions, especially for some Virtual Reality applications, such as mental health (both for promotion and treatment goals). This research will help to understand better the development of some psychopathological phenomena and the development of new correcting experiences and learning to cope with those psychopathological experiences.

Mood devices are special hardware and software configurations able to induce different forms of mood enhancement. In this paper, we describe one of the "mood devices" that have been developed inside the project: the EMMA room.

Keywords--- Mood devices, presence, emotions.

1. Introduction

1.1. Presence and emotions

A participants’ sense of "being there" in a mediated experience is determined by a variety of characteristics or components. There is a consensus that it is formed through interplay of raw sensory data and various cognitive processes. In particular, the most common definition found in literature is the “sense of being there” in one place or environment (i.e. a virtual environment) even when one is physically situated in another [1]. Starting from this view, Lombard and Ditton [2] reviewed a broad body of literature related to presence and identified six different conceptualizations of presence: realism, immersion, transportation, social richness, social actor within medium, and medium as social actor. Presence in a mediated environment can be defined as the sense of being there arising from a perceptual illusion of non-mediation: the extent to which a person fails to perceive or acknowledge the existence of a medium during a technologically mediated experience.

There are many key determinants of presence. However, they have been mainly classified in two groups in the literature: media characteristics and user characteristics. Regarding media characteristics, there are two different groups of factors that can influence the level of presence of the user: media form (which includes the extent and fidelity of sensory information, and the match between sensors and display) and media content (including factors such as familiarity, naturalness, interest, meaning, narratives). Regarding user characteristics, we can point out factors such as immersive tendencies and prior experience.

Other studies [3] have distinguished between external factors (resulting from the virtual environment system itself) and internal factors (ranging from structural aspects to psychological aspects).

The majority of presence studies conducted to date have focused on manipulations of media form and not media content. However, our position is that both media form and media content can affect the level of presence achieved by the user. We intend to investigate how media content manipulations can affect and have influence on presence. In this study, we will focus on the role of emotions, analyzing the relationship between presence and emotions.

Literature on psychological aspects of Virtual Reality (VR) has studied preferentially basic and common psychological processes (perception, memory, etc.) influencing important issues in VR. However, there are other personality and psychological variables that are also relevant to the VR field (suggestionability, dissociation, imagery, absorption, expectations, etc.), one of these being emotions. These variables are important in order to generate and enhance presence. This is especially true for mental health applications of VR. To design virtual environments with therapeutic goals, it might be necessary to put a good part of resources on achieving the highest emotional involvement. The corroboration of this issue would guide researchers in their efforts to achieve virtual worlds that work.

As some studies have pointed out [4], definitions on presence have mostly been cognitively or environmentally
based, generally ignoring the emotional aspects of presence. But emotions play an important role in our subjective judgments and automatic responses, influencing our learning and how we understand, describe and react to the world and ourselves. Important differences have been found [5] in the responses to VR environments between non-patients and (mental health) patients, which proved the importance of emotions for clinical users. Emotions may play a role both as causes and as consequences of presence. It may be said that the higher presence and reality judgment, the higher intensity of emotions the user experiences. Then, if the focus is on designing applications to elicit emotions with the goal to reduce or modify them, the environments have to be able to produce in the users the feeling of being there and of being "real". However, it could also be said the opposite: the higher intensity of emotions and feelings, the higher presence and reality judgment. From this point of view, the focus would lay on designing affectively significant environments. We would need to include those elements with the potential of activating emotions.

Emotions affect behaviors and cognitions, and they can have an important impact on presence. Users “feel” presence. And as presence is unstable, emotions are also continuously changing. If we are able to understand better presence and emotional reactions to VE, we will be able to design more effective “virtual” experiences. Especially in clinical psychology field, but not only, we need to know how to generate and optimize the emotional impact that the virtual experiences have from a therapeutic point of view.

In order to analyze these factors, we have developed several mood devices, which are special software (virtual environments) and hardware configurations designed to generate and enhance presence and emotions. They will be applied in selected populations: people who suffer from emotional problems, people who suffer from restricted mobility and general population.

1.2. Mood disorders

The target of EMMA is not focused on mental disorders, but on emotional experiences that are present in these selected populations. The purpose is to study the mood devices’ usefulness to improve emotional welfare.

The word “depression” or “anxiety” refers to distressful and uncomfortable moods. Often used to describe moods that all of us experience from time to time, these emotions can also be symptomatic of true clinical-medical disorders. Periods of depression and anxiety are quite common. Secondary to other medical or psychiatric illnesses, some individuals experience lengthy, often repeated or chronic episodes of these emotions for no apparent reason when they are otherwise physically well. These people are said to suffer from a mood disorder, where altered moods are central or “primary” and are accompanied by other symptoms, such as disturbances of sleep, weight changes, lack of energy, exaggerated worry and tension, low self-esteem, and so on. As mental disorders, both anxiety and depression are considered the most common mental illness, whose effects bring significant and costly problems. Mood disorders are important public health problems, due to its high prevalence, its comorbidity with other disorders, and their important negative consequences on the sufferers' quality of life. Epidemiological studies throughout the world [6] indicate that the life time risk for Major Depression ranges from 10% to 25% for women and 5% to 10% for men, and the prevalence rates appear to be unrelated to ethnicity, education, income, or marital status. In the case of anxiety, several studies have also shown high prevalence rates (for instance, 13% for social phobia prevalence, 10% for panic attacks). Furthermore, Adjustment Disorders, characterized by clinically significant emotional or behavioral symptoms in response to identifiable psychosocial stressors, are very common, and epidemiological figures vary widely as a function of the population studied and the assessment methods used [6].

1.3. Goals

Our goal is the investigation of the use of engaging media for the development of non-addictive, mood-stabilizing experiences, in particular, after analyzing the possible emotional impact of high compelling synthetic experiences characterized by a high level of presence.

“Mood devices” will provide innovative ways of coping with distressful emotions, that will be better than existing approaches, for different users: users who suffer from psychological problems (affective disorders, anxiety disorder, adjustment disorders), users with acute restricted mobility (the emotional mediated experiences that bedridden patients could have by means of mood devices may help relieving their anxieties, reducing their pain, and encouraging them in their fight against diseases), and mood enhancement for general population (relaxation environments through TV or VR; presence-enhanced synthetic environments for entertainment, etc.). We intend to achieve a more complete understanding of presence and reactions to mediated experiences. This will help us in creating more effective experiences for emotional learning, which could be useful in many different contexts.

Another innovation is to use VR as a “new realistic laboratory” [7] where to study behaviors, emotions, thoughts, basic psychopathological processes, individual differences and emotions. This “realistic lab” will allow to do research with a high degree of validity. It is classic the dilemma between the different types of validity. It seems that we usually sacrifice something regarding the internal or external validity: As a greater control is needed, it seems necessary to turn to the “artificiality” and/or “simplicity” of the lab. The “virtual laboratory” could help overcoming this dilemma creating significant contexts, with high external and ecologic validity, in which certain questions can be tested with a high degree of control and accuracy.

2. EMMA room

We are going to describe in this paper one of the mood devices that we have developed: the EMMA room. The main characteristic of this mood device is that its aspect can be changed dynamically controlled by the therapist.
depending on the emotions that the user is feeling at each moment. We intend to study the role that this environmental changes can have on the level of presence experienced by the user.

In order to visualize the virtual environment, we have selected a big retro-projection screen where the virtual environment is projected. The navigation and interaction device is a wireless joystick.

In the following points, we are going to describe the main elements that compose the environment.

2.1. General description

The room developed will help the user to remember and re-live past experiences.

It has been modeled as a big hall with circular shape, with no walls so the outer environment can be visualized.

![Figure 1 Room external aspect](image1)

**Figure 1 Room external aspect**

![Figure 2 Room internal aspect](image2)

**Figure 2 Room internal aspect**

The environment includes a set of different systems that will interoperate allowing the user to express ideas by means of items, mixing or modifying their characteristics and changing their positions. Most of the systems have been programmed as object holders. They are the places where database elements can be stored.

There are different types of object holders:

- The environment object holders allow the user to show the three-dimensional aspect of an element, or to play the associated sound or video.
- The database object holders only show the element iconic views and store them in the database screen.
- The living book object holders. This special 3D object allows the user to show and classify the element iconic views in the different pages and chapters defined by the user.
- The discharge area object holders allow the user to put any element on them to interact with it.
- The drain, which in fact does not hold any element, has an interface that is exactly the same as the other object holders.

A virtual keyboard has been created that allows the user to label elements of the environment object holders or chapters in the living book.

There is also a special module called the inventory. This interface consists of a set of object holders that are always shown on the screen. At any moment on the session one of the object holders is active and will receive the next picked element or give the next dropped one depending on what the user does. It is used as a way...
to translate elements from one place to another in the environment.

In the following points, we are going to describe in more detail the main elements of the environment and their functionality.

2.2. The database screen

2.2.1. Description
The database screen floats around the big hall. It is composed of several tabs that give access to the different element categories of objects that can be hold and used inside the system: sounds, videos, images, three-dimensional objects and colors.

2.2.2. Interaction details
Each category is made up of an array of special object holders; its special feature is that they do not disappear from the screen when the user picks them to the inventory tool. Instead, a copy of the element is made and put on the inventory tool.

A three-dimensional scrollbar module has also been used on the database screen development. A tab module has been created that allows several pages on the screen. This module has also been used with some modifications on the living book development.

![Figure 3. Database screen aspect](image)

2.3. The living book

2.3.1. Description
The living book is the most important piece of the environment. It is also formed of object holder arrays and, as the database screen, it is organized in pages. Each page represents a chapter of the living book. A title can be introduced for each chapter by using the virtual keyboard or if no title is introduced the page is considered to form part of the previous chapter.

The main use of the living book is to help the user to re-experience or re-live the past as it happens with family photographs and home videos. It is the instrument that the user will use to keep and put in order all the contents he/she has worked with the psychologist during the session in the virtual environment.

The main reason to design this tool as a book is that it permits the same functions in the real world so the users will understand its use in a natural way. The living book, however, is not written mainly with words. The user writes it with the different kinds of materials used in the EMMA environment: 3D objects, images, films… will be manipulated in the environment, so the user can move them, keep them in the book and recover them from the book at any moment. The user can change the situation of any object inside the book.

The living book is a completely flexible tool, every user can adapt it, and it is useful to keep together all symbols with emotional meaning for the user.

2.3.2. Interaction details
Initially, the living book is empty. The user can select the elements that will be introduced at each of the chapters directly from the database screen or from an environment object holder. The way of acting is the same: the user has to select the object just by clicking on it; then, the element will be introduced in the inventory; finally, clicking on the correct slot inside the living book, the element will be copied there.

Once the elements have been copied in the book, their order inside a chapter can be changed at any moment on the session just by picking an element (to the inventory) and dropping it on a different slot. When an element is dropped on a slot previously occupied a swap is made between the inventory element and the living book one, making it very fast and easy to change the order of elements in the book.

![Figure 4. The living book and the database screen](image)

2.4. The environment object holders

2.4.1. Description
The environmental object holders will be a means to personalize the environment. The user will be able to select the different elements (sounds, images, videos, colors, 3D objects) directly from the database or from the living book. These elements will be relevant for the patient, and will be selected with the help of the therapist during the session.

2.4.2. Interaction details
An environment object holder can serve as a mixer tool to combine several elements to form a new complex element that will be used exactly in the same way as the
others. When a complex element is placed on an environment object holder, all the forming elements are shown simultaneously floating around the object holder.

The way in which the user interacts with elements consists on picking them from any object holder to the inventory tool and dropping them in any other object holder. When an element is dropped on an environment object holder that already has an element inside, a composed element is created with the sum of the elements, for example, sound and image or geometry and colour.

2.5. The drain

2.5.1. Description

The drain is an element of the environment where the user will throw any element that is not needed anymore. Objects that are placed in the living book or in the room can be dropped to the drain in order to destroy them.

The drain is placed in a corner of the room.

The user puts inside the drain any object he/she wants to get rid of. At that moment, an animation is launched and some noises are heard that show that the element is being destroyed.

2.5.2. Interaction details

As it has been explained before, the drain behaves in a very similar way to the rest of object holders. When the user drops an element from the inventory to the drain the element is deleted from the inventory. The only difference is that no element is created on the drain object holder.

2.6. The discharge area

2.6.1. Description

One of the targets in EMMA is to give the environments an interactive way to behave and the ability to be modified regarding the emotions of the user returning him/her a feedback and establishing a communication.

The emotional discharge system provides a personal and private space where people could give free rein to their emotion with nobody watching or evaluating them with the exception of the psychologist. With this system, it is possible to modify the characteristics of the virtual environment in function of the emotions of the users, returning to them some kind of response, as it happens in the communication with another human being, and avoiding the possibility that the users could “hide” their emotions. However, the answer of the virtual environment is different from the answer that the user could receive from another human being. This answer is expressed in modifications of the environment the user is immersed in.

With the emotional discharge system, the user can manipulate the objects of the environment. This is the way to carry out the change of the meanings that the objects have for the user. This is not different from the rest of the environment, as for example working with the living book. The previous work will teach the user to give personal meanings to the objects inside the system.

During the work in the discharge area, as it happens in the rest of the environment, the user will be accompanied by the therapist. The psychologist and the user will work together this experience from a psychological point of view.

2.6.2. Interaction details

There are three special object holders that are in the balcony of the room. In these object holders, the user can...
modify the shape and aspect of the objects that are placed on them by means of the voice (shouting in a louder or softer way). A system has been programmed that detects the loudness of the input sounds and modifies accordingly the size of the objects placed on those special object holders. More than one object holder can be active at the same time. In this case, the aspect of the objects placed in all of them will change simultaneously.

2.7. Dynamic changes of the environment

2.7.1. Description

One of the main interests of this environment is that its aspect can be changed dynamically depending on the emotions of the user. As long as the therapist will stay with the user during each session, he/she will be the person that will select the aspect of the environment depending on the emotions that the user is feeling at each moment during the session.

We have designed four pre-defined aspects for the outer part of the virtual environment corresponding to four different emotions: joy, relaxation, rage and anxiety. The therapist can select between this four options. And, besides, it is also possible to select between different effects that can be applied to the environment: fog, rainbow, rain, snow, earthquake, and so on.

Figure 8. Aspect of the outer environment corresponding to joy with the EMMA avatar

Figure 9. External aspect of the environment corresponding to anxiety
2.7.2. Interaction details

In order to achieve these effects, a special interface has been prepared that allows the therapist to control several aspects of the appearance of the outer part of the EMMA room. The application for controlling this appearance is running on a computer different from the computer where the virtual environment is running. The commands that the therapist introduces are sent using TCP/IP to the environment computer, and the appearance of the environment changes depending on the command that the computer has received.
Conclusions

It is our position that variations in both media form and media content can influence presence. The fact that content can affect both presence and emotions suggests a need for research to understand the relation between the two.

Using this environment, we are measuring aspects of presence and emotion in order to study the influence of the emotiveness of a media experience—manipulated through variations in media content—on presence and we expect to obtain soon conclusive results.

References

Towards Emotionally Adapted Games

Timo Saari, Niklas Ravaja, Jari Laarni, Kari Kallinen, Marko Turpeinen
M.I.N.D. Lab/CKIR/HSE & HIIT, CKIR/Helsinki School of Economics, Helsinki Institute for
Information Technology
{saari@hkkk.fi, ravaja@hkkk.fi, laarni@hkkk.fi, kallinen@hkkk.fi, marko.turpeinen@hiit.fi}

Abstract

In this paper, we present a framework for a gaming personalization system to systematically facilitate desired emotional states of individual players of games. Psychological Customization entails personalization of the way of presenting information (user interface, visual layouts, modalities, narrative structures and other factors) per user or user group to create desired transient psychological effects and states, such as emotion, attention, involvement, presence, persuasion and learning. By varying the form of information presented in a game in an emotionally intelligent way it may be possible to achieve such effects. Theory, key concepts, available empiric evidence and an example of an application area in emotional gaming as well as a basic system design are presented.

1. Introduction

Gaming research is often conducted on the basis of game content and genre analysis, typologies of gaming styles or consumption and sales of games. Little research is done on the actual user experience of games, such as presence and emotion. Most of this research does not conceptualize user experience, such as emotion, sufficiently from the point of view of emotion psychology. It is often also concentrating on some overly technical aspects of gaming such as ways to detect emotions of players with sensor technology. Further, very little research is available to understand how various individual differences may influence emotions during gaming. Also, the emotional influence of the way of presenting the basic elements of a game has not sufficiently been studied. The field also lacks a more general framework of seeing how one may match individual differences to the way of presenting and adapting games to create emotional effects.

This article explores these issues and attempts to provide a preliminary approach to understand emotionally adapted games.

User experience is seen within this article as the transient attentional and emotional states, presence, moods, information processing, learning, flow, persuasion and various other subjective experiences occurring i) just before a user engages with technology, ii) as a result of the user perceiving and processing information mediated by technology during a session of use and iii) immediately after the use of technology within a given task and context of use. User experience is related to action, i.e. the user has an experience and may perform a certain action or decide not to act, i.e. the user has creative and autonomous degrees of freedom for action.

When perceiving external stimuli like information or a game via a communication technology users have a feeling of presence. In presence, the mediated information becomes the focused object of perception, while the immediate, external context, including the technological device, fades into the background [3, 28, 29]. Empirical studies show that information experienced in presence has real psychological effects on perceivers, such as emotional responses based on the events described or cognitive processing and learning from the events [see 40]. It is likely that when playing games people experience presence with the content of the game.

2. Psychological Customization

Media- and communication technologies, such as systems for gaming, as special cases of information technology may be considered as consisting of three layers [1]. At the bottom lies a physical layer that includes the physical technological device and the connection channel that is used to transmit communication signals. In the middle is a code layer that consists of the protocols and software that make the physical layer run. At the top is a content layer that consists of multimodal information. The content layer includes both the substance and the form of multimedia content [2, 44]. Substance refers to the core message of the information. Form implies aesthetic and expressive ways of organizing the substance, such as using different modalities and structures of information [44]. Naturally, these are difficult to separate.

Technologies are most often designed from the point of view of available communication capacity, software and hardware around a certain task. In addition to this there is another approach to design of technology taking into account the user experience of the users of technology and the goals users may have regarding a certain technology. For instance, a user may wish to perform a certain task as efficiently as possible, or a user may wish to be in a good mood when performing the task. Hence, there is another
layer for design of technology in addition to the task-based approach.

This type of design may be called as being Mind-Based, i.e. paying attention to the needs and goals of users to achieve desired user experiences, such as positive mood and efficiency of cognition. It also involves personalization and customization that entail the automatic or semi-automatic adaptation of information per user in an intelligent way with information technology [42, 54].

Media- and communication technologies may be called Mind-Based if they simultaneously take into account the interaction of three different key components: i) the individual differences of perceptual processing and sense making of different segments of users, ii) the elements and factors inherent in information and technology that may produce psychological effects (physical, code and content layers), and iii) the consequent transient psychological effects emerging based on perception and processing of information at the level of each individual. [51]. Hence, Mind-Based design takes into account individual differences in processing information in order to be able to offer a particular user a particular type of experience. Naturally task and for instance the content or substance of information, such as email, Mobile Multimedia Messaging (MMS) or a news article, influence the user experience.

However, it may be feasible to facilitate desired user experiences just by varying the form of information. For instance, with Mind-Based Technologies one may vary the form of information per user profile, which may systematically produce, amplify, or shade different psychological effects. This type of system design approach may be of practical use, as it is known that individual differences in processing information may produce sometimes quite large variance in the intensity or type of psychological effects, such as depth of learning, positive emotion, persuasion, presence, social presence and other types of psychological states and effects [44, 45, 46, 47].

Hence, the Mind-Based Technologies approach may be valuable also when creating systems for facilitating clear and cognitively optimised or emotionally focused interactive and adaptive systems. One operationalization of Mind-Based Technologies in system design is Psychological Customization [51, 55].

Psychological Customization includes modeling of individuals, groups, and communities to create psychological profiles and other profiles based on which customization may be conducted. In addition, a database of design rules is needed to define the desired cognitive and emotional effects for different types of profiles. Once these components are in place, content management technologies can be extended to cover variations of form and substance of information based on psychological profiles and design rules to create the desired psychological effects.

The key idea is that matching a particular individual difference, such as user’s fluency in processing textual information, with a desired user experience, such as the need for efficient cognitive processing. This results the system presenting the user with as much text-based modality within a certain task as possible. This would then enhance information processing related to task of that particular user. Another user may be more fluent in processing audio information or video material with the same task to achieve more efficient processing of information, for example.

Psychological Customization can be applied to various areas of HCI, such as Augmentation Systems, Notification Systems, Affective Computing, Collaborative Filtering, Persuasive Technology and Messaging Systems. It can be hypothesized that the selection and manipulation of substance of information takes place through the technologies of the various application areas of Psychological Customization. Underlying the application areas is a basic technology layer for customizing design. This implies that within some limits one may automatically vary the form of information per a certain category of substance of information. The design space for Psychological Customization is formed in the interaction of a particular application area and the possibilities of the technical implementation of automated design variation [51].

The particular focus of this paper is gaming. Even though no actual system has been implemented yet for Psychological Customization related to gaming, empirical evidence supports the feasibility and validity of this idea. First, it is well established that there are individual differences in cognitive processes such as attention, memory and language abilities. These individual differences have a considerable effect on computer-based performance [e.g. 8]. For example, individual differences in memory capacity have an effect on people’s behavior in many types of activities [57]. This suggests the need for Psychological Customization Systems that optimize the presentation of information to different target groups having different psychological profiles. There is considerable evidence in literature that varying the form of information creates for instance emotional and cognitive effects [e.g. 15, 16, 18].

In media studies it has been found that different modalities, such as visual and auditory, may lead to different kinds of psychological influences and the valence of a preceding subliminal stimulus influences the subsequent evaluation of a person evaluated [5, 14]. In educational studies it has been shown that different ways of processing information influence learning and emotion of stimuli with certain modality [41]. Research concerning emotional influences on the cognitive processing of information has often concentrated on how different emotions related to information change the way users pay attention to, evaluate and remember the mediated message. This research has results on the influence of emotional information as increasing the user’s self-reported emotion [22]; attention (physiological and self-reported) [21] and memory for mediated messages, particularly arousing messages [20, 21, 22]. Studies in experimental psychology have shown that recognition and memory can be influenced or even enhanced by previous exposure to subliminal visual or auditory images of which the subjects are not consciously aware [13]. Some of these effects are produced in interaction with individual differences, such as cognitive style, personality, age and gender.
In our own research on the influence of form factors of colour screen PDA’s and mobile phones of information (such as news, games, messaging content and entertainment content) on psychological effects has yielded many results. Regarding emotional responses has also shown that subliminal exposure to happy affective primes in connection with video messages presented on a small screen has several putatively positive influences (i.e., increased pleasure, perceived message trustworthiness, and memory) [38]. Further, recent studies in our laboratory have empirically confirmed that media messages can be modified in terms of audio characteristics [12, 37] and the presence of image motion [35] to meet the personality (as defined in terms of dispositional behavioral activation system sensitivity) of the user, thereby enhancing his or her attentional engagement, information processing, and enjoyment. The role of hardware should not be neglected. A device with a large screen or a portable device with smaller screen with user-changeable covers may also influence the emerging effects [e.g. 17].

Saari (2001) has grouped the clusters of form related variables relevant to psychological effects as: i) hardware layer (size, proximity, fixed place/carried by user), ii) code layer (way of interaction and degree of user control, ways of presenting visual-functional controls in the user interface) and iii) content layer (substance: essence of the event described in the message, form: modalities, visual layouts and temporal structures).

The empiric evidence presented here partly validates the possibility for Psychological Customization Systems at least with mobile devices and user interface prototypes used in our own research. Typical experiments we have conducted on the influence of form of information on psychological effects have included such manipulations as animation and movement (for orientation response), fonts of text, layout of text, background colors of text, user interface navigation element shapes (round vs. sharp), user interface layout directions, adding background music to reading text, use of subliminal affective priming in the user interface (emotionally loaded faces) and use of different modalities of information, for instance.

As the task of capturing and predicting user’s psychological state in real time is highly complex, one possible realization for capturing user’s psychological state is to have the user linked to a sufficient number of measurement channels of various i) psychophysiological signals (electroencephalography [EEG], facial electromyography [EMG], electrodermal activity [EDA], cardiovascular activity, other), ii) eye-based measures (eye blinks, pupil dilation, eye movements) and iii) behavioural measures (response speed, response quality, voice pitch analysis etc.). An index based on these signals then would verify to the system whether a desired psychological effect has been realized.

Another approach would be to conduct a large number of user studies on certain tasks with contexts with certain user groups, psychological profiles and content-form variations and measure various psychological effects as objectively as possible. Here, both subjective methods (questionnaires and interviews) and objective measures (psychophysiological measures or eye-based methods) may be used as well interviews [for a review on the use of psychophysiological methods in media research, see 36]. This would constitute a database of design-rules for automatic adaptations of information per user profile to create similar effects in highly similar situations with real applications. Naturally, a hybrid approach would combine both of these methods for capturing and facilitating the user’s likely psychological state.

3. Applications in Gaming

3.1. Basic Elements of Games

Games have often been researched from the point of view of narrative, consisting of a dramaturgical structure focused on crisis and the resolution [32]. However, it may be that gaming is not easily understood as a linear narrative. For instance, a gamer may be more interested in collecting points and more powers for his character inside the game and mere survival between different levels of the game than in moving along a story line coherently towards a climax [25].

Similarly, it has been argued that the participatory aspect of gaming is the key to the experience of gaming [7]. It may be stated that the algorithm of the game is another key source of experiencing a game [31]. This implies that as the player learns the hidden rules and logic behind the game and is therefore successful in playing it, a state of satisfaction may arise. Further, it is evident that the skills of the gamer vs. the challenges presented in the game should be in balance [11]. If a game is too difficult or too easy to play, it may not be involving, but rather frustrating or boring. Another difference between narrative and games is that the tensions in narrative are dependent on the irreversibility of the consequences of the events of the narrative [33]. For instance, if the hero dies in the end of a book, it creates emotions because it cannot be reversed. However, in gaming it is often possible to gain a “new life” and return to a situation where one has “died”. Player’s knowledge of this may lessen the involvement with a narrative or a dramatic event.

Despite the differences between traditional narratives and games, many similarities exist. The basic structure and elements of the narrative may be present also in games although the reactions of a player of a game may be different than those of the reader of the book to a particular linear sequence of a storyline due to the possibility of reversibility of effects of events or their consequenses. It is proposed that narrative is a basic way of organizing people’s experiences of various events of the world into sequences with a beginning, middle and end [4]. Further, memory research has demonstrated that schemas related to even complex events are often organized according to such a narrative structure [see 30].

A narrative schema may have the following structure: i) introduction of a setting and key characters, ii) explanation of the current state of affairs or the situation at hand, iii) initiating event leading to a motivation to act or change the state of affairs, iv) emotional response of the
protagonist and a goal for acting or changing the state of affairs, v) the difficulty experienced by the protagonist while performing actions to change the state of affairs and vi) the outcome of the action of the protagonist, i.e. success or failure in changing the state of affairs [4]. This implies that i) the role of the characters in the game is of key importance, including the role and point of view of the player and his character or role and ii) it is possible to create emotional reactions and motivation in the player to act in a desired manner by introducing events in a certain manner, or by offering a chance to succeed in attaining a goal. Hence, from the point of view of emotions, manipulating the events within a particular sequence of the game as well as introducing the situation and creating basic tensions and motivations as a basis for the task of the user in the game are important.

Outside narrative elements of a game, also the factors related to the presentation of the substance of the game or the form of the game, such as visual representations of the gaming events, amount and pace of image motion, audio effects and background music, and the level of interactivity offered to the player, are important from the point of view of emotion.

However, despite the amount of research on gaming the systematic adaptation of elements of the game to produce psychological effects has not been presented. A basic approach to an element to be adapted inside a game is a psychologically validated template that is embedded inside the game to create a particular psychological effect. A broad view of templates may be that the whole game consists of a database of psychologically validated templates that are presented in sequences. A limited view entails that a smaller collection of templates is used. The element of psychological evaluation means that the selected psychological influence (such an emotional response) of the template on a particular type of user is sufficiently well predictable.

These psychologically evaluated templates may consist of i) manipulating the substance of a game, such as story line (initiating events, new characters etc.) and manipulating the situations specifically related to the character of the player (such as putting the character into sudden and dangerous situations inside the game) and ii) manipulating the form or way of presentation of the game (such as visual elements, shapes, colours, types of objects, sound effects, background music, level of interactivity and feedback etc.). The difficulty level of the game may also be continuously automatically be adjusted, thereby keeping the skills and challenges in balance, which results in a maintenance of an optimal emotional experience and possibly also a flow-state.

3.2. Gaming, Presence, Emotion and Personality

How then to conceptualize emotion and related psychological constructs in gaming to understand how to psychologically validate the influence of various gaming templates? Emotions are biologically based action dispositions that have an important role in the determination of behavior [23]. This would be expected to be the case also in connection with gaming behavior, which makes emotional responses elicited by games an important target of investigation. That is, people seek, and are eager to pay for, games that elicit optimal emotional responses (or response profiles).

Emotions can be seen as constituted by three aspects or components: subjective feeling, expressive behavior, and physiological arousal; others add motivational state or action tendency and/or cognitive processing [23, 53]. A dimensional theory of emotion holds that all emotions (e.g., joy, fear, depressed feeling, pleasant relaxation) can be located in a two-dimensional space, as coordinates of valence and arousal (or bodily activation); [e.g., 23, 26]. 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.

Presence [i.e., the perceptual illusion that a mediated environment is not mediated; 27] may also be related to emotions. Factors such as a first-person view and naturalness may contribute to the experience of presence. Video games that engender a greater sense of presence are likely to elicit greater physiological arousal (at least when the game content is arousing), involvement, and attention (real-world stimuli, or stimuli that are perceived more or less as such, are likely to elicit greater attentional engagement compared to stimuli that are readily perceived as mediated presentations of real-world stimuli; [see e.g., 34, 35]. Importantly, games contributing to a high sense of presence have been suggested to be highly entertaining and, simply put, fun [greater enjoyment; 27]. The sense of presence may also facilitate game performance [see, 27].

People’s responses to different kinds of media have often been shown to vary as a function of their personality [e.g. 35]. Video games vary wildly in terms of thrill and violence, for example. It is likely that whether players find a thrilling, violent game as more entertaining and engaging compared to a nonviolent game with a happy atmosphere depends on the player’s enduring personality traits. In this connection, the Sensation Seeking trait might be particularly relevant, given that high sensation seekers have a high general need for thrills and excitement [39, 59].

How then to verify emotional reactions to validate the emotional influences of chosen gaming templates on different types of users when playing games? One solution is to have the user linked to a psychophysiological measurement system. An important advantage of psychophysiological measurements is that they can be performed continuously during game playing and have a high level of temporal precision.

It is well established that tasks requiring cognitive effort or active coping (e.g., many games) elicit emotional arousal accompanied by sympathetic nervous system (SNS) arousal and tonic heart rate (HR) acceleration (Ravaja, in press). Thus, in this connection, a tonic HR acceleration indexes emotional arousal. EDA, commonly known as skin conductance, is another psychophysiological index of arousal. As people experience arousal their SNS is
activated, resulting in increased sweat gland activity and skin conductance. [see 36]

Although tonic HR acceleration is related to emotional arousal, phasic changes in HR are sensitive indicators of attention. There is a considerable body of research showing that HR decelerates when attention is paid to an external stimulus or information is taken in [e.g. 19, 56]. The phasic HR deceleration is an important component of the orienting response (initiates the processing of external stimuli), and is mediated by the parasympathetic nervous system [36].

Facial electromyography (EMG) is frequently used as a psychophysiological index of hedonic valence [e.g. 24]. Increased activity at the zygomaticus major (cheek) and corrugator supercilii (brow) muscle regions has been associated with positive emotions and negative emotions, respectively, during affective imagery and when viewing media [24, 36]. Several studies have also found that tonic activity at the orbicularis oculi (periocular) muscle area is greater during high-arousal than during low-arousal emotion conditions [e.g. 58]; orbicularis oculi activity appears to be particularly high during positively valenced high-arousal emotions [36].

3.3. Example: Emotionally Adapted Gaming

The question then remains, how can a Psychological Customization system modify the emotional responses of various gamers? What elements of each gaming template may be varied per type of player to enhance various effects? Without detailed empiric evidence this is difficult to answer. However, at the conceptual level three typical emotionally focused scenarios of gaming are presented to shed light on this issue.

In the scenario, the user is sitting home alone at night with nothing special to do. He is a bit bored and needs some excitement and uplifting experiences. He opens his computer and selects a 3D first person shooter type of computer game to be played. From the available settings to control the type and intensity of his emotional experience in the game he decides to set the general levels to maximum arousal and selects the valence levels to be set also to maximum whether positive or negative. He thinks this should lift his mood and provide some exciting action. He further selects a medium level of difficulty of playing the game, resulting in a good balance between his skills and possibility to have engaging experiences inside the game, not being bored or not getting killed all the time.

The system has previously tested and modeled his personality and various other psychologically relevant factors. He then hooks himself up with a sensor system linked to the system measuring his level of emotional valence and arousal (such as measuring HR and skin conductance from his finger and EMG from his facial muscles in a non-intrusive and simple way). The game begins. The user is introduced to the starting point of the game: he needs to save the earth from alien invaders by using futuristic weaponry. While playing the user is suddenly disturbed by a high-speed and seemingly hostile alien attack. The user is in quite a hurry to point his weapon and shoot the alien and he just barely manages to do this.

Next, the user navigates the terrain in the game and is presented with an especially grim and scary scenery of ruins of a well known city. The lighting conditions, shadows, appearing nightfall, shapes and colouring of objects seen in the visual field, the audio effects and the quiet background music are together creating quite an eerie and sinister effect. As the user ranks quite high on the sensation seeking scale, the level and intensity of stimuli to create a particular emotional reaction and mood is increased throughout the game. The system may also use subliminal stimuli, such as smiling or angry faces not detectible to the user to increase particular emotional states in the game.

In another scene, a desperate little girl character suddenly appears to very persuasively plead for his help to save a group of humans captured by aliens in a nearby village. This creates an urgent need in the user to go and release the prisoners. The user goes about this and succeeds in the task well over expectations. Consequent praise of the released prisoners, including several quite attractive females, and the simultaneous subtle change in the tone of the present contextual scenery to be more uplifting (lighting, types of objects, music, audio effects) add to the feeling of high satisfaction and positive valence of the player. Hence, all emotional reactions are taken to the extreme inside the game.

From the point of view of ecological validity it may be stated that the key to a “good” fighting or war game is the optimal division of different types of emotional experiences while gaming, rather than just intensifying for instance excitement and arousal all the time. For instance, fear and hatred may be skilfully interlaced with joy and positive emotion. In other words, some parts of the game contain hatred and fear but there also have to be parts in which these are relieved and moments of victory and joy can be experienced (a terrible enemy has finally been devastated by the player).

Another approach not concentrating on maximizing and interlacing for instance negative high-arousal states and positive emotions would be a parental control of the level of emotional reactions elicited by a game. In such a scenario, parents may set the valence and arousal levels desired in a game to a particular level, such as avoiding too much arousal and negative valence, hence making most of the game either neutral or positive and only moderately arousing.

Yet another idea would be to maximize entertainment and fun within a genre of game that is quite positive by basic attitude, such as a level-playing game (Mario or similar). In this type of game scenarios and elements may be varied to create a flow of experience mostly characterized by moderate arousal and high positive valence, when possible. Naturally one may think of several other feasible scenarios.

Evidently, these scenarios may represent some mainstream uses of gaming. How then can emotionally adapted gaming add value in these scenarios? To answer this, first a matrix including the technological possibilities of creating emotionally focused gaming templates is presented in Table 1.
Currently, if the capture of user’s psychological states with psychophysiological and behavioral methods is realized with future gaming technologies, it may be possible to more objectively capture the effects of a given emotionally positioned template on the receiver designed to induce positive emotion, for instance. Such recording technologies would make the system more reliable and allow for fine-tuning of effects and allow the system to learn each individual’s patterns of responses in order to conduct better Psychological Customization over time.

As noted above, there are several psychophysiological measures that may be used to capture player’s emotional and attentional responses. First, tonic and phasic HR can be used to index emotional arousal and attention, respectively. Second, EDA is also a very sensitive index of emotional arousal. Third, facial EMG measured from the zygomaticus major and corrugator supercilii muscle areas can be used to index positive and negative emotions, respectively (i.e., the valence of an emotional experience). Finally, EEG can be used to measure both emotional valence and attention.

4. Conclusion

The authors know of no other comprehensive framework to systematically adapt elements of games to create desired psychological effects, such as emotions.

It should be noted that to build a smoothly functioning Psychological Customization system one should do much more research and gain more evidence of the systematic relationships of user profiles, information forms and psychological effects than what is currently reported in scientific experiments with available methods of acquiring such complex information. More empiric evidence is needed per selected focus area before conclusions can be provided. However, in our research for the past four years in varying the form of multimodal information, such as news, entertainment and games presented through color-screen high-capacity mobile phones we have found many emerging commonalities and possible rules for altering designs and user interfaces to reliably facilitate emotions, presence, attention, and learning.

Consequently, clear and conclusive hypothesis, best practices for design or other low-level and explicit recommendation on how exactly to build and best use a Psychological Customization system in gaming is beyond the scope of this conceptual paper.

Despite this, the authors argue that the approach to system design presented in this paper may be beneficial in gaming application areas because: i) it provides a possibility to personalize the form of information that may be more transparent and “acceptable” to the users than adapting the substance of information in gaming, ii) it offers a way of more systematically accessing and controlling transient psychological effects of users of games by the users themselves, and iii) it is potentially compatible with existing and new systems as an add-on or a middleware layer in software with many potential application areas.

The potential drawbacks of the framework include the following: i) it may be costly to build the design-rule

---

<table>
<thead>
<tr>
<th>Layer of Technology</th>
<th>Emotionally Adapted Gaming Templates</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Physical</td>
<td>-Mobile device: user changeable covers in colours and shapes that facilitate desired emotion. -PC: colours and shapes that facilitate desired emotions.</td>
</tr>
<tr>
<td>2. Code</td>
<td>-The user interface elements (background color, forms, shapes, directions of navigation buttons etc.) may be varied in real-time per page per user in which a certain advertisement is located to create various emotions and ease of perceptual processing. -audio channel may be used to create emotional effects (using audio input/output sound, varying pitch, tone, background music, audio effects etc.). -the interaction modalities may be adapted to suit the nature of the task.</td>
</tr>
<tr>
<td>3. Content</td>
<td>-The genre of the game or type of game should be taken into account (first person shooter, simulation game, level playing game, other). -Various emotionally engaging story lines may be used to create different types of emotional reactions. -The role of the user in the story can be varied to create emotional reactions. -Adding subliminal extra content to create desired emotion while playing.</td>
</tr>
<tr>
<td>A. Substance</td>
<td>-Multimedia content created by authors or generated by game algorithms.</td>
</tr>
<tr>
<td>B. Form Modality</td>
<td>-Modality may be matched to cognitive style or pre-existing mood of the receiver to create ease of processing. -Background music, audio effects or ringing tones may be used as a separate modality to facilitate desired emotions and moods.</td>
</tr>
<tr>
<td>Visual presentation</td>
<td>-Emotionally evaluated and positioned layout designs and templates for ads (colors, shapes and textures) may be utilized per type of user segment.</td>
</tr>
<tr>
<td>Structure</td>
<td>-Using emotionally evaluated and positioned narrative templates for creating emotionally engaging stories.</td>
</tr>
</tbody>
</table>

Table 1. Technological possibilities of Psychological Customization in emotionally adapted gaming.

Based on Table 1 it may be then said that Psychological Customization in creating emotionally adapted templates for adapted gaming may be used in several quite feasible ways. Further, while the use scenarios presented are somewhat pushed to the extreme corners of possibilities of Psychological Customization, they illustrate some aspects that may add value to the users of such systems in emotionally adapted gaming. In the example provided the system offered more opportunities to be playful, have fun, to have more arousing experiences and to be able to control children’s experiences of games.
databases and actually working real-life systems for creating systematic psychological effects, i) the rule-databases may have to be adapted also locally and culturally, ii) the method needed to create a rule-database is not easy to use and may be suspect to ecological validity (eye-tracking, behavioral and psychophysiological measures, self-report, field tests are needed to verify laboratory results etc.) and iv) if the system works sufficiently it may raise privacy issues, such as the intimacy of a personal psychological user profile (personality, cognitive style, values, other) or even issues related to mental health and mind-control (such as clinically depressed patients using emotional controls of the game to induce negative emotional states and become even more depressed or children experiencing unwanted emotional states).

These issues should be addressed in order to achieve technically, commercially and ethically feasible systems for emotionally adapted gaming. Regarding future research, we aim to build, evaluate and field-test prototypes of emotionally adapted gaming. Finally, it is evident that the area of Psychological Customization of gaming is still in its infancy. International-level collaboration of multidisciplinary research groups focused on gaming is much needed to fully realize the framework proposed in this article.

References


Saari, T., Turpeinen, M., Laarni, J., Ravaja, N. and Kallinen, K. User experience-based adaptation of form of information in mobile contexts for mobile messaging. A manuscript submitted to publication.


How Putting Yourself into the Other Person’s Virtual Shoes Enhances Collaboration

Maria Spante, Ralph Schroeder, Ann-Sofie Axelsson, Michael Christie*
Department of Technology and Society, Chalmers University of Technology, Gothenburg Sweden
*CKK, Center for Learning and Communication, Chalmers University of Technology, Gothenburg, Sweden
{marspa@mot.chalmers.se, ralsch@mot.chalmers.se, asax@mot.chalmers.se, michael.christie@ckk.chalmers.se}

Abstract
Working together in collaborative virtual environments (CVEs) with different systems is often problematic; different capabilities can lead – among other things – to an unequal distribution of tasks and to misunderstandings. This paper reports on an experiment in which participants worked together on a task and switched halfway through from an immersive system to a desktop one (and vice versa) – and exchanged ideas about this with their partners afterwards. It was found that there are several advantages in collaborating in connection with experiencing different and unequal systems: partners learn not only about the strengths and limitations of the different systems, but also about collaborating with others and about the implications of using different technologies. The paper concludes with the implications of this ‘good inequality’ for the design and use of CVEs.

Keywords—virtual environments, collaboration, copresence, learning, immersive projection technology, desktop systems.

1. Background and Aim
Online collaboration is becoming increasingly common, and collaborative virtual environments (CVEs) are an important part of this development [1]. When people work together at a distance via computers or in CVEs, they often use systems with different capabilities, typically without being aware of the type of system their partner(s) are using.

The consequences of this asymmetry can be positive or they can go unnoticed: partners may ‘divide the labour’ between themselves, taking on different tasks without being aware of this. Or the consequences may be problematical insofar as the inequality between systems may lead to differences in leadership in carrying out a task or in status – again, without being aware that this inequality has been introduced or shaped by the technology.

One of the reasons for these effects is the absence of social cues in computer-mediated communication and other media. This effect has been studied at least since the research of Short, Williams and Christie [2], who made comparisons between different communication technologies and face-to-face communication in terms of ‘social presence’. CVEs are potentially a ‘richer’ medium in terms of social presence since they support a sense of ‘being there together’ in an environment other than the one in which you are physically present (the commonly used definitions of presence, copresence, and social presence; see [3], [4], [5]).

In this paper we investigate collaboration in VEs, and in particular whether having the experience of different systems can enhance – rather than being a problem for – collaboration. The question was investigated by means of an experiment in which participants switched system halfway through the task. Their comments afterwards – both about their individual experience and about their experience with others – show that ‘putting yourself in the other person’s virtual shoes’ can be a valuable learning experience in CVEs, and perhaps in online collaboration generally. Since asymmetrical setups will continue to be commonplace in the everyday usage of CVEs, it is important to study how this inequality can be managed in supporting online collaboration. The aim of this study is thus to investigate how people experience to solve a task together using both immersive and non-immersive systems, and how changing systems influences their experience of the two systems and their sense of the collaboration.

1.1. Previous Research
Previous research about collaboration in VEs with different systems has shown a variety of effects. A study by Slater and colleagues [6] of small group collaboration showed that the person using an immersive system (in this case a head-mounted display) was considered to be the leader in a group working together with two persons on desktop systems where they did not know what type of system the others were using. There were similar findings in studies by Schroeder and colleagues [7] for pairs working together with one person in a Cave-type or immersive projection technology (IPT) system working with a partner using a desktop system. Again, the person
using the IPT was considered the leader and as contributing the greater share to the task, even though there was no such leadership or unequal contribution when doing the same task face-to-face, desktop-to-desktop, or IPT system networked with another IPT system. An additional finding was that the participants ‘naturally’ divided the task or the labour between themselves, with the immersed person taking a more active role in the spatial aspects of the task and in manipulating objects, whereas the desktop person took a more ‘supervisory’ role – again, without being aware of the differences in the system that their partner was using and without being aware that they were ‘dividing the labour’ between them [8]. Finally, Axelsson [9] has analysed the findings of studies from both immersive and desktop systems and pointed out the problems when people in shared virtual environments are not aware of the status differences introduced by asymmetrical technologies.

A further study that is relevant here are the findings of Hindmarsh et al. [10] about problems of working together on networked desktop systems on a spatial task because of the restricted field of view and also because partners are not aware of what their collaborators can and cannot see. Heldal et al. [11] by contrast found that this problem on the whole does not apply with collaboration in networked IPT systems.

In contrast to these studies with their between subject design, the following study used a within subject design, giving users a first hand experience of both types of technologies. This is important to investigate since, in practice, people will often be confronted with asymmetrical setups.

2. Method

The experiment used a within subject design. 18 subjects arranged into nine pairs participated in the experiment. Each pair met in a virtual environment to solve a Rubik's-cube type puzzle (see figure 1) using an immersive and a non-immersive system. The trial was limited to twenty minutes and subjects changed systems half way through the trial. The subjects were 17 postgraduate students who were taking a pedagogical course and one teacher (the quotations below are translated or taken from subjects whose English was not their first language – hence some quotes have awkward phrasing). There were 4 females and 14 males from various disciplines at a technical university. The subjects had all seen each other during the course, but they had no previous experience of working together.

2.1. Technology and Task

The immersive system used was an IPT system, a 3x3x3 meter TAN VR-CUBE with stereo projection on five walls (no ceiling). The application was run on a Silicon Graphics Onyx2 Infinity Reality with 14 250MHz R10000 MIPS processors, 2GB RAM and 3 Infinite Reality2 graphics pipes. The participant wore CrystalEyes shutter glasses and used a 3-D wand for navigation. A Polhemus magnetic tracking device tracked the head via the glasses and the hand via the wand.

The non-immersive desktop system consisted of a Silicon Graphics O2 with one MIPS R10000 processor and 256MB RAM and a 19-inch screen display.

The dVise 6.0 software was used.

With the IPT systems, the subjects could move the blocks or cubes by putting their virtual hand into the virtual cube and pressing the button of the 3-D wand. On the desktop system, participants could navigate by moving the middle mouse button and could select the cubes by clicking on them with the left mouse button. To move the cubes, they had to keep the right mouse button pressed and move the mouse in the desired direction. They could also rotate the cube by pressing the right mouse button combined with the shift key. The movements of the avatar in the desktop system that was transmitted through the technology showed only the position of the avatar (no pointing) in relation to the virtual objects, visualized with a static avatar, whereas the avatar in the immersive system was dynamic and represented the user’s tracked movements.

Both the IPT and desktop systems allowed the participants to ‘mark’ the cubes by selecting them, which made their outlines appear as dotted lines (which was also visible to their partner).

Audio communication was via headsets with microphones and earphones.

The task was to solve a puzzle involving 8 blocks with different colours on different sides and to rearrange the blocks such that each side of the finished cube would display a single colour. The colours on the sides of the 8 blocks were red, blue, green, orange, yellow, white, and black.

Figure 1 The Rubiks Cube Puzzle

2.2. Procedure and Experimental Design

Before the trial session started, all subjects were given verbal instructions about the experiment. They had 5 minutes to familiarize themselves with the system, but were not allowed to start communicating with the partner. The total time for doing the task was 20 minutes and they changed system half-way through. Post-trial interviews with individual users took between 5 and 15 minutes and small focus group discussion of between 4-6 participants too between 45-60 minutes.
2.3. Data Collection and Analysis

Apart from the interviews and discussions, observations were also made during the trial. Post-trial interviews and discussions were audiorecorded and the trial sessions were audio- and videotaped.

The analysis presented here is based on the transcribed post-trial interviews. We were particularly interested in the subjects’ experience of changing systems and how that influenced their experience of the technology and their perceived sense of collaboration. We used content analysis as presented by Altheide [12] to interpret the interviews. As we were particularly interested in how subjects had experienced using the systems and working with their partner, we asked them about this. This allowed us to lift this theme out of the transcribed text, and to divide this theme, in turn, into technical and social aspects. These themes were subdivided further according to the following schema:

![Figure 2 Schema of interpretation](image)

3. Method

In the following section we present, first, the material related to technology, followed by social aspects. Quotes from subjects are coded as follows: I1 represents a quote from the first subject starting the trial in the IPT system, and D1 from the first subject starting the desktop system etc.

3.1. Experience of Using Different Systems:

Technical Aspects

One question concerned how the subjects experienced to use the two different systems during the trial. Typically the subjects remarked on the different technical functionalities. In general they experienced the IPT system as intuitive to use and manipulation of the cubes as easy. The desktop system was experienced as difficult to use because of the need for button pressing for manipulating objects. Typically subjects commented on similarities and differences:

“I started at the workstation [referring to the desktop system]...When I came down here [room with the IPT system] it was more intuitive in a way what I should do. I saw where I was and I had only one type of control apart from my own movements. The only thing I needed to do was to grasp and release. There I could twist and do the turns. At the workstation I had to do it with control –alt-shift or control-shift and a mouse click so it was harder for me to do the task there. I thought I noticed that for Sharon as well. When she came up there [to use the desktop system] she didn’t really know how to move either” (D1)

“One was more handicapped there [on the desktop system]. One can not do everything. The function one could have is to stand aside and look since one do not have the same functionalities. Here [IPT system] one is much smoother.” (D7)

“I think you get a better view from here, from the computer [desktop system], but you can not handle the things from here very well. I don’t know if there is a problem with a cable or with the special joystick. But I think when you are there on the stage [IPT system] it is easier to move and manipulate object.” (I1)

Regardless of which system they started with, subjects shared the same view. However, departing from majority opinion, one subject said:

“I think that the computer [desktop system] was easier, but one is used to computers. At the same time, here [IPT system], some things were easier to do here such as walking around faster and see the colors. That one could do faster down here, but the mere use of it, the computer was easier, but that is surely a habit issue, one is used to control [with the keyboard]. Here it was more of trying to see the hand and, well, click on the right spot. I experienced that as harder.” (D6)

Most subjects could intuitively and easily use the IPT system even though they had no previous experience with the technology. A few subjects expressed minor difficulties in grasping objects with their virtual hand using the wand, but the general experience was that it was easy to manipulate the virtual objects in the IPT system in comparison with using the desktop system.

Another experience linked to the technology was the difference in immersion. Even though the majority of subjects said that the IPT was more intuitive to use in comparison with the desktop for actions such twisting and turning the virtual objects, some subjects experienced the IPT as too immersive:

“My feeling is that you can manage better the system trough the computer [desktop system]. That cube [IPT system] is...it...cause a lot of
difficulties. You feel surrounded by cubes. And you sort of…you can grab one of them. But for me from this monitor I can see everything and probably I can manage my tools. I think so.” (I13)

Another subject said in the same vein:

“There they are all around you [talking about the virtual cubes] so it is hard to get a real overview.”(I16).

Even subjects that did not mention the feeling of being “surrounded by cubes” said that the desktop provided a better overview of the cube puzzle. The following two quotes illustrate the general view of the subjects about the advantage of using the desktop in comparison with the IPT system:

“I think it was easier in front of the desktop using the mouse and keyboard to have an overview and perhaps help out a bit and check it out and think a bit”(D9).

“I think you get a better view from here from the computer but you can not handle the things from here [the desktop system] very well”(I11).

It was not the case that subjects were in total favor of one system compared to the other. The subjects appreciated the two different systems in different ways. The different technical functionalities of the IPT and desktop systems were useful for different purposes in solving the puzzle together. The IPT system was experienced as useful for manipulating objects and the desktop system for getting a clearer overview of the puzzle. At the same time, the different technical functionalities also caused different types of difficulties: the desktop was associated with manipulation difficulties with problems moving objects because of the need for button pressing; in the IPT system, on the other hand, it was difficult to get an overview of the puzzle. This yields the following picture:

<table>
<thead>
<tr>
<th></th>
<th>IPT</th>
<th>Desktop</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manipulation</td>
<td>Easy</td>
<td>Difficult</td>
</tr>
<tr>
<td>Overview</td>
<td>Difficult</td>
<td>Easy</td>
</tr>
</tbody>
</table>

3.2. Experience of Using Different Systems: Social Aspects

The subjects were also asked a number of questions concerning their experience of using the two systems in relation to social aspects. We will present, first, their views of collaboration and then their views about communication.

Most of the subjects experienced the trial as a highly collaborative situation and expressed themselves in a positive way about working together. Regardless of whether they solved the puzzle or not, working with a partner was a good thing. But although 11 of 18 subjects commented in a positive way about collaboration, three felt that they could have solved the task without a partner. Only two reported that their collaboration was not working.

Most thought that their collaboration improved after having changed systems, and thought that they used this knowledge about how the different systems worked to improve the collaboration:

“I thought it [the collaboration] worked well. I thought it worked very well when one knew, when one had tried out each other’s tools. In the first instance one did not know what kind of capabilities the other had. I noticed that he could move around much easier but I did not know if that was because of him being better to manage the terminal or what it was. I didn’t know that he was down here [IPT system], that he had a tool like this. It became much easier after, when one knew, then we could divide the work better between us.”(D7)

“I think the collaboration with my partner was really fruitful and especially because we had two different views. From the computer I can see above better, and he can handle better the cubes, so I think the collaboration is necessary to solve the task faster.” (I11)

“You know, we started with no strategy at all. That was actually bad because we didn’t see what next. But during this final stage we understood better each other and that was a relief.”(I13)

Even subjects who thought they did not really make active use of their knowledge about the different technologies believed that it would have improved their collaboration.

“I think it is better than working with my self because obviously there some task that is more difficult to do from computer but its easier for me to move around and he can turn around more easier different side of the block. So I think that it is good to compensate but all we need is to have a better plan if we know the task earlier or in the middle. We should have some time in the beginning to just [talk about] - how we should do the task?...It’s easier to control when you sit here in front of your computer [desktop system] of course. So maybe it is good to have a strategy and then do some work from the computer first and then do down stairs [IPT system] to make the detail.” (I10)
In relation to the experience of collaboration, most subjects thought that collaboration was useful and it was improved by the fact that they changed systems. The change of system led to an increased understanding of each other’s perspectives and capabilities. This understanding enabled them to divide the labor based on the capabilities of the technology - such that the IPT person manipulated the objects and took a more active role in moving the objects, and the desktop person had an overview and took a more 'supervisory' role.

As for communication, subjects regarded verbal communication via the audio channel as crucial, but they also considered it to be as important to see their partner’s avatar movements and actions. Typically they commented about the way their partner moved around in the environment. In particular, those who started off using the desktop system found it remarkable that their partner moved around so easily or smoothly in comparison with themselves:

“In some way I realized that he had a different tool. One understands that at once when one see how smoothly he can move. One understood that quite quickly. Then it took a while before we talked about what kind of tool the other one had, but that became obvious when we changed.”(D7)

This quote also illustrates how some subjects attributed the differences in movements to the technology without knowledge about the differences between the systems. Some subjects, however, associated this to a difference in their partners’ skills:

“I thought it was a superman I had met that could do exactly as he pleased with his keyboard.”(D3)

The ability to refer to objects by pointing to or moving them facilitated communication about which object was being handled at the moment. However, the ability to refer to objects was different in the two systems, as one subject noticed:

“He was there in a way. It was really hard to express when one was upstairs [desktop system]. Then one had to grab a cube and say – “I am over here”. But for him [in the IPT] he could say “where I am”, or, “where I am going”. In some way he was there but I was not.”(D8)

This example indicates the subtle mix of verbal and non-verbal communication in a CVE with an asymmetrical setup. Referring to objects depended on the system: in the desktop systems it was necessary to grab a cube and also mark it visually to indicate to the partner what object one is talking about. In the IPT, subjects could point in referring to a cube. Movement could be conveyed by means of the dynamic avatar, which was not possible on the desktop system. Not only was action more intuitive in the IPT system, but language use was more intuitive in the sense that “here” and “there” could be conveyed through the interface as in the physical world. Subjects realized that knowledge of the different system also improved the way they communicated:

[Changing systems halfway through] “was fun. One could see these different possibilities. But that also made, given that one had tried both systems, one could more easily communicate with the systems and [also] communicate better with each other.”(D6)

Changing systems was thus important for a better understanding of each other’s possibilities and constraints, which helped subjects to agree about who should do what.

Finally, the experience of collaboration and communication is also reflected in subjects’ comments related to ‘being there together’ or copresence:

“Without voice communication it would have been difficult, so it was crucial.”(D8)

However, this same subject also felt that he sometimes forgot his partner when he was in the IPT when working with the objects:

“But also, since I did not see him, or rather he was over there so to speak, he was not close to the cubes. Then it was very easy to forget [him] …not until I was working alone I thought: oops, now I’m doing too much!”(D8)

This quote also illustrates that it is the position of avatars that subjects responded to, and in this case, when the subject was preoccupied with the cubes, he felt that he lost awareness of his partner.

4. Discussion

Previous studies have highlighted the disadvantages of asymmetrical setups, but as we have shown, if users can become aware of the differences entailed by different systems, this can also provide users with a better understanding of the possibilities and constraints of different technologies and thereby enhance collaboration. The view of our subjects was clearly that changing systems was positive; they thought that they could collaborate and communicate better after the change. They also realized that they made use of the different capabilities of the technology and could improve their strategy for solving the task by making use of these different capabilities. They recognized that the IPT gave them better possibilities for object manipulation and the desktop system gave them a better overview.

Interestingly, subjects recognized the benefits as well as the drawbacks of each system. Their better understanding thus not only made them aware of each other’s possibilities and constraints, but also enhanced their interpersonal interaction. These were insights that in some cases they were able to implement during the second sessions after switching systems for the first time, and in other cases this occurred to them only after both sessions
and they would have been able to implement what they had learned in future collaboration. (One idea suggested by this study is that this type of collaborative exercise with unequal systems could be a good pedagogical tool for learning about the difference that different technological capabilities can make.)

Before we discuss the implications of these findings further, it will be useful to recall some of the disadvantages of unequal systems. One of the main disadvantages is that partners may not be aware of the different capabilities that they and their partners have. This can lead to misunderstandings in communication, to adopting a poor strategy for collaborating on the task, and it raises the issue of fairness – at least in the case where two people should be collaborating on an equal basis and yet they are unaware of their different capabilities.

Another point to make here is that overcoming these disadvantages by ‘trading places’ will often not be possible. The point of distributed work is that partners can work together at a distance, but even if many people will be familiar with (or have access to) desktop systems in which they can collaborate in a 3D environment and where they can manipulate objects with a mouse, the same cannot be said for immersive systems. And although immersive systems may become more accessible in the future, they may also become more powerful and remain costly. Further, the general point – that different capabilities may be more invisible in CVEs than in the case of other technologies - will still apply to asymmetrical systems (since this asymmetry or inequality will not be conveyed through the interface) regardless of how accessible they become (and there are likely to be some asymmetries in most systems). Finally, it is worth mentioning that although for many tasks, symmetrical or same system setups will be an advantage for collaboration, for other tasks, there may be advantages for two or more participants to have different technologies and play different roles (for example, when people need to perform different complementary tasks).

One result of using of unequal systems is that, as mentioned earlier, when collaborators are unaware of the type of system that their partner is using, they may divide the labour between themselves – again, being unaware that they are doing this. In this study, when the participants did know the reason for their unequal participation, they said that they could make use of this knowledge to figure out a better strategy to do the task. In other words, creating a ‘common ground’ in a situation of missing social cues allowed them to collaborate better [13]. When they did not know about the technical reason for the different behaviors that was conveyed through the interface, on the other hand, they interpreted the difference as either personal (working with a more skilled person) or technical (working with different type of system). It can therefore be seen that technology is not only a tool for social interaction, but also an important feature in social interaction [14].

**Conclusions**

The finding that collaboration can be enhanced by awareness of different technological capabilities will have obvious relevance for the design of systems and their uses: how can systems provide knowledge about different capabilities? Can it be built into systems, or should task sessions be structured so as to allow for ‘wearing the other person’s virtual shoes’? This study will also have ramifications for research on presence and copresence: could the sense of copresence be enhanced – or possibly diminished - with the awareness that one’s collaborator may have a different affordance for copresence? (One reason for mentioning the latter possibility is that it has been shown that a person’s ‘copresence’ may not only be related to one’s own system, but also to one’s partner’s sense of ‘presence’ and ‘copresence’ in their system, [7]).

One obvious suggestion that one might be tempted to make in response to this study is that the differences between the capabilities of the technology should always be made obvious to users, and that this can easily be done with CVE technology. For example, avatars could have labels that specify what type of system and input/output devices they are using. Note, however, that this solution would also have drawbacks: apart from possible creating a cognitive overload on the users’ part (how much such information could the user ‘take in’?), it may also be that in focusing on figuring out what capacities they and their partners have, collaborators may lose the advantage of ‘naturally’ dividing the labour between them and thus add to rather than improving the time they take for the task.

A limitation of our trial was that it was short and subjects had only one chance to solve the task together. It may be that collaborators are able to adapt to the different capabilities, or to the absence of social cues which make them aware of these differences, over the course of time [16]. It would be interesting in future research both to test whether such adaptation takes place, and whether longer sessions with different systems could mitigate the need for ‘trading places’ or if this could be even more valuable with longer sessions. It would also be interesting to find out whether simply communicating the different capabilities verbally or by the partner’s demonstrating them to each other remotely could be just as effective as experiencing the different systems.

To sum up: putting yourself into the other person’s virtual shoes can enhance the interaction and the strategy in a collaborative task, as well as providing people with valuable insights into the use of CVE systems. In other words, ‘the good inequality’ can enhance collaboration. As the discussion has shown, however, such a setup for ‘trading places’ may not always be possible or desirable to implement. Future research will show under what circumstances the experience of different technological capabilities can be useful – for users, for research about CVE systems and their design, and when VEs are used for practical collaborative work beyond the trial setting of this study.
Acknowledgements

This research was part of the project ‘Living in Virtual Worlds: Implications of the Uses of Virtual Reality in Long-Term Settings’ supported by VINNOVA, and the Chalmers C-CeLT project ‘Engendering good learning’.

References


Healing Media: The moderating role of presence in restoring from stress in a mediated environment

Anne-Mie Sponselee, Yvonne de Kort, & Anneloes Meijnders
Eindhoven University of Technology, The Netherlands
{y.a.w.d.kort@tm.tue.nl, a.l.meijnders@tm.tue.nl}

Abstract

Restorative environments are environments that can help restore directed attention or reduce emotional and psychophysiological stress. The present study investigates the importance of presence in relation to restoration. We hypothesized that presence would moderate the stress-reducing effects of a mediated restorative environment. After performing a stress-inducing task, participants watched a restorative film on either a small or big screen. Physiological measurements (heart period and skin conductance level) were taken throughout the experiment. In addition we measured self-reported affect and presence using the ITC-Sense of Presence Questionnaire. Significant results appeared for skin conductance level (SCL) showing an interaction between screen size and time on restoration. No remaining effects of screen size were found on restoration.

1. Introduction

Hassles and troubles of everyday life make us feel stressed and mentally fatigued. In order to reduce the unhealthy consequences of stress, it is important to improve our mental, affective, and physical state. In the recent past there have been numerous studies indicating the beneficial effects of restorative – often natural - environments. These are said to help people reduce psychophysiological stress and renew attentional and other psychological resources [1].

Unfortunately, restorative environments are not always handy. An alternative solution would be to restore in a simulated (e.g. virtual) or mediated environment (e.g. slides and film). In fact, most of the studies performed in the research domain of restoration make use of simulations of environments instead of actual ones and have been effective in changing mood.

The present study investigates the importance of presence in relation to restorative environments. We hypothesize that presence – the sense of actually being there in the presented environment – may moderate the effect the restorative environment has. It might increase the psychological feeling of being away from work routine, demands, and obligations. This feeling of ‘being away’ is considered one of the central components of a restorative experience.

2. Theoretical background

2.1. Restoration

Restoration involves renewing diminished functional resources and capabilities [2]. It enhances the ability to focus attention [3], it reduces stress [4], and leads to positive affective states [5]. Although the field has not reached consensus regarding the basic mechanism behind these effects – two prominent theories coexist, one claiming recovery from psychophysiological stress as the central process [4], the other recovery from directed attention fatigue [6] – the number of publications reporting restorative effects of certain types of physical environments is growing.

In his psychoevolutionary theory, Ulrich [4] focuses mainly on the visual perception of certain environments, and the aesthetic and affective reactions associated with it. He emphasizes that most affective reactions are precognitive. The visual properties influencing aesthetic preference and interest Ulrich discusses are: complexity, structure, depth, ground surface texture, threat or tension, deflected vistas, and water. According to Ulrich’s evolutionary approach, such surroundings optimally support approach and avoidance behaviors which are relevant for people’s well-being and survival. The experience of visually pleasant physical surroundings is thought to reduce stress by eliciting positive emotions, sustaining non-vigilant attention, restricting negative thoughts, and returning physiological arousal to more moderate levels [1], see also [7].

Restoration Theory (ART, Kaplan & Kaplan [6]) postulates that to direct or focus attention requires a certain capacity that can become depleted, resulting in directed attention fatigue. This resource can slowly regenerate in the absence of an ongoing need for directed attention, but ART
suggested that it will be restored faster by what Kaplan and Kaplan have called ‘involuntary attention’. Involuntary attention is attention that is drawn by stimuli that are fascinating in themselves, not requiring the resources directed attention does. Fascination is drawn by stimuli that are reasonably complex, coherent, and legible and yet hold some mystery. Fascination however, is one necessary but not sufficient prerequisite for restoration. Ideally, the environment should also afford a feeling of ‘being away’, have ‘extent’ and be compatible with the viewer’s wishes and capacities.

Although these two theories attribute the restorative effects to different processes, both support the idea that nature functions well as a restorative environment. These expectations have been supported by a number of empirical studies, reporting stronger reductions of negative feelings such as anger and aggression and stronger increase of overall positive affects such as happiness, friendliness, or elation after viewing natural vs. urban scenes [3,8,9,10]. In addition, similar empirical evidence exists of physiological restoration in terms of skin conductance, muscle tension, and pulse transit time and blood pressure e.g. [7,10,11,12].

Only a small number of studies on restorative effects have actually taken participants on a visit to natural places, e.g. [3]; most of these studies have been performed in psychological laboratories, employing photographs, slides, or videos, under the implicit assumption that this will result in similar effects as experiencing the real environment in its full sensorial richness – in other words assuming experiential isomorphism for mediated and unmediated stimuli. Strikingly, only little or attention is paid to optimizing the mediation or experience of actually being there.

### 2.2. Presence as a moderator

In using mediated or simulated environments for evaluative and therapeutic purposes, response similarity with regard to real environments is considered a prerequisite, e.g. [13,14,15]. The presence experience thus becomes an important means – even the key - to valid and effective use of mediated environments, following the response similarity approach [16] stating that ‘it is reasonable to expect that as the fidelity of the displayed environment increases, responses to that environment will be increasingly similar to responses we exhibit to the same objects, agents and events in real environments’ (pp 202).

This same assumption underlies the use of psychophysiological measures as potential objective indicators of presence, as for instance suggested by [17, 18], and studied by – among others – [13,19] (for a thorough review see [20]. If a real environment causes certain (psychophysiological) responses, then the similarity (or often strength) with which a mediated environment engenders the same effects can serve as an indicator of the amount of experienced presence in that environment. Here presence is the ultimate goal and psychophysiological measures are a means to measure presence.

The present study adheres more strongly to the rationale presented in the first paragraph than that of the second one (though both share common grounds and in fact represent two sides of the same medal): in parallel to the use of mediated or simulated environments in therapeutic settings – e.g. the treatment of phobias [21] in [13] – we expect that treatment in a restorative environment will be more successful as the person experiences more presence. Presence thus is a means to enhance restorative effects. An experiment was set up to test this assumption, in which we manipulated presence in a restorative environment. Both self-report measures of affect and psychophysiological measures were taken to investigate moderating effects of presence on restoration. We hypothesized that as presence in the mediated environment increased, restorative effects would become stronger.

### 3. Method

#### 3.1. Design

The effect of presence on restoration was studied in an experiment in which after a stressful episode, participants watched a film of a restorative environment under low or high presence conditions. Presence was manipulated between-subjects by varying screen size. In the low presence condition the film of the restorative environment was shown on a small screen, whereas in the high presence condition it was shown on a large screen. To assess whether the restorative effect of the nature film depended on the level of presence, we measured changes in participants’ skin conductance level (SCL), inter beat interval (IBI), and self-reported positive and negative affect.

#### 3.2. Participants

A student sample of N = 80 participated in the experiment, of which 36% was female and 64% was male. Their mean age was 23. The participants were paid € 8 in exchange for their effort. Participants were randomly assigned to the two presence conditions.

#### 3.3. Setting

The experiment was conducted at the Eindhoven University of Technology, in a laboratory. Participants were seated at a table in front of a large back projection screen. The back projection screen of 110 by 145 centimeters (72”) was positioned with its center on eye level, at a 2.25 meters distance approximately. On the table was a computer on which participants received instructions, performed the stress task, and filled in the self-report measures.

#### 3.4. Stressor task

In this experiment we made use of the Markus & Peters Arithmetic Test (MPATest, [22,23]) to increase participants’ level of stress prior to exposing them to the film of the restorative environment. This stressor consists of
a mental arithmetic task in combination with uncontrollable industrial noise. The effectiveness of the stressor has been confirmed by previous research, showing that the stressor brings about heightened heart rate [24], increased skin conductance and a negative mood [22,23]. The stressor took between 18 and 32 minutes. The variation in duration was due to the time participants took to read the instructions, and the first exercises of the task, as the task only continued after three correct answers. The actual stressor took about 16 minutes, as it included 16 1-minute trials.

3.5. Restorative film

In the two presence conditions the same 10 minutes film without sound was presented. The film was a compilation of nature scenes from two DVD’s created under the authority of “Vereniging Natuurmonumenten”, a Dutch nature reserve association. The duration of the restorative film was exactly ten minutes.

3.6. Presence manipulation

We chose to manipulate presence by varying screen size, because this manipulation does not change the content of the sensory input. In the low presence condition the restorative film covered 47 x 60 centimetres (31”) of a 110 x 145 centimetres (72”) screen. In the high presence condition the film covered the entire screen.

3.7. Presence measure

The subjective state of presence was measured using the ITC-Sense Of Presence Inventory (ITC-SOPI, [25]). This inventory taps four different factors (spatial presence, engagement, ecological validity / naturalness, and negative effects) with 44 items in total. The items are statements and participants are asked to indicate their degree of agreement with these statements on scales ranging from 1 “Strongly disagree” to 5 “Strongly agree”. The reliability of the subscales was acceptable to high (alpha’s from .60 to .92).

The questionnaire was administered immediately after the film.

3.8. Psychophysiological measures

Skin conductance level (SCL) was recorded directly using the constant voltage technique. A BioPac Electrodermal Activity Amplifier Module (GRS100B) measured the absolute skin conductance for every 5 milliseconds (200 samples/second). The lowpass filter was set to 1 Hz, the gain was set to 20µmho/V. Conductance was measured from the non-dominant hand by placing BioPac Electrodermal Activity transducer (TSD103A) Ag-AgCl electrodes on the first phalange of the index and middle fingers. A non-irritating electrode gel (Parker Signa gel) was used as the electrolyte.

Heart period (inter beat interval, IBI) was derived from an electrocardiogram (EKG), which was recorded from two BioPac Ag-AgCl disposable shielded electrodes (10 mm contact area) placed on each wrist after preparing the skin with alcohol. Another unshielded electrode was placed on an ankle for the ground. To obtain heart rate, the EKG signal was amplified with a BioPac Electrocardiogram Amplifier module (ECG100B), which detected the occurrence of the ‘R’ wave. The ‘R’ wave detector circuitry consisted of a high Q (Q = 5), 17Hz band pass filter followed by a full wave rectifier, followed by a 10.0Hz three pole, low pass filter. The gain was set to 500 (40mV), the high pass filter switched to 0.05Hz the sample rate was 200 samples per second.

The physiological measures were taken during a baseline period, during the stress episode, and during the restoration episode.

3.9. Affect measures

An affect questionnaire was developed to measure positive and negative affect. The entire questionnaire consisted of 16 affect words. These affect words appeared in random order one by one on the computer screen, embedded in the sentence: “I feel ...”, followed by a 7-point answering scale, ranging from 0 “Not” to 6 “Very much”. Examples of positive affect words are “relaxed” and “cheerful”. Examples of negative affect words are “tense” and “irritated”. The mean score on the eight positive items was used as a measure of positive affect (alpha = .73) and the mean score on the eight negative items was used as a measure of negative affect (alpha = .71).

Participants completed the affect questionnaire three times during the experiment: prior to the stressor, after the stressor, and after the restorative film.

4. Results

4.1. Effectiveness of stress task

To test whether the stress task was effective in eliciting stress, we conducted a repeated measures analysis of variance with skin conductance level as the dependent variable and Screen size and Time of measurement (baseline and at the end of the stressor) as the independent variables. This analysis showed a significant effect of Time of measurement, $F(1, 78) = 108.84, p = .00$, while the interaction effect with Screen size was insignificant, $F < 1$. The stress task increased skin conductance level (baseline $M = 10.57$ mumho, $SD = 6.06$, stressor $M = 14.92$ mumho, $SD = 8.25$).

A similar analysis was performed with heart period measured during baseline and during the stressor. Again, the effect of Time of measurement was highly significant, $F (1, 78) = 19.56, p = .00$; the interaction-effect with Screen size was insignificant, $F < 1$. The stress task decreased heart period (baseline $M = 0.85$ s, $SD = 0.13$, stressor $M = 0.81$ s, $SD = 0.10$).

We also tested the effect of the stress task on negative affect. A repeated measures analysis of variance with negative affect as the dependent variable and Screen size and Time of measurement (before and directly after the
stress task) as the independent variables showed a significant effect of Time, $F(1, 78) = 244.34$, $p = .00$; the interaction-effect with screen size was insignificant, $F < 1$. Negative affect increased due to the stress task (before stressor $M = 2.00$, $SD = 0.62$, after stressor $M = 3.44$, $SD = 0.84$).

A repeated measures analysis of variance with positive affect as the dependent variable and Screen size and Time of measurement (before and directly after the stress task) as the independent variables also showed a significant effect of Time, $F(1, 78) = 196.73$, $p = .00$; the interaction-effect with screen size was insignificant, $F < 1$. The stress task resulted in a decrease of positive affect (before stressor $M = 4.05$, $SD = 0.61$, after stressor $M = 2.62$, $SD = 1.02$).

Hence all stress indicators point in the same direction, namely that the stressor task was successful in eliciting stress.

### 4.2. Effects of screen size on experienced presence

To examine whether the screen size manipulation had an impact on experienced presence, a multivariate analysis of variance was conducted with the four experienced presence factors (spatial presence, engagement, ecological validity / naturalness, and negative effects) as the dependent variables and screen size as the independent variable. This analysis showed no significant result, $F(4, 75) = 1.59$, $p = .19$. Hence screen size seems to have failed to influence experienced presence. It should be noted that levels of experienced presence were very low in both conditions. Spatial presence was rated $M = 0.81$ ($SD = 0.53$) in the small screen condition and $M = 0.95$ ($SD = 0.63$) in the large screen condition. Results are reported in Table 1.

### 4.3. Effect of screen size on restorative impact

To find out whether presence moderated the restorative impact of the nature film, a repeated measures analysis of variance was performed with skin conductance level as the dependent variable and Screen size and Time of measurement (before the film, and three times during the film) as the independent variables. A significant effect of Time of measurement was found, $F(1, 76) = 8.43$, $p = .00$, as well as a significant interaction-effect with Screen size, $F(1, 76) = 2.79$, $p = .05$. As can be seen in Figure 1, skin conductance level decreased faster and more in the large screen condition as compared with the small screen condition.

A similar analysis was conducted with inter beat interval as the dependent variable. This analysis showed a significant main effect of Time of measurement, $F(1, 76) = 63.99$, $p = .00$, but the interaction-effect with Screen size failed to reach significance, $F < 1$. This indicates that inter beat interval did not develop differently for the two experimental conditions. Results are shown in Figure 2. Means are reported in Table 1.

![Figure 1](image1.jpg)

**Figure 1** Skin conductance level (SCL) during restorative film, measured in mumho, corrected for last level during stressor (i.e. increase of SCL).

![Figure 2](image2.jpg)

**Figure 2** Inter beat interval (IBI) during restorative film, measured in seconds, corrected for last level during stressor (i.e. increase of IBI in s).

![Figure 3](image3.jpg)

**Figure 3** Positive and negative affect scores following the stressor and following the restorative film, for both experimental conditions.
To examine whether presence moderated changes in affect, a repeated measures analysis of variance was conducted with negative affect as the dependent variable, and Screen size and Time of measurement (before and after the film) as the independent variables. The main effect of Time was significant, $F(1, 78) = 134.76, p = .00$, but the interaction with Screen size was not, $F(1, 78) = 1.29, p = .26$, indicating that negative affect was lower after the restorative film than before, regardless of screen size.

The same analysis was conducted with positive affect as the dependent variable. Again a main effect of Time was found, $F(1, 78) = 14.66, p = .00$, but no interaction with Screen size, $F<1$. This indicates that positive affect was higher after the restorative film than before, regardless of whether it was a small or a large projection. Results for positive and negative affect are shown in Figure 3.

### 5. Discussion

The present study showed some results in the expected directions but left us with some questions as well. In this discussion we will try to address some of the most important issues, but would like to inform the reader that data-analyses – especially of the physiological data – are still ongoing. We hope to present the finalized analysis at the conference.

Unfortunately, the manipulation of screen size was not successful in producing a significant difference on our manipulation check, although all components of presence showed trends in the expected direction. This was striking, since the manipulation was considerable and screen size manipulations have been shown to influence presence significantly in earlier studies [e.g. 26, 27].

However, the screen size manipulation did show a significant interaction with time on skin conductance level. Although no main effect emerged, results did indicate that arousal – as indicated by SCL – followed a clear downward slope for the big screen condition and did not do so in the small screen condition. Additional analyses on the electrodermal data still have to be performed. This particularly involves investigation of skin conductance responses, like the number and amplitude of fluctuations in skin conductance.

Similar results did not prove significant on the heart period (IBI) data. Looking at the trajectories we are tempted to conclude that reductions of heart rate to baseline level had already been realized in the first phase of the restorative film in both experimental conditions. Judging from the average increase in heart rate (or decrease in heart period) during the stressor we expect that the stressor task – although it did show effects on all measures in the expected direction – was not strong enough to require a lengthy restoration period for IBI data. We plan to look into the trajectories of IBI data in more detail, especially in the first four minutes during the restorative film.

Results of affect measures also show complete restoration and return to baseline level after ten minutes for both experimental conditions. Again we suspect that a more stressful task – requiring more and longer restoration might have shown differences between the two presence conditions. In line with this finding, other researchers in the domain of restoration also report great difficulties in engendering high levels of psychophysiological stress or attention fatigue in experimental settings. Some are now turning to naturally appearing stress e.g. as exists among students directly after exams. Higher levels of induced experimental stress would certainly increase the potential to study effects of restorative conditions.

### Conclusions

In spite of the mixed results of the study, preliminary findings do point out the relevance of investigating moderating effects of presence in research on restorative environments. This and similar studies should prove useful for gaining a deeper understanding of all interrelationships between psychophysiology, restoration and presence and, in addition, may result in implications for the development of media technology that can help people restore from stress in settings as diverse as offices (work stressors), homes (restoring from daily hassles or negative life events), and

### Table 1: Means and SD of ITC-SOPI scales, SCL, IBI and positive and negative affect for both experimental conditions

<table>
<thead>
<tr>
<th></th>
<th>Small screen</th>
<th>Big screen</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spatial presence</td>
<td>M 1.81</td>
<td>M 1.95</td>
</tr>
<tr>
<td></td>
<td>SD .53</td>
<td>SD .63</td>
</tr>
<tr>
<td>Engagement</td>
<td>M 2.56</td>
<td>M 2.77</td>
</tr>
<tr>
<td></td>
<td>SD .60</td>
<td>SD .63</td>
</tr>
<tr>
<td>Naturalness</td>
<td>M 3.41</td>
<td>M 3.57</td>
</tr>
<tr>
<td></td>
<td>SD .78</td>
<td>SD .51</td>
</tr>
<tr>
<td>Negative effects</td>
<td>M 2.17</td>
<td>M 1.88</td>
</tr>
<tr>
<td></td>
<td>SD .59</td>
<td>SD .51</td>
</tr>
<tr>
<td>SCL scores</td>
<td></td>
<td></td>
</tr>
<tr>
<td>corrected for last</td>
<td></td>
<td></td>
</tr>
<tr>
<td>stressor level:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCL restoration 1</td>
<td>-7.97</td>
<td>1.20</td>
</tr>
<tr>
<td></td>
<td>-0.565</td>
<td>1.22</td>
</tr>
<tr>
<td>SCL restoration 2</td>
<td>-6.16</td>
<td>1.68</td>
</tr>
<tr>
<td></td>
<td>-0.944</td>
<td>1.59</td>
</tr>
<tr>
<td>SCL restoration 3</td>
<td>-.896</td>
<td>1.97</td>
</tr>
<tr>
<td></td>
<td>-1.009</td>
<td>1.83</td>
</tr>
<tr>
<td>IBI scores</td>
<td></td>
<td></td>
</tr>
<tr>
<td>corrected for last</td>
<td></td>
<td></td>
</tr>
<tr>
<td>stressor level:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IBI restoration 1</td>
<td>0.089</td>
<td>.068</td>
</tr>
<tr>
<td></td>
<td>0.102</td>
<td>.06</td>
</tr>
<tr>
<td>IBI restoration 2</td>
<td>0.089</td>
<td>.061</td>
</tr>
<tr>
<td></td>
<td>0.094</td>
<td>.06</td>
</tr>
<tr>
<td>IBI restoration 3</td>
<td>0.080</td>
<td>.066</td>
</tr>
<tr>
<td></td>
<td>0.083</td>
<td>.06</td>
</tr>
<tr>
<td>Positive affect post</td>
<td></td>
<td></td>
</tr>
<tr>
<td>stressor</td>
<td>2.55</td>
<td>1.12</td>
</tr>
<tr>
<td></td>
<td>2.69</td>
<td>0.92</td>
</tr>
<tr>
<td>Negative affect post</td>
<td></td>
<td></td>
</tr>
<tr>
<td>stressor</td>
<td>3.43</td>
<td>1.02</td>
</tr>
<tr>
<td></td>
<td>3.49</td>
<td>0.94</td>
</tr>
<tr>
<td>Positive affect</td>
<td></td>
<td></td>
</tr>
<tr>
<td>post restoration</td>
<td>3.68</td>
<td>0.65</td>
</tr>
<tr>
<td></td>
<td>3.87</td>
<td>0.74</td>
</tr>
<tr>
<td>Negative affect post</td>
<td></td>
<td></td>
</tr>
<tr>
<td>restoration</td>
<td>4.78</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td>5.04</td>
<td>0.81</td>
</tr>
</tbody>
</table>

Note: Skin conductance level (SCL) measured in mumho; Inter beat interval (IBI) measured in seconds; affect stores vary from 0 to 6.
even medical purposes as restoring from treatment-related and post-surgical stress.

Correspondence

Please direct correspondence concerning this article to: Yvonne de Kort, Human – Technology Interaction, Faculty of Technology Management, Eindhoven University of Technology, PO Box 513, Eindhoven, The Netherlands.

References


Comparing Differences in Presence during Social Interaction in Augmented Reality versus Virtual Reality Environments: An Exploratory Study

Arthur Tang, Frank Biocca, Lynette Lim

Media Interface and Network Design (M.I.N.D.) Labs,
Dept. of Telecommunication, Michigan State University, East Lansing, MI 48824
Tel: (517)355 5073 Fax: (517)355 1292 Web-page: http://mind.org
E-mail: tangkwo1@msu.edu, biocca@msu.edu, lynette@msu.edu

Abstract

In augmented reality (AR) environments, users experience the physical environment and other users directly along with mediated virtual objects embedded in the environment. In immersive virtual reality (VR), users experience of a visual environment (and sometimes other senses) is completely mediated. The representation of user’s body in virtual environments granted us a new research territory in dualistic interaction between the mind and body: how do the virtual body and user’s mind interact one another, and eventually affect user’s behaviors to the environment? An experiment was conducted to explore potential effect of users and interactant’s body to sense of presence in VR and AR environments. Results from the study suggest that the absence of representations of the user’s body in VR environment may lessen sense of spatial presence comparing with AR environment.

Keywords: Augmented reality, Presence, Body and Mind, Spatial cognition.

1. Presence and the continuum between mediated and unmediated experience

Presence, the sense of being there in a space, may be a psychological state associated with both mediated and unmediated experience; it is defined by some as the perceptual illusion of non-mediation [1, 2]. The user fails to perceive or acknowledge the existence of a medium and responds as if the medium is not there.

Media differ as well in the degree to which they mediate experience; that is the degree to which direct sensory experience of the environment is replaced with stimuli and representations generated by the medium. In the continuum between direct, unmediated perception of the environment and completely mediated perception, augmented reality (AR) is closer to unmediated perception than virtual reality (VR). In most typical configurations, VR and AR systems use head-mounted displays (HMD) and motion trackers to display virtual objects, environments, or other user. The first person perspective and tight coupling of the human body to the computer interface generate a compelling sense of being in the mediated space and the consciousness of user’s body in the computing environment [3].

Marsh and others have argued that the degree of presence of an experience is dependent on the transparency and continuity of the interface [4]. Transparency is the elimination of mediation [1]: the lack of consciousness of the medium itself. Continuity is lack of disruption during interaction. Disruption may occur when the user becomes overly aware of the medium and the physical interface. For example, the sense of presence in the narrative world presented by the television disappears when the observer becomes aware of the television set. Slater and his colleagues have refer to this as breaks-in-presence [5, 6].

In typical VR and AR interfaces, users interact with the medium in a three dimensional space instead of a two dimensional surface in typical interfaces and media. The elimination of the frame of the detached media, such as interface screen and paper, provides a higher transparency and continuity. The use of head motion and hands gesture to interact with the computer transforms the human body to the primordial communication interface to the environment [7]. VR and AR technologies show promise for the ultimate interface of high degree of presence.

1.1. Mediation and the experience of one’s body, other bodies, and the environment.

The representation of user’s body in the virtual environment granted us a new research territory in dualistic interaction between the mind and body: how do the virtual body and user’s mind interact with one another and affect user’s behaviors in the environment, and eventually the body schema of the user in reality? The sensory cues of direct perception and action within an environment help locate the body of the user [8]. The first person viewpoint of immersive VR and AR environments provide visual information about one’s view and head movement in space, but the rest of the user’s phenomenal body in immersive VR is usually largely absent or distorted. It is technically difficult to capture all the movement and features of the user and project it to the avatar in the VR environment in real time. Most VR systems also have a mismatch body parts position, physical scale, and motion scale between the phenomenal body in VR and the actual user’s body. For example, a VR user interface that use hand manipulation of objects typically shows only a limited virtual representation.
of user’s disembodied hands, neglecting the rest of the body. Those VR interfaces that implement a full body avatar for the user usually use a crude and primitive human-like figure. It may not be in the right size, shape, color and gender [9, 10].

The fundamental difference between VR and AR environments is that AR user is able to perceive the real world as well as the user’s body and the body of an interactant. Instead of replacing perception in the natural surroundings, AR systems augment the human visual channel with computer generated graphics. So instead of relying on proprioception memory, the user has real time visual perception of every body motion. It is observed that AR systems users are generally more confident in making body motion than VR systems users. The increased consciousness of the user’s body in AR environments also facilitates a more natural body movement.

Presence is part the sense of the location of one’s phenomenal body when coupled to a medium. Presence may be increased by not only the perception of the environment but by the experience of ones own moving body and other bodies in the environment. Slater found that adding a fully body avatar and providing even primitive arm movement for the full body avatar increased the sense of presence [11]. But this experiment looked only at the presence or absence of an avatar. It did not compare the direct experience of a fully natural body.

1.2. Research question: Does the presence of unmediated cues increased presence

AR environments allow us to explore the role of unmediated direct experience of the body and other cues increases presence. The authors are not aware of any previous study on the sense of presence using AR system. A study was conducted to explore the following research questions: Does the direct visual experience of the one’s body and that of partner increase the sense of presence? Or to put it another way, does the absence of visual cues of ones body and that of other in a virtual environments decrease presence?

2. Methodology

A within-subjects experiment was conducted to explore any potential difference in sense of presence between VR and AR environments. There was one independent variable, the interaction environment, with two levels: (a) AR Environment, and (b) VR Environment. The intent of the experiment is to measure the extent to which users are able to experience being physically present when communicating with a human in an AR environment or with a virtual agent in an AR environment.

2.1. Participants

16 participants from an undergraduate class at a university volunteered to participate in the study. None had previous experience in any VR/AR environment. 11 of the participants are male, and 5 are female.

2.2. Materials

2.2.1. Virtual Environments

The study was conducted using hardware and software that could create a virtual or an augmented reality environment. Stimulus materials were displayed in stereo using the Sony Glasstron LDI-100B. Subject’s head motion was tracked using the Intersense IS-900 ultrasonic/inertia hybrid tracker. Stereo graphics were rendered in real time based on the data from the tracker. Presentation of stimulus materials was written using ImageTcAR [12].

For the AR environment (see Figure 1), subjects directly experienced a simply black room where a set of virtual cell phones was present on physical table. Across the table was a partner, a confederate of the experimenter, directly visible and also wearing a head-mounted display. In the VR version of this environment (See Figure 2) all aspects of the visual environment were virtual. An avatar of the partner was presented in a virtual black room with a matching virtual table. In both cases, the participants interacted with the same female confederate.

Figure 1. A photograph taken from the see-through HMD in the AR condition. The cellular phones on the table were virtual objects but the confederate and environment were directly visible and unmediated.

Figure 2. Stimulus material displayed to the participant in the VR environment. In this environment, user’s body is not visible. The cellular phones, partner, and environment were all mediated and virtual. The virtual partner was a full body avatar puppeteered by the confederate.
2.2.2. Measures

Participants were administered the ITC-Sense of Presence Inventory (ITC-SOPI) to evaluate their levels of physical presence felt [13]. The ITC-SOPI is a validated 44-item self-report questionnaire that was used in this study to measure how physically located users feel within any mediated space, how the mediated environment compares to the real world, and how realistic the environment feels. The items generate four factors: (1) Spatial Presence – how physically present users feel in the virtual environment; (2) Engagement – how involved users would feel toward the content of the virtual environment; (3) Ecological Validity – the level of realism and naturalness of the environment; and (4) Negative Effects – any harmful physical effects, such as eye-strain or nausea, that users may experience by being within the environment.

2.3. Procedure

The experimental procedure consisted of an initial demographic questionnaire and two user study phases. After filling out the demographic questionnaire, participants were brought to a cylindrical room covered with black cloth, and were immersed into an AR and a VR environment in turn. The sequence of the environment presenting to each subject was randomized and counter-balanced.

In each of the environment, participants were asked to carry out a social discussion with a confederate about personal preference about two cellular phone models. A typical discussion includes comparison of colors, buttons, shape of the two cellular phone and other personal preferences. The set of cellular phones displayed in each environment is also randomized and counter-balanced. The participants were asked to fill out the ITC-SOPI questionnaire right after the discussion in each environment. Participants were told that the confederate is one of the subjects. All experimental sessions took less than 30 minutes.

3. Results and Analysis

Table 1 and figure 3 illustrate the four mean factor scores generated by the ITC-SOPI questionnaire. Resulted were further analyzed using repeated measures ANOVA.

The effect of spatial presence between the two treatment conditions is significant \( F(1, 16) = 5.33, p = 0.04 \). Users experienced greater spatial presence in the augmented reality condition.

On the other hand, comparing engagement, naturalness and negative effect factor scores between the two treatment conditions, no significant differences were found between the two conditions \( F(1, 16) = 0.34, p = 0.86; F(1, 16) = 0.55, p = 0.47 \) and \( F(1, 16) = 0.20, p = 0.66 \) respectively.

4. Discussion

4.1. Difference in spatial presence between AR and VR

The simplest explanation for the difference in spatial presence between AR and VR is that the AR condition is largely unmediated. Therefore, there are more sensory cues as to ones’ spatial location. But there may be more to this than the lesser level of sensory cues.

Human brain uses both sensory and motor information to construct an internal representation of the space we perceive. Research results from spatial cognition suggest that objects in the environment are represented in egocentric and allocentric references frames [14, 15]. Since all visual input is egocentric, all objects in the environment are originally egocentric before some of them are encoded and clustered into the allocentric reference frame.

We argue that a first person VR visual simulation without a user avatar interferes the behavior of the human brain’s visuo-spatial perception system. This may weaken user’s ability to perceive egocentric space accurately and negatively influence user’s sense of spatial presence. By extension, we would predict that adding a realistic body for the user in the VR condition would make the level of presence more similar.

4.2. Lack of difference in Engagement and Naturalness

We were originally expected that the ecological validity/naturalness factor score of AR environment would be significantly higher than that of VR environment. Although the AR environment does score higher on both dimensions, the differences are not significance. This result
6. Conclusions

An experiment was conducted to evaluate the difference of sense of presence generated by AR and VR environment using the ITC-SOPI inventory. There is some evidence for the proposition that users’ perception of spatial presence is higher in AR environments than in VR environment. We speculate that the failure to find a difference between the AR and VR environments on the naturalness dimension may be caused by a possible ambiguity regarding what is being referenced (evaluated) in the scale items of the ITC SOPI measure of presence used in this experiment. Because the focal content of the interaction was the virtual cellular phones, the users may interpret that the questions to refer to the VR objects and not the whole environment. This would need to be confirmed by debriefing.

7. Acknowledgements

This material is based upon work supported by the National Science Foundation under Grants No. 02-22831. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. We thank Betsy McKeon for her assistance in conducting the experiment.

8. References

Facilitating Interconnectedness between Body and Space for Full-bodied Presence
- Utilization of “Lazy Susan” video projection communication system -

Shigeru Wesugi, Yoshiyuki Miwa
School of Science and Engineering, Waseda University
{wesugi@computer.org, miwa@waseda.jp}

Abstract
A significance of a sense of togetherness or co-existence among remote communication partners has been pointed out in computer-mediated communication. In order to create such a co-existing situation, full-bodied presence of participants should be supported at each place. We devise two representation methods; representing self reflection in a shared video space, and projecting reflection of a remote partner onto the local tabletop. Then we construct “Lazy Susan” video projection communication system composed of a shared disk system and a video projection system based on this method. The results of experiments suggest our system can enhance an interconnectedness between self and a remote place, and simultaneously enhance a sense of co-existence. Based on this result we discuss and propose a novel approach focusing on duality of embodiment to facilitate interconnecting body with space for full-bodied presence.

Keywords--- Co-existing space, Embodiment, Self reflection, Visual interaction, Tangible interaction, Interconnectedness

1. Introduction
Many telecommunication systems have been proposed and developed in the past few decades. These systems can support remote awareness - who is participating, what they are doing, and where they are - by expressing bodily action of a remote partner such as gaze and gesture [1]. A problem of disembodiment, however, occurs in such computer-mediated communication as Dreyfus pointed out [2]. For example, even when a life-sized image of a remote partner appears on a large screen at a local place, the psychological distance to the person in the screen differs from that in a face-to-face situation [3]. We consider that image-only projection of a remote partner to create much less of a sense of presence. It is obvious to the local partner that the remote partner is not there.

In contrast, people are bodily present at the same place in face-to-face situations. This full-bodied presence is crucial to everyday human encounters [2]. We can attune ourselves to mood, when we are bodily present in a situation. The situation -we are embedded with others and things in a common setting- is called as “Ba” in Japanese [4][5]. “Ba” is not a physical place but a co-existing situation in mind. We believe that creation of a co-existing situation is fundamental before interpersonal communication takes place [4][5]. We define “co-existing space” as a common place at which people are full-bodied present.

A few communication systems have been proposed and received attention to support a sense of connectedness among remote families and a remote watching an elderly person living alone warmly in several years [6][7]. A requirement of communication technology creating a co-existing space between remote places will increase more than ever. In order to address this challenge, the first step is to enhance a sense of “interconnectedness” among remote people in a shared virtual environment. Therefore we devised a networked “Lazy Susan” communication system; integrating tangible interaction with physical disk and visual interaction in a shared virtual environment [8][9]. Consequently, we found its capability to enhance a sense of co-existence in a shared virtual space.

This paper describes our next step of developing this design method toward creating a virtual co-existing space in a physical place where our body exists. The paper explains two approaches and our novel “Lazy Susan” video projection communication system composed of a shared disk system and a video projection system; representing reflection of self in a common video space, and projecting reflection of a remote partner onto a local tabletop. The results of experiments suggest our system can enhance a sense of “being present in a video space of a remote place” and “a remote participant is being present at a local place”, and simultaneously enhance a sense of co-existence. At the last it discusses and proposes a novel approach focusing on duality of embodiment in order to facilitate interconnecting body with space for full-bodied presence.
2. Design of a virtual co-existing space

There are few design frameworks and approaches to create a virtual co-existing space, since most communication systems discussed in current literature are intended to support primarily how to convey bodily action such as gaze awareness [10][11][12] and gesture in a shared workspace [11][13][14]. When we look at how to connect among remote people in a shared space, however, a few challenging systems are proposed. VirtualActor [15] can create a shared virtual environment where virtual reflections of remote others and self appear and act corresponding to actions of real person. Then HyperMirror [16] uses a metaphor of mirror and can create a shared video space by synthesizing reversed self reflection into a remote place. These communication systems are intended to support a sense of interconnectedness between remote partner and self, not necessarily to imitate a real face-to-face situation. In particular, a feature of these systems is to represent reflections of others and self in a common situation.

We have devised a novel communication system in order to create a virtual co-existing space in a physical place as well as in a virtual environment based on this approach representing reflections of others and self at a common place. This system is based on our “Lazy Susan” communication system we have already developed.

Figure 1(a) illustrates a design approach of the “Lazy Susan” communication system, and Figure 1(b) shows the communication system. “Lazy Susan” is a wooden revolving disk, much like what you see in a Chinese restaurant as shown in Figure 1(b). This system supports visual interaction, -in a shared virtual space representing bodily interactions with a physical disk visually, and a virtual disk linked with the physical one-, and tangible interaction - with a physical disk which can be rotated by hand at each site, and their rotations are synchronized with each other-. Results of communication experiments suggested that this system could enhance a sense of co-existence in a shared virtual environment [9].

In order to realize consistently representing reflections of others and self in a common space, and tangible interaction with a networked physical disk, we devise two approaches. The first one is to synthesize the reflection of self with that of others in a common video space, the second one is to project reflections of remote partners onto a local place. Figure 2(a) illustrates the first approach that a synthesized video including reflection of a remote partner and self is projected onto a screen behind a table. We call this expression as Representation(a) hereafter in this paper. Figure 2(b) illustrates the second approach that reflection of a remote partner is projected onto a local tabletop directly. We call the expression as Representation(b) in this paper.

3. “Lazy Susan” video projection communication system

The "Lazy Susan" video projection communication system is composed of a shared disk system and a video projection system as illustrated in Figure 3.

The shared disk system is packed on a wooden table (600×450×770[mm]) on the top of which a rotating wooden disk (280[mm] in diameter) is embedded as shown in Figure 4. The disk can be rotated by hand, and its rotations are synchronized with the movements of the corresponding disk on a remote table. The disk connects with one rotary encoder (NEMICON, OME-360-2MC) and with one DC servomotor (Japan Servo Co., Ltd., DME34S36G10B) as shown in Figure 4, and its motor controller (iXs Research Corp., iMCs01) communicates with the host PC via USB. The Host PC transmits rotation angular data to the remote PC through an IP network, and then each PC applies a feedback loop to control the rotation angle of the disk according to the remote disk’s rotation. When participants interact with the two networked disks, the two disks will behave as if they were coupled by a spring coil. Both disks remain motionless until one of the participants rotates his/her local disk; at that point the corresponding disk will move
When one rotating disk is stopped, the other disk stops at the same time and in the same position. If there is a conflict—if, for example, one participant attempts to rotate the disk clockwise while the other participant attempts to rotate counterclockwise—then each participant will feel torque in the opposite direction.

We explain the video projection system as seen in Figure 3. This video projection system can represent reflection of participants according to Representation(a) and Representation(b) as described in the previous chapter. Representation(a) is installed at one site PlaceA and concurrently Representation(b) is installed at the other site PlaceB. Installing Representation(a) or Representation(b) at both sites can be available. In order to investigate the availability of these representations, each representation is installed at each site. At PlaceA, CCD camera1, which is installed over the table, captures a participant at PlaceA and physical objects on a tabletop. Simultaneously, the video is projected onto a tabletop by video projector1 at PlaceB. On the other hand, CCD camera2 at PlaceB, which is installed at the height of eye view, captures a tabletop and participants around the table. At once the video of the perspective of remote PlaceB is projected by video projector2 onto a screen behind a table at PlaceA. A participant at PlaceA can communicate with a remote participant at PlaceB by viewing reflection of the other and self in a video of a remote place and by rotating the physical disk linked with remote one. On the other hand, the participant at PlaceB can communicate with the remote participant at PlaceA, whose reflection is projected on the local tabletop while rotating the remote linked disk. Figure 5 shows a scene of employing our communication system at both sites.

### 4. Design of Experiments

Our experiments are designed to investigate the extent of connectedness between self and remote things—participant, disk, table, and object—and between a remote participant and local things with or without self-reflection in a synthesized video, projecting a reflection of the remote participant on a real table, and tangible...
interation with remote-linked disk on Representation(a) and Representation(b).

The experiments were conducted under the three conditions on Representation(a) as displayed in Table 1 and Figure 6, and under the three conditions on Representation(b) as displayed in Table 2 and Figure 7. Eleven pairs of adult students (aged 20-24) participated in the experiments. One of the pairs experienced Representation(a) in one place and simultaneously the other experienced Representation(b) in another remote place. Experimental participants were required to rotate the disk alternately while making sure that an object on the disk wasn’t going to fall down. They spent three minutes experiencing each condition. Figure 5 shows a scene of communication experiment under Condition1. After three conditions, each participant answered a questionnaire and wrote down some comments on a sense of co-existence and interaction with the rotating disk under each of the three conditions. The questionnaire includes 12 items for Representation(a) as displayed in Table 3 and for Representation(b) as displayed in Table 5, and each item is rated on a scale from -3 to +3 (0 neutral). After one experiment, the same pair changed the place with each other, and experienced three conditions in the other representation again. Each participant experienced Representation(a) and Representation(b) with the same partner. The order of these conditions was shuffled to each pair. Figure 8 and 11 illustrate the results of average and standard deviation under 12 items of the questionnaire. A Wilcoxon signed-rank test is executed to calculate a significant difference. Table 4 and Table 6 sum up comments obtained from participants under each condition on both representations.

We call condition1 “dual” interaction mode, condition2 “visual” interaction mode and condition3 “tangible” interaction mode.

5. Result

5.1. Results of Representation(a)

In Representation(a), a questionnaire including 12 items in Table 3 is designed to investigate to what extent connectedness between self and a remote place, and also between self and a remote participant, are influenced with or without self reflection in a synthesized video, and with or without the subject’s operating a video disk by physically rotating it. Table 4 sums up comments from all participants.

Questions 1 through 6 ask about the connectedness between self and remote things in a video - a rotating disk that can be connected with physical one, a fixed table, physical objects, and communication partner. Questions 7 and 8 ask about the connectedness between a physical object actually located in front of the subject and its video in a remote place. Questions 9 through 12 ask about a sense of “being co-located” and closeness.

Figure 8 shows most of the participants rated positively all items under “dual” interaction mode. Additionally a highly significant difference can be found between “dual” interaction mode and visual interaction mode, and between “dual” interaction mode and tangible interaction mode. Comments from participants in Table 4 also suggest superiority of “dual” interaction mode over visual interaction mode and tangible interaction mode. Most of the participants reported a sense of connectedness between self and a remote place and between self and remote partner was enhanced. Additionally, we observed an interesting situation in which a participant at Place A almost stretched out to a falling statue on a remote disk when he/she rotated the local physical disk by his/her hand or when he/she touched the local physical disk a remote partner controlled at Place B as shown in Figure 9. Another situation is that when a remote statue was falling down toward reflection of own hand, he/she moved his/her hand away at once as shown in Figure 10. Afterwards,
most of them reported that they felt as if their hands expanded to the remote place. These results indicate a connectedness between self and remote things is enhanced in “dual” interaction mode over in visual interaction mode and in tangible interaction mode.

In visual interaction mode without tangible interaction with video disk, Figure 8 shows participants rated positively Q1 and Q2 inquiring a connectedness between self and a remote disk, and between self and a remote table. However, they didn’t necessarily rate positively Q3,4,5 and Q6 inquiring a connectedness between self and remote objects, and between self and remote partner, since each average of those scores indicates near neutral zero. In view of all comments, these results suggest a connectedness between self and remote things is weak only in visual interaction mode.

In tangible interaction mode without self reflection, Figure 8 shows participants rated positively only Q1 inquiring a connectedness between self and a remote disk. However, each average of the other items’ scores indicates around or below neutral zero. In view of all comments, these results suggest a connectedness between self and remote things is also weak only in tangible interaction mode.

Table 3 Items of questionnaire on Representation(a)

| Q1. | To what extent if they sensed as if they touched a remote disk. |
| Q2. | To what extent if they sensed as if they touched a remote table. |
| Q3. | To what extent if they sensed as if they touched an object on a remote disk. |
| Q4. | To what extent if they sensed as if they touched an object on a remote table. |
| Q5. | To what extent if they sensed as if they touched a participant on a remote disk. |
| Q6. | To what extent if they sensed as if they touched a participant on a remote table. |
| Q7. | To what extent if they sensed as if a local object were on a remote disk. |
| Q8. | To what extent if they sensed as if a local object were on a remote table. |
| Q9. | To what extent if they sensed as if they were in a remote place. |
| Q10. | To what extent if they sensed as if a remote participant were here in a local place. |
| Q11. | To what extent if they sensed as if they were co-located in the same place. |
| Q12. | To what extent if they felt a closeness of a remote participant. |

Table 4 Summary of comments on communication experiments

<table>
<thead>
<tr>
<th>Condition 1</th>
<th>“Dual” interaction mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>They felt as if their bodies expanded toward a remote space.</td>
<td></td>
</tr>
<tr>
<td>They turned aside their hands not to put their hands on a video of remote objects.</td>
<td></td>
</tr>
<tr>
<td>They were surprised that they turned over an object on a remote disk when they rotated the disk.</td>
<td></td>
</tr>
<tr>
<td>They felt as if they touched a remote participant while rotating the disk.</td>
<td></td>
</tr>
<tr>
<td>They felt a sense of disconnectedness between a local table and a video when a remote participant rotated the disk.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Condition 2</th>
<th>Visual interaction mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>They felt a sense of connectedness was weak between self and video of remote table.</td>
<td></td>
</tr>
<tr>
<td>They didn’t feel well as if they touched a remote object.</td>
<td></td>
</tr>
<tr>
<td>They felt their hands passed through an object.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Condition 3</th>
<th>Tangible interaction mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>They felt they rotated the remote disk through a controller device indirectly.</td>
<td></td>
</tr>
<tr>
<td>They felt a sense of connectedness was weak.</td>
<td></td>
</tr>
<tr>
<td>They felt it was obscure a remote participant rotated the local disk or the disk rotated automatically.</td>
<td></td>
</tr>
<tr>
<td>They didn’t feel they rotate the remote disk when they rotated the disk.</td>
<td></td>
</tr>
<tr>
<td>They felt their timing was off.</td>
<td></td>
</tr>
</tbody>
</table>

Figure 8 Result of questionnaire on Representation(a)

Figure 9 A scene of reaching out to a remote object that is going to fall down

Figure 10 A scene of moving own hand away from a falling down statue
5.2. Results of Representation(b)

In Representation(b), a questionnaire including 12 items in Table 5 is designed to investigate to what extent

Table 5 Items of questionnaire on Representation(b)

<table>
<thead>
<tr>
<th>Question</th>
<th>Condition 1</th>
<th>Condition 2</th>
<th>Condition 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>To what extent if they sensed as if a remote participant touched a local disk.</td>
<td>-3</td>
<td>-3</td>
<td>-3</td>
</tr>
<tr>
<td>To what extent if they sensed as if a remote participant touched an object on a local table.</td>
<td>-2</td>
<td>-2</td>
<td>-2</td>
</tr>
<tr>
<td>To what extent if they sensed as if a remote participant touched a local table.</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>To what extent if they sensed as if a remote participant touched an object on a local disk.</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>To what extent if they sensed as if a remote partner touched their hands on a local disk.</td>
<td>+1</td>
<td>+1</td>
<td>+1</td>
</tr>
<tr>
<td>To what extent if they sensed as if a remote object were on a local disk.</td>
<td>+2</td>
<td>+2</td>
<td>+2</td>
</tr>
<tr>
<td>To what extent if they sensed as if a remote object were on a local table.</td>
<td>+3</td>
<td>+3</td>
<td>+3</td>
</tr>
<tr>
<td>To what extent if they sensed as if they were in a remote place.</td>
<td>+3</td>
<td>+3</td>
<td>+3</td>
</tr>
<tr>
<td>To what extent if they sensed as if they were co-located in the same place.</td>
<td>+3</td>
<td>+3</td>
<td>+3</td>
</tr>
<tr>
<td>To what extent if they felt a closeness of a remote partner.</td>
<td>-3</td>
<td>-3</td>
<td>-3</td>
</tr>
</tbody>
</table>

Table 6 Summary of comments on communication experiments

<table>
<thead>
<tr>
<th>Condition 1</th>
<th>“Dual” interaction mode</th>
<th>Condition 2</th>
<th>Visual interaction mode</th>
<th>Condition 3</th>
<th>Tangible interaction mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>They felt as if a remote participant rotated the disk actually.</td>
<td>They felt as if a remote participant touched an object on a local table.</td>
<td>They felt as if a remote participant touched an object on a local table.</td>
<td>They felt as if a remote partner touched their hands on a local disk.</td>
<td>They felt as if a remote participant touched a local table.</td>
<td>They felt as if a remote participant rotated a local disk.</td>
</tr>
<tr>
<td>They felt a connectedness between remote participant and local things is enhanced in “dual” interaction mode.</td>
<td>They felt as if they rotated disk in face-to-face.</td>
<td>They felt as if a remote partner touched their hands on a local disk.</td>
<td>They felt as if a remote hand came in.</td>
<td>They felt as if they rotated disk in face-to-face.</td>
<td>They felt as if a remote partner touched their hands on a local disk.</td>
</tr>
<tr>
<td>These results suggest a connectedness between remote participant and local things is enhanced in “dual” interaction mode.</td>
<td>They felt a sense of connectedness between remote partner and local things is enhanced in “dual” interaction mode.</td>
<td>They felt as if a remote participant touched their hands on a local disk.</td>
<td>They felt as if a remote hand came in.</td>
<td>They felt as if they rotated disk in face-to-face.</td>
<td>They felt as if a remote partner touched their hands on a local disk.</td>
</tr>
<tr>
<td>Questions 1 through 6 ask about the connectedness between remote participant and local place, and also between remote participant and self are influenced with or without projecting reflection of remote participant on a real table, and with or without remote partner’s rotating a physical disk networked with a local disk. Table 6 sums up comments of all participants.</td>
<td>Questions 7 and 8 ask about the connectedness between a physical object at a remote place and its video projected in front of the participant. Questions 9 through 12 ask about the sense of “being co-located” and closenesses.</td>
<td>Questions 9 through 12 ask about the sense of “being co-located” and closenesses.</td>
<td>Questions 9 through 12 ask about the sense of “being co-located” and closenesses.</td>
<td>Questions 9 through 12 ask about the sense of “being co-located” and closenesses.</td>
<td>Questions 9 through 12 ask about the sense of “being co-located” and closenesses.</td>
</tr>
<tr>
<td>Figure 11 shows most of the participants rated positively all items under “dual” interaction mode. Additionally a highly significant difference can be found between “dual” interaction mode and visual interaction mode, and between “dual” interaction mode and tangible interaction mode. Comments from participants in Table 6 also suggest superiority of “dual” interaction mode over visual interaction mode and tangible interaction mode. Additionally, we observed an interesting situation during which a participant at PlaceB was going to touch a video hand, but moved his hand away from the video hand when the video hand was going to hit his own hand. These results suggest a connectedness between remote partner and local things is enhanced in “dual” interaction mode over in visual interaction mode and in tangible interaction mode. However, some participants in visual interaction mode pointed out that they felt a sense of discomfort viewing a two-dimensional video object especially when the object was tall.</td>
<td>Some participants in visual interaction mode pointed out that they felt a sense of discomfort viewing a two-dimensional video object especially when the object was tall.</td>
<td>Some participants in visual interaction mode pointed out that they felt a sense of discomfort viewing a two-dimensional video object especially when the object was tall.</td>
<td>Some participants in visual interaction mode pointed out that they felt a sense of discomfort viewing a two-dimensional video object especially when the object was tall.</td>
<td>Some participants in visual interaction mode pointed out that they felt a sense of discomfort viewing a two-dimensional video object especially when the object was tall.</td>
<td>Some participants in visual interaction mode pointed out that they felt a sense of discomfort viewing a two-dimensional video object especially when the object was tall.</td>
</tr>
<tr>
<td>In visual interaction mode without remote partner’s rotating a physical disk networked with a local disk, Figure 11 shows participants didn’t rate positively all items, since each average of all items’ scores indicates around or below neutral zero. In view of all comments, these results suggest a connectedness between remote partner and local things is enhanced in “dual” interaction mode over in visual interaction mode and in tangible interaction mode. However, some participants in visual interaction mode pointed out that they felt a sense of discomfort viewing a two-dimensional video object especially when the object was tall.</td>
<td>Some participants in visual interaction mode pointed out that they felt a sense of discomfort viewing a two-dimensional video object especially when the object was tall.</td>
<td>Some participants in visual interaction mode pointed out that they felt a sense of discomfort viewing a two-dimensional video object especially when the object was tall.</td>
<td>Some participants in visual interaction mode pointed out that they felt a sense of discomfort viewing a two-dimensional video object especially when the object was tall.</td>
<td>Some participants in visual interaction mode pointed out that they felt a sense of discomfort viewing a two-dimensional video object especially when the object was tall.</td>
<td>Some participants in visual interaction mode pointed out that they felt a sense of discomfort viewing a two-dimensional video object especially when the object was tall.</td>
</tr>
</tbody>
</table>
5.3. Results of Representation(a) and (b)

The results described in the previous section indicate in “dual” embodied interaction mode that Representation(a) can enhance a connectedness between self and a remote place, and simultaneously Representation(b) can enhance a connectedness between a remote participant and a local place. In other words, Representation(a) can create a situation of “being present at a remote place”, and Representation(b) can create a situation of “a remote participant is being present at a local place”. From this view, we consider the result of common items Q9, Q10, Q11, and Q12 in both representation methods. At Q9 inquiring as if they were in a remote place, Representation(a) is evaluated more highly than Representation(b), and a highly significant difference can be found (p<.001) between them. On the other hand, at Q10 inquiring as if a remote participant were here in a local place, Representation(b) is evaluated more highly than Representation(a), and a highly significant difference can be found (p<.01) between them. Additionally both Representation(a) and Representation(b) are highly rated at Q11 as if they were co-located in the same place, and at Q12 on a closeness of a remote participant. These results mean each representation method proposes two approaches where a co-existing space is created.

6. Discussion

We discuss our two representations to create a virtual co-existing space and its fundamental design approach.

First, we consider a method of representing reflection of others and self in a common space. VirtualActor [15] and HyperMirror [16] are constructed based on this approach. A CG avatar appears in VirtualActor [15], and life-sized reversed reflection of an upper body or full body appears in Hyper Mirror [16]. These systems can support a connectedness between others and self by representing their reflections at a common space. In our “Lazy Susan” system, a reflection of one’s own hand and arm appears in a video from the perspective of viewing a table, and in addition the participant can interact with the physical disk. Our experiment results indicate a sense of “as if own body expanded toward a remote place” is enhanced in “dual” interaction mode more than that in only visual interaction and that in tangible interaction. Recent studies in brain science report interesting discoveries on expansion of body image by measuring brain activity when a monkey sees himself reflected in video monitor and when the monkey is using a tool [17]. It is significant to consider a fundamental design method of interface system expanding embodiment and its evaluation from the knowledge on body image in brain science.

Second, we consider a method of representing reflection of a remote communication partner at a local place. Plenty of communication systems have been proposed based on this idea. Hydra [10] and MAJIC [12] can support gaze awareness during conversation. AGORA [14] can support remote collaborative work with physical objects. Clearboard [11] can support gaze awareness during collaborative drawing on a shared board. These systems report availability of video conference and remote collaborative work; however, they hardly evaluate at all a sense of connectedness among participants and common space. Additionally, inTouch [18] proposes to arouse a sense of presence of a remote participant by rotating wooden rollers with each other without representing reflection of participants visually. Our experiment results indicate a sense of “as if a remote participant were present at a local place” is enhanced in “dual” interaction mode more than that in only visual interaction and that in tangible interaction.

It’s interesting that experiment results suggest presence of a remote partner is enhanced even when both participants don’t touch each physical disk concurrently. Additionally, the results also suggest presence of a remote object is enhanced even when it is projected on a fixed table as well as projected on a rotating disk. Experiment results in an approach representing reflection of self in a video space also indicate, although participants can only operate the disk that is one small part of common video space, a sense of remote presence is enhanced when one’s own hand and objects are located on the fixed table as well as when they are on the rotating disk. In our opinion, these results propose a design approach that participants shouldn’t necessarily touch a shared tool all the time, and all of things they can touch shouldn’t necessarily be tele-operated in order to enhance connectedness between a remote place and self, and between remote participants and a local place. This method for connecting a physical tool with a remote space or a virtual environment has also been applied in other domains, for example Tactile Augmentation for conducting virtual therapy for arachnophobia and the reduction of pain during treatment for burns [19].

At the last we propose a fundamental design approach in order to create a virtual co-existing space between remote places. We believe embodied interaction has two roles [4][5]; one is a function to convey intentions with each other by explicitly expressing bodily action such as gaze, facial expression and gesture, another is a function to interconnect self with others,
things and situation implicitly in the background by representing reflections of others and self in a common space or by interacting with a shared tool. Most of previous communication systems are intended to support the former function for remote collaborative work. In contrast, our system, VirtualActor [15] and HyperMirror [16] are intended to enhance an interconnectedness between others and self. Our system has the special ability to enhance an interconnectedness between self and a remote place not only by representing reflection of self but also by the participant’s ability to operate a video disk by rotating a corresponding physical one: and between a remote partner and a local place not only by representing reflection of a remote participant but also by that remote participant rotating a local disk. Tangible interface framework also proposes “duality” expressing information in foreground and background, however, there has been proposed and evaluated few communication systems yet. We propose that “duality” of explicit and implicit embodied interaction should be supported for creating a virtual co-existing space at which people are full-bodied present.

7. Conclusions

A significance of a sense of co-existence and connectedness among remote participants has increased in computer-mediated communication. Therefore a design approach to create a co-existing space at which remote participants are bodily present is necessary. However, previous communication systems have not been developed in the view of creating such a co-existing situation. Therefore, we devise two representation methods; representing a self reflection in a shared video space, and projecting reflection of a remote partner onto the local tabletop. Then based on these approaches we construct a “Lazy Susan” video projection communication system composed of a shared disk system and a video projection system. The results of experiments suggest that our system can enhance an interconnectedness between self and a remote place, and between a remote partner and a local place. We introduce the concept of “duality” of embodied interaction -conveying intentions by expressing explicitly bodily action, and interconnecting self and others implicitly. We believe our system can support this “duality” of embodied interaction and propose an approach based on “duality” of embodied interaction for creating a virtual co-existing space. The direction of our future work will be that we construct and evaluate the design framework by referring to knowledge on extension of body image when using a tool in brain science [17].

Acknowledgements

This research is supported partially by a Grant-in-Aid for the WABOT-HOUSE Project by Gifu Prefecture. The authors would like to thank graduate students Tomofumi Katayama and Masaaki Iyoda for their helps.

We would also like to thank Lydia Timmins for her help in editing the final version of this paper.

References


Jeremy N. Bailenson  
Department of Communication  
Stanford University  
Stanford, CA 94305  
+1 650-723-0701  
bailenson@stanford.edu

Eyal Aharoni  
Department of Psychology  
University of California  
Santa Barbara, CA 93106  
+1 805-893-5798  
aharoni@psych.ucsb.edu

Andrew C. Beall  
Department of Psychology  
University of California  
Santa Barbara, CA 93106  
+1 805-893-5957  
beall@psych.ucsb.edu

Rosanna E. Guadagno  
Department of Psychology  
University of California  
Santa Barbara, CA 93106  
+1 805-893-5082  
guadagno@psych.ucsb.edu

Aleksandar Dimov  
Department of Psychology  
University of California  
Santa Barbara, CA 93106  
+1 805-893-5798  
dimov@psych.ucsb.edu

Jim Blascovich  
Department of Psychology  
University of California  
Santa Barbara, CA 93106  
+1 805-893-6124  
blascovich@psych.ucsb.edu

Abstract

Recent work [1, 2, 3] has argued that subjective questionnaires may be ineffective at measuring copresence towards agents and avatars in immersive virtual environments (IVEs). The current work directly compares self-report and behavioral measures of copresence. In two studies, we measured the interpersonal distance between participants and either an embodied tutoring agent or an unfamiliar embodied agent as they walked through an IVE. We found that participants yielded more personal space to embodied tutors compared to other embodied agents in both studies. However, self-report measures of copresence, likability, status, or interest did not reveal any differences between embodied tutors and strangers. These findings suggest that nonverbal behavior may be a more sensitive measure of the copresence and general influence of embodied agents than self-report measures. While alternative explanations for these findings certainly exist, there are clearly strong advantages of using behavioral measures to study copresence as a compliment to other measures. Given that a large portion of current research evaluating collaborative environments utilizes self-report measures only, the current findings are particularly notable.

Keywords

Embodied virtual agents, virtual environments, social presence, copresence, social interaction, virtual reality

1. Introduction

As collaborative interactions between humans and embodied agents become more common, it becomes critical to gain a thorough understanding of the nature of those interactions. In order to do so, reliable measurement tools that quantify the parameters of these interactions (i.e., copresence or social presence) must be established [4, 5, 6, 7]. Researchers have begun to examine nonverbal behavior as a mechanism to quantify the copresence that embodied agents inspire in humans during collaboration. Such measures include proxemics [8, 9, 10 11], eye gaze [12, 13, 14, 15, 16, 17, 18] and other gestures [19].

Regarding copresence towards embodied agents, we prefer to utilize behavioral measures such as nonverbal gestures, eye-gaze, and task performance [20, 2], as opposed to self-report measures such as questionnaires and anecdotal accounts because behavioral measures have the potential to offer greater sensitivity and reliability over self-report ratings [21]. Moreover, we believe these behavioral measures may be the most appropriate way to detect affective responses [22] that tend to be difficult to explicate verbally or could fall victim to demand
characteristics, where participants act in accordance with their perception of the experimenters’ goals. In our previous research, we demonstrated that interpersonal distance behavior in IVEs was quite similar to interpersonal distance behavior in physical environments [9, 10]. In other words, people tend to leave a personal space bubble of similar size and shape around people, both virtual and physical alike. In the current set of studies, we use personal space as a metric to compare questionnaire-based and behavioral measures of copresence.

Previous research on spacing behavior in the physical world indicates that individuals leave differential amounts of personal space between themselves and others depending on factors such as familiarity and status. For instance, interactions between familiar persons are characterized by a smaller interpersonal distance than interactions between strangers [9]. Other research demonstrates that individuals leave larger personal space bubbles around those who are high in status than around those who are low in status [23]. For example, observations of student-teacher interaction show this pattern with respect to peer-peer interaction [24]. Specifically, these differences in status among interactants tend to correlate positively with interpersonal distance.

Recent studies by the present authors have validated the use of distance cues to measure copresence behaviors in virtual environments [2, 9, 10]. This approach, combined with our knowledge of real-world spacing behavior, lead us to examine participants' experience of presence with embodied agents during collaborations by comparing their interpersonal distance behavior with an embodied tutoring agent [25, 26] to their behavior with other virtual humans.

The strategy of the current work was to have participants interact with agents who vary in both familiarity and status. The literature on nonverbal behavior in face-to-face interaction discussed in the previous paragraph predicts notable differences in personal space behavior based on these characteristics. Consequently, these differences should carry over to the ways in which participants rate these agents via questionnaires as well as how the participants interact nonverbally with the agents. On the other hand, previous research [2] has demonstrated that questionnaires are not always sensitive enough to measure differences in copresence and affective responses towards agents. Consequently, it may be the case that differences between types of agents will only manifest itself with a nonverbal behavior measure.

In Experiment 1, participants collaborated with a desktop tutoring system based on one previously developed by Nass, Moon, and Carney [27]. Then, participants entered an Immersive Virtual Environment (IVE) and walked around an embodiment of that tutor or around an embodiment of a stranger. Experiment 2 is a replication of Experiment 1. The purpose of reporting both replications is to demonstrate the strength of the current finding, as well as its persistence over slightly changed experimental conditions. In both studies we collected interpersonal distance data as well as questionnaire ratings, including participants’ perceptions of copresence, status, interest, and likeability of the embodied agent.

Given the research discussed above, we predicted that participants would respect their tutors’ personal space more than another virtual human’s (e.g., a stranger) personal space. Furthermore, we predicted that the behavioral measure of personal space regulation would detect differences between tutors and strangers that self-report measures would not.

2. Experiment 1

2.1. Method

We manipulated the identity (tutor vs. stranger) of an embodied agent between participants. For half of the participants, the virtual agent was identified as “the virtual tutor.” For the other half, the same virtual agent was identified as “a virtual stranger.” The virtual tutor was understood to be a virtual embodiment of the same procedural algorithm that had been used to train participants during a desktop tutoring session earlier in the study. The virtual stranger was understood to be a virtual embodiment of an unfamiliar algorithm.

Participants consisted of 72 psychology students (36 females, 36 males) from the University of California, Santa Barbara who either received course credit for an introductory psychology course or were paid $10 for their participation. Participants’ age ranged from 15-30 (Median = 20).

The details of the system are fully described in an earlier publication [9], but essentially consisted of a head-mounted display (HMD), dual pipe (stereoscopic) OpenGL PC graphics updated at 60 Hz, and 6DOF head tracking via inertial orientation (Intersense IS300) and video position tracking (WorldViz PPT, millimeter resolution). The explorable space measured 2.6 x 2.5 x 2.5 m (length, width, height). The average latency between head motions and actual update of images in the HMD was 55 ms. Figure 1 illustrates a participant using the IVE equipment in the room where the experiment occurred.

The participant’s view of the IVE at the beginning of the experimental session is depicted below in Figure 2. The virtual agent was based on a deformable mesh model representing a Caucasian male. In this study, it exhibited no behaviors other than blinking. The virtual agent always faced the participant’s starting position. The participant’s eye height was matched to the 1.65 m eye height of the agent.
In the first phase of the study, participants entered a room and sat down with a tutoring algorithm. Half of the participants interacted with a computer program and the other half interacted with a non-linear book fashioned after the “Choose your own adventure” series children books and written to approximately match the interaction with the electronic computer. There were no major differences in the results between these two conditions so this manipulation will not be discussed further in this paper.

The tutorial then presented a series of 20 facts about American culture and later tested participants on similar facts. The tutorial was modeled in great detail after the system employed by Nass, Moon, and Carney [27]. During the tutorial there was no representation or embodiment of the tutoring algorithm aside from text appearing on the display.

After completing the tutorial, participants left the room and were escorted to another room with an IVE system. They were then instructed on how the IVE equipment functioned and asked to join a “virtual person” in a virtual environment. Half of the participants were told that they were to meet a virtual embodiment of a computing program. The other half were told they were to meet a virtual representation of an unfamiliar person. Once immersed, participants were first instructed on how to navigate in the virtual world, and received approximately one minute of practice in walking. They then were asked to examine the virtual person by walking up to him: first to the left side, then to the right side, then to the front and center. From this final location, participants were instructed to read aloud a label positioned on the virtual person’s chest. Each participant performed this sequence of behaviors twice; each time, a different label appeared on the chest of the virtual agent. In the virtual tutor condition, both labels consisted of familiar keywords sampled from the original twenty facts administered in the tutorial (i.e., prom, kitty). In the virtual stranger condition, labels consisted of novel keywords, functionally and syntactically equivalent to those in the virtual tutor condition (i.e., game, bunny) but not words that had appeared during the tutoring session. In both cases, the label was designed to be large enough that participants could read it clearly from their starting position. The label-reading task has been used previously to facilitate proxemic interaction [9, 10]. In all conditions, while the participant was immersed in the IVE, there was always a single experimenter in the room administering the experiment and ensuring that the participants did not walk into any physical walls.

We sampled each participant’s location at 12 Hz. After completing the two walking trials, participants remained in the virtual environment and responded verbally to a questionnaire with thirteen 7-point items on a Likert-type scale ranging from -3 (strongly disagree) to +3 (strongly agree) with a mid-point of 0 (neither agree nor disagree). This questionnaire was designed to assess perceptions of copresence [5, 28], likeability of the embodied agent, the perceived status of the embodied agent, and finally the degree of interest that the agent elicited from the participant. Some of these scales (i.e., copresence and likeability) were previously validated in other work by Bailenson et al. [9], and were designed to explore potential reasons why one might regulate their interpersonal distance behavior in front of an embodied agent. All items appear in Appendix A.

2.2. Results and Conclusions

The tracking equipment automatically and unobtrusively collected position data from the participant
as he or she traversed the IVE. We derived a measure of the position tracking data based on the minimum distance, that is, the single shortest line in space between the center of the participant’s head and the center of the agent’s head during a given trial. We ran an analysis of variance with virtual agent identity (tutor or stranger) as a between-subjects variable and minimum distance as the dependent variable. There were no significant effects. However, as Figure 3 demonstrates, participants appeared to interact with the virtual agent differently over the two walking trials, demonstrating a larger interpersonal distance with the tutor than with the stranger on the first walk but not the second.

![Figure 3. Mean minimum distance to each agent type for walks 1 and 2.](image)

Consequently, we ran a mixed analysis of variance, with trial (first walk vs. second walk) as a within-subjects variable, and virtual agent identity as a between-subjects variable, and minimum distance as the dependent variable. We found a significant interaction between trial and virtual agent identity, $F(1, 69) = 4.89, p < .03$. On the first walk, participants went closer to the stranger. On the second walk, however, there was no significant difference. There were no main effects in this mixed analysis. On their first walk, however, there was no significant difference. On the second walk, there was no notable patterns of gender differences.

We next examined the participants’ responses to the questionnaire items designed to assess perceptions of the embodied agent. Items on three of the four scales (liking, interest, and copresence) were moderate in reliability so we averaged the relevant questions listed in Appendix A, adjusting for the directionality of reverse-coded items. Items comprising the fourth scale containing measures of perceptions of agent status were analyzed separately due to low scale reliability. Nonetheless, the status items produced convergent results. When explicitly asked if the virtual agent was higher in social status, there was no difference in ratings based on an independent samples t-test ($M = -1.28, SD = 1.37$) for the tutor and ($M = -1.19, SD = 1.47$) for the stranger, $p > .8$). Similarly, participant ratings of the formality of their relationship with the agent showed no difference between tutor and stranger, using a reverse-code ($M = -2.27, SD = .198$) for the tutor and ($M = -1.67 SD = .185$) for the stranger, $p > .6$). Also, when asked how comfortable they would be using slang terms when speaking in front of the agent, no difference emerged, adjusting for reverse-coding ($M = -1.31, SD = .182$) for the tutor and ($M = 1.25, SD = .197$) for the stranger, $p > .8$). Furthermore, the ratings of agent status did not correlate with minimum distance, $r = -.072,.058,.184$ respectively, all $p > .1$.

An independent samples t-test showed no difference between the two agent conditions on the likeability scale ($\alpha = .52$); participants disliked the tutor ($M = -.29, SD = .91$) as much as the stranger ($M = -.56, SD = 1.04$), $p > .2$. Furthermore, likeability did not significantly correlate with minimum distance, $r = -.10, p > .4$. The analyses on the additional items revealed a similar lack of significant difference between the two types of embodied agents. An independent samples t-test showed no difference between the two agent conditions on the interest scale ($\alpha = .57$); the tutor ($M = .89, SD = .94$) was rated similarly to the stranger ($M = 1.11, SD = 1.04, p > .2$), and interest did not correlate significantly with minimum distance, $r = -.08, p > .3$. Finally, an independent samples t-test showed no difference between the two conditions on the copresence scale ($\alpha = .73$); the tutor ($M = -1.00, SD = .99$) was rated similarly to the stranger ($M = -1.94, SD = 1.06, p > .8$). Copresence did not significantly correlate with minimum distance, $r = .07, p > .5$. In all analyses, there were no notable patterns of gender differences.

Because we obtained a significant difference in the distance participants maintained from the embodied agent only with the first trial in Experiment 1, we revised our experimental measures and collected data from 48 additional participants in Experiment 2, extracting more data from each participant so that it would be possible to examine their personal space behavior in greater detail.

3. Experiment 2

3.1. Method
The design was identical to the design of Experiment 1. Participants interacted with a learning algorithm, and then interacted with either a virtual embodiment of that algorithm or a virtual embodiment of a stranger. There were 32 participants in the tutor condition and 16 participants in the stranger condition. These numbers are disproportionate due to counterbalancing with two different types of tutors. There was no difference between the two tutors, and this effect is not discussed further in this paper. All significant differences in personal space reported hold up if we include only the first 16 subjects run in the tutor condition. However, we include the extra 16 subjects in the analyses in order to increase the probability of finding differences with the self-report data.

Participants consisted of 48 psychology students (27 males, 21 females) from the University of California, Santa Barbara who received course credit in an introductory psychology course for their participation. Participants’ age ranged from 18-23 (Median = 19).

The materials and apparatuses were identical to those of Experiment 1.

The procedures were identical to the procedures of Experiment 1 except that in order to more thoroughly examine trial order effects, position data was collected over the span of six walking trials. Consequently, there were also six different labels for each participant to read from the virtual agent’s chest. As before, labels were either taken from the tutor session (tutor condition) or were random (stranger condition). We administered only likeability and copresence of the original sets of items on the self-reported perceptions of the embodied agent in this study due to time constraints in a given session resulting from the four additional walks around the agent.

3.2. Results and Conclusion

We ran an analysis of variance with virtual agent identity (tutor vs. stranger) as a between-subjects variable, and the average minimum distance of from the six trials as the dependent variable. There was a significant effect of virtual agent identity, F(1, 42)=8.78, p<.005. As Figure 4 indicates, participants maintained a larger interpersonal distance with the tutor than with the stranger, both on the first walk and on the other five walks as well. Furthermore, there was a linear trend for participants to leave more interpersonal distance on later walks than on earlier walks, F(1,44)=6.32, p<.05.

We next examined participants’ ratings of likeability and copresence. First, we examined the reliability of each scale. The two likeability items were low in reliability and, thus, were analyzed separately. Independent samples t-tests for each item showed no difference in likeability between the two conditions; participants disliked the tutor (M = .72, SD = .81) as much as the stranger (M = .56, SD = .81), p>.5, and they reported equal attractiveness ratings for the tutor (M = -1.16, SD = 1.42) and the stranger (M = -1.06, SD = 1.65), p>.8. Furthermore, the likeability items did not significantly correlate with minimum distance, r = .041, p>.7, and r = -.078, p>.6 respectively. An independent samples t-test showed no difference on the copresence scale (α = .59) between the two conditions; the tutor (M = -1.43, SD = 1.20) was rated similarly to the stranger (M = -1.66, SD = 1.09, p>.6). Furthermore, copresence did not significantly correlate with minimum distance, r = .12, p>.4. In all analyses conducted, there were no notable patterns of gender differences.

Consistent with Experiment 1, participants maintained a significantly larger bubble of personal space around the virtual tutor than around the virtual stranger, and this effect was validated across all six walking trials. However, again, self-report measures on perceptions of the embodied agent did not demonstrate any differences between these two conditions.

4. Conclusions

In the current set of studies, two findings emerge consistently. First, participants who interacted with an embodied tutoring agent in an IVE demonstrate larger interpersonal distances between themselves and the agent than participants who interacted with an unfamiliar embodied agent. Second, this behavioral measure detected differences between tutors and strangers that questionnaire-
Participants left larger personal space bubbles around embodied tutoring agents than around embodied strangers. There are many explanations for such an effect given what researchers have demonstrated in vivo with regard to interpersonal distance. First, participants may have been more interested in the stranger than the tutor because they had already met the familiar tutor via text interface on a desktop computer. However, this interpretation is in conflict with earlier findings that greater familiarity is associated with closer interpersonal distance [9]. Second, we consider that a tutor is someone who provides knowledge and in doing so is higher in status over those that he or she teaches [27]. Thus, in staying farther away from the tutor than the stranger, participants may have been demonstrating politeness towards their tutors or deferring to their status. Finally, participants may have recognized the word from the tutorial session more easily in the tutor condition, and consequently may have not walked as close to examine the words themselves. However, that explanation would not explain the participants approach differences on the profile sides of the embodied agents. Future research could examine this further by designing interactions with an embodied agent in which personal distance could be more accurately interpreted as a sign of deference, dislike, disinterest, or prior recognition.

In addition, across both studies, there was a trend for participants to maintain a larger personal space bubble between themselves and the embodied agent as the number of trials increased. One potential explanation for this trend is that participants became disconcerted while viewing a virtual agent that did not perform any behavior aside from blinking, and after a number of trials they shied away from the representations. Future research could examine this issue further by varying the level of behavioral realism exhibited by the virtual agent.

However, the goals of the current studies were not to provide an underlying theory of interpersonal distance behaviors within IVEs. Instead, the goals were to demonstrate that behavioral measures were successful in detecting differences in the way participants interacted with different types of agents, while questionnaires alone were not. Consequently, to attempt to measure the nuances of human interaction with agents and avatars using self-report data alone may be short-sighted.

Of course, this claim needs to be thoroughly qualified given the current data set. It could be the case that we did not have enough statistical power to properly demonstrate the effects of the self-report data. Furthermore, we may have failed in choosing the correct questionnaire measures—while we attempted to use a wide variety of scales as well as ones with previous validity, it was not possible to examine every single potential set of questions. Indeed, the reliability index on most of our scales was moderate at best. Moreover, the context of the current study—an agent who did talk, express emotion, or walk about—certainly limits our findings to such low-level types of interactions. Finally, it is not clear what the difference in personal space between the two conditions actually represents—status, familiarity, copresence or some other type of affective inclination participants maintained towards the tutor. As a result, it could be the case that the nonverbal behavior has little to do with the copresence, status, likability, or interest self-report measures. In sum, we are not claiming that personal space is a direct proxy for copresence. However, clearly people behave nonverbally towards tutors and strangers differently in IVEs, and this behavior does not map onto any obvious questionnaire rating.

Nonetheless, one thing is clear. In both studies, we observed notable differences in nonverbal behavior towards the two types of agents, but could not match that difference with any type of questionnaire. This is by no means evidence that questionnaires are not useful. However, these data, along with data from previous studies [2], indicate that it is crucial to augment self-report data with some kind of behavioral measure. In future work we plan to examine a larger scope of questionnaires as well as to compare behavioral responses to open-ended self report measures. In addition, we plan to scrutinize more involved nonverbal behaviors, such as eye-gaze and facial expressions.

In addition to providing information to researchers studying presence and copresence measurement, these two studies provide evidence that participants treated embodied agents that they had prior exposure to qualitatively differently than they treated embodied strangers. Given the increasingly common utilization of embodied agents, these results have implications for designers of collaborative systems, in that certain agents may be more appropriate than other types, depending on the type of collaboration. These results suggest that people may initially investigate a virtual stranger more closely than a virtual agent they have prior experience with, even if that experience is not in a virtual environment.

In sum, in these two studies, self-reported perceptions of the embodied agent did not detect differences in our two conditions. This is not to say that one should not utilize self-report questionnaire-based measures in research in this area; countless studies concerning collaborative environments as well as IVEs have demonstrated notable and valuable findings using questionnaire-based and open-ended response forms of self-report data. The purpose of this paper was to provide new empirical data that encourage augmenting questionnaires with behavioral measures when possible. Questionnaires have the advantage of being relatively easy to administer, to
validate, to determine reliability through psychometric techniques, and to share with other research groups [30]. However, one of the greatest limitations in questionnaire-based studies is that participants are not always the most accurate judges of their own thoughts and feelings, so they often misreport affective and cognitive responses to stimuli. Therefore, dependent measures based on self-report questionnaires are best used in conjunction with other measures. The convenience of questionnaire-based dependent measures should not come at the expense of measurement power, which behavioral variables, at least in the current study, have as an advantage over self-report.

References


APPENDIX A

Social Presence
I perceive that I am in the presence of another person in the virtual room with me.
I feel that the [person OR tutor] in the virtual room is watching me and is aware of my presence.
The thought that the [person OR tutor] is not a real person crosses my mind often.
The [person OR tutor] appears to be sentient, conscious, and alive to me.
I perceive the [person OR tutor] as being only a computerized image, not as a real person.

Likeability
I like the virtual [person OR tutor].
I think the virtual [person OR tutor] is attractive.

Status
The virtual [person OR tutor] is of higher social status than I am.
My relationship with the virtual [person OR tutor] is a casual and informal one.
I would feel comfortable using slang words in front of the virtual [person OR tutor].

Interest
I am interested in the virtual [person OR tutor].
I feel that the virtual [person OR tutor] is interesting to look at.
Theoretical and Empirical Support for Distinctions Between Components and Conditions of Spatial Presence

Saskia Böcking, Andre Gysbers, Werner Wirth, Christoph Klimmt, Tilo Hartmann, Holger Schramm, Jari Laarni, Ana Sacau, & Peter Vorderer
University of Zurich, Hanover University of Music and Drama (HMT), University of Zurich, Hanover University of Music and Drama (HMT), Hanover University of Music and Drama (HMT), University of Zurich, Helsinki School of Economics and Business Administration (HKKK), University Fernando Pessoa (UFP) Porto, University of Southern California (USC) Los Angeles
{s.boecking@ipmz.unizh.ch, andre.gysbers@ijk.hmt-hannover.de, w.wirth@ipmz.unizh.ch, christoph.klimmt@ijk.hmt-hannover.de, tilo.hartmann@ijk.hmt-hannover.de, h.schramm@ipmz.unizh.ch, jari.laarni@hkkk.fi, pssacau@ufp.pt, vorderer@usc.edu}

Abstract
One of the most important measures of Presence is the questionnaire. Different instruments have been introduced; however, they are based on different and partly implicit theoretical assumptions. The MEC model of Spatial Presence has been proposed as a theoretical framework for the unification and simplification of the existing Presence research. Based on a short explication of this model, the definitions and operationalization of selected constructs (i.e. involvement and presence) within the MEC Spatial Presence Questionnaire (MEC SPQ) will be explained and distinguished from their former use in presence research. Finally, we will present some findings from pretest studies that were conducted with 290 students from three different countries (U.S., Portugal, Finland) to demonstrate the differences between presence and involvement. Testing four different media (linear text, hypertext, film, virtual environment) the data not only supported the constructs’ validity but also allowed to create highly consistent and homogeneous scale versions for these constructs.

Keywords--- Spatial Presence, Components, Theory, Measurement, Questionnaire

1. Introduction

One important contribution of social-scientific approaches to Presence research is the development of robust and valid measures of the construct. To be able to generate reliable and valid data on Spatial Presence is a crucial condition for successful basic research, application development, service evaluation and product optimization. Consequently, the demand for usable and effective measures of Presence is strong both in academic and industrial contexts.

So far, a lot of different approaches have been used for measuring presence. Usually, subjective and objective measures are differentiated [1]. Even if it is questionable whether objective measures assess presence well – presence is mostly assumed to be a subjective sensation that it is not so amenable to objective physiological measurement [2] – the main problem of existing measurement tools is the lack of a solid theoretical foundation. This point is critical as by which way presence is measured depends on the theory used [3]. MEC’s model of Spatial Presence [4] has been introduced to close the theoretical gap in presence research. However, due to the theoretical framework developed within this model, some constructs (i.e. presence, involvement) have been defined somewhat different to their former use in presence research. Because of the close connection between presence theory and the way of measuring presence, it was necessary to develop new measurement tools that fit the theoretical framework of the MEC model.

In the following sections we will report the development of this new measurement tool1. We start with a short summary of MEC’s model of Spatial Presence which was used as a theoretical framework (section 2). We then explain in more detail MEC’s definitions of presence and involvement (section 3). In doing so, we will explicate our understanding of these constructs in comparison to existing understandings in presence research. At the same time, the conceptual clarification is the basis for the operationalization and thus the development of the scales for measuring these two constructs. Finally, we present empirical findings on the reliability and validity of the scales developed for measuring presence, and involvement (section 4).

2. The MEC model of Spatial Presence

1 As we consider presence to be a subjective experience, we developed a questionnaire that encompasses the different constructs of the MEC model, i.e. the MEC Spatial Presence Questionnaire (MEC SPQ).
The MEC model of Spatial Presence [4] has been proposed to serve as a unifying theoretical approach in explicating the development of feeling physical – or as it is named in the model – spatial present in a mediated environment. Integrating both media and user factors, it is applicable not only to virtual environments, but to less interactive and immersive media settings as well. Moreover, by introducing processes of attention and perception it integrates mental mechanisms which enable humans to feel spatially present.

According to the MEC model, Spatial Presence is considered to be a specific part of the experience of presence. It arises when media users think that they are put in an environment offered by the medium. The model suggests that spatial presence emerges in two steps (cf. figure 1). In the first step, the user constructs a spatial situational model (SSM) of the mediated environment (cf. [5]). Decisive for the construction of the SSM are automatic and controlled attention processes. They result in the user’s fading out of environmental stimuli and his focusing of the mediated environment. The user perceives the spatial cues of the mediated environment and constructs a mental model out of them. It is assumed that individual user characteristics such as spatial ability [6] influence the form of the SSM. The second step of the model comprises perceptual processes that are based on the SSM and guided by hypotheses [7]. In the course of hypothesis testing, the user scrutinizes the assumption that the mediated spatial environment is the primary egocentric-reference frame [8] [9]. If this assumption is affirmed, the user feels Spatial Presence. Within this process of hypothesis testing different user characteristics are considered to be important. The model here concentrates on processes of cognitive involvement [9], and suspension of disbelief [10], which is affected by trait variables, such as absorption [11].

Figure 1 MEC’s two-level-model of Spatial Presence

3. Conceptual clarifications of MEC’s understanding of Spatial Presence and involvement

In contrast to existing theoretical and empirical modellings of presence experiences, the MEC model explicitly distinguishes presence, involvement, and attention by definition. This distinction and the integration of mental mechanisms allow for empirically testable predictions about the formation of Spatial Presence experiences. However, MEC’s understanding of presence, involvement, and attention is in some way different to existing definitions and usage. Therefore, we intend to clarify these differences.

3.1. The presence concept in the existing presence literature

In general, the origin of the presence concept is seen in the term telepresence that was first introduced by Minsky [12]. In its original formulation it meant the illusion of being transported via telecommunication systems to a real, physical location experienced synchronously. Since the introduction by Minski, however, this term has been generalized to a sense of transportation to any “space” created by media [13] and finally has been broadened to the shorter and more common term “presence”. Recently, presence has been generalized to the illusion of “being there” or, to name it more precisely, the “perceptual illusion of nonmediation” [14]. Lee [15] has presented the latest review of existing definitions of presence and came up with a similar definition which concentrates on aspects of perception and thus on subjective experiences.

Although most presence researchers agree in regarding presence as a subjective experience and although the broad definition of Lombard and Ditton [14] is widely accepted, there is still some confusion regarding the components of presence. Several subtypes of presence are defined which are not labelled uniformly.

When analyzing the use of the general term presence, Lombard and Ditton [14], for example, identified two broad theoretical categories: physical and social presence. While physical presence refers to the sense of being physically located in a mediated environment, social presence refers to the feeling of being together (and communicating) with a mediated person. In addition, Lombard and Ditton [14] report presence being linked to media or stimuli factors as social richness, realism, immersion, or link this concept to user’s reactions to media (such as parasocial interaction). Draper, Kaber, and Usher [16] distinguish three different types of presence: simple, cybernetic, and experiential. While the first is qua definition simply the ability to operate in virtual environments, the second is concerned with aspects of the human-computer interface. The third, finally, is the one most scholars think of when talking of presence by using the general term: a mental state in which a user feels physically present within a computer-mediated environment. Heeter [17] again, splits the general term presence in three subdimensions: personal, social and environmental.
presence. While the first is defined as the sense of being in a virtual world – including the reasons why one feels like being there (e.g., a dynamic, artificial representation of some part of oneself) –, social presence is seen as a subset of personal presence with the user being together with others in the mediated world, too, and interacting with them. Environmental presence refers to the extent to which the environment itself appears to know that the user is there and to react to him/her. Biocca [18], finally, distinguishes theoretically between physical, social and self presence. While his understanding of social and physical presence is almost identical with its use by Lombard and Ditton [14], he defines self presence as the users’ mental model of himself inside the virtual world.

Consequently, different authors define the general term presence differently. Nonetheless, most of them agree on the basis of theoretical considerations that presence is a subjective experience of which the feeling of being physically located in a mediated environment is one main component.

However, what presence, i.e. the feeling of “being there”, exactly is, has not been answered uniformly. Different authors have defined this experience in terms of hardware or stimulus effects [10] [13] [19] [20]. Moreover, empirical findings propose presence to be a complex, multidimensional construct. Using cluster analyses, Witmer and Singer [20], for example, identify control (i.e., perceived control of events in the VE), responsiveness of the VE to user-initiated actions, the involving power of the visual aspects and the participant’s involvement – which the authors define similarly to attention – in the VE), naturalness (i.e., the extent to which interactions with the VE feel natural, the extent to which the VE is consistent with reality, and the degree of naturalness of the control of locomotion through the VE) and interface quality (i.e., interference of or distraction by control or display devices from task performance, extent to which the participant feels able to concentrate on the task) as components of presence. Lessiter, Freeman, Keogh, Davidoff [21], applying the ITC-SOPI, extract the factors sense of physical space (i.e., a sense of physical placement in the mediated environment, interaction with, and control over parts of this environment), engagement (i.e., the tendency to feel psychologically involved and to enjoy the content), naturalness (i.e., a tendency to perceive the mediated environment as life-like and real), and negative effects (i.e., adverse physiological reactions caused by the mediated environment). Similarly, Schubert, Friedmann, and Regenbrecht [22] identify a three-component structure of presence, containing the factors spatial presence (i.e., sense of “being there” and possible actions in the VE), involvement (i.e., awareness and attention processes) and realness (i.e., in how far the virtual and the real world are perceived to be similar). The authors prove them to be on the one hand different from, but on the other hand closely related to each other. Besides these three components that they regard as facets of presence, five other factors (i.e., quality of immersion, drama, interface awareness, exploration of VE, predictability and interaction) have been identified. In contrast to the presence factors, they tap descriptions of the stimuli offered by the VE and the interface, and the interaction with them. However, when taking a closer look at all these components, it has to be taken into account that factor and cluster analyses can only structure the items derived theoretically. This is problematic as both Witmer and Singer [20] as well as Lessiter et al. [21] did not base their items on a theoretical framework. Moreover, there are other problems which we address in the following section.

3.2. Problems of the existing understanding of the presence concept

At first glance, the three investigations analyzing components of presence revealed different components. However, if one examines these components more deeply, most of them turn out not to be part of the presence concept, i.e., the feeling of being in a mediated environment. Rather, they are hardware, software, subject or task variables which, when combined, will add to the sensation of presence. For example, the items of the Witmer and Singer questionnaire [20] do not directly assess subjective presence experiences, but address factors that influence involvement (which the authors defined as awareness and attention processes) and immersion. By doing so they rather measure subjective evaluations of immersive technology than presence [22]. Thus, Witmer and Singer [20] do not properly distinguish between the mere experience of presence and factors leading to this experience. The same holds true for Lessiter et al. [20] who did not clearly define what they mean by the term “presence”, but used 15 different content areas deemed relevant to presence on the basis of theoretical and empirical papers instead. Thus, their components of presence – as that of many other authors – are mainly related constructs which seem to be rather perceptions of different hardware components enhancing presence or effects of presence, but not the experience of being present itself.

Altogether, the different factors mentioned above are usually regarded as being part of the presence experience. However, some scholars prove that, for example, realness is closely related to presence but at the same time is different to it [22]. According to these results, presence experiences are influenced by, but are distinct from evaluations of the immersive technology or evaluations of interaction. Therefore, it is more plausible to assume that constructs such as involvement or reality judgment are antecedents of presence experiences rather than components. In this context, it is pointed to the possibility of feeling present in a media environment without assigning reality status to it [23]. Schubert [24] highlights that the sense of presence has to be distinguished from the technological quality of the virtual environment. “The former is subjective experience similar to a feeling. The latter is commonly called immersion ..., referring to objective descriptions of the technology” (p. 69). Similarly, Kalawsky [25] states: “It is clear that presence is a cognitive factor that must be treated differently than other perceptual aspects of a human-computer
interface such as brightness or contrast of an image.” (p. 5)
Additionally, sometimes negative effects have been considered to be one component of presence [20]. However, as the word “effects” suggests, behaviour such as becoming motion-sick may be an outcome of presence rather than an element of it.

These problems of existing conceptualizations of different presence components are due to the lack of a theoretical framework which underlies the definitions and empirical investigations. The MEC model of Spatial Presence has been proposed to close this gap (cf. section 2). However, this conceptualization defines presence and involvement differently from their former use in presence research, i.e. as conditions of presence.

3.3. Presence, and involvement within the MEC model of Spatial Presence

The feeling of being physically located in a mediated environment is considered to be a subtype of the general term presence (section 3.1.). Most scholars, however, when talking about this experience, do not define in which way the feeling of being present manifest itself. Rather, most of them mention this experience by labeling it as the feeling of “being there” or they misleadingly include different factors or perceptions leading to presence into their definitions (section 3.2.).

Schubert et al. [22] refer to cognitive processes and propose more thorough considerations when trying to explain the formation of presence experiences. According to them, two cognitive processes are involved in the emergence of presence: the construction of a mental model and attention allocation. In applying the embodied cognition framework [26], they argue that media users feel present when the mentally represented actions are bodily actions within the space depicted. Attention allocation is crucial for this process, as without attention is not possible to construct a mental model of the mediated environment. Using factor analyses Schubert et al. demonstrate that presence experiences in fact consist of spatial-constructive and attention facets.

Following these considerations, the MEC model distinguishes between attention processes, the construction of a mental model of the mediated environment, and the experience of presence itself. In contrast to other scholars, presence – which we call Spatial Presence in order to distinguish it from the general term and to emphasize the aspect of feeling physically present in the mediated environment – in a first step is defined as the subjective experience of being in the mediated environment. However, this mere verbal description does not offer any possibility for formulating concrete items – except items which directly address the feeling of being present in the mediated environment. Taking this problem into account and regarding that the embodied cognition framework suggests that mentally represented actions of the own body in the space depicted are important for experiencing Spatial Presence, we would like to widen the classical definition of Spatial Presence for this aspect. Thus, Spatial Presence consists of two dimensions: (1) the classic description of presence, i.e. the sensation of being physically situated within the spatial environment portrayed by the medium (self-location), and (2) perceived possibilities to act in the mediated environment (possible actions). Following the cognitive processes proposed by the model, a user experiences Spatial Presence, if he perceives himself to be in and connects his action possibilities to the mediated environment. The user’s mental capacities are bound by the mediated environment instead of reality. He perceives only those action possibilities that are relevant to the mediated space, but will not be aware of actions that are linked to his real environment.

In doing so, we limit Spatial Presence to its core, i.e. the concrete experience. Cognitive processes as attention allocation or the construction of a mental model of the mediated environment are regarded as preconditions of this experience. Different media and user factors that can affect those preconditions of Spatial Presence experiences are integrated in the MEC model. However, they are not considered to be part of this feeling.

One of the user factors taken into account in the MEC model is involvement. So far, in presence research involvement has mainly be conceptualized as awareness and attention processes [22], immersion into the virtual environment [28] [27] or a mixture between attention and immersion [20]. However, according to our model, attention processes are assumed to take effect especially in the first stage. Thus, involvement as it is understood here implies more than just mere attention. Following conceptualizations from advertising [28] [29] and communication research [30] we therefore regard involvement as higher forms of information processing that may have cognitive, affective and conative aspects. However, as affective and conative aspects are strongly connected with cognitive processes, we consider involvement to be the intense cognitive engagement with a media environment that can be observed via processes of appraisal, elaboration, evaluations, and mental explorations. Thus, involvement is sharply distinguished from Spatial Presence. While involvement means the active and intense processing of the world presented by the media, Spatial Presence emphasizes the experience of being solely within the mediated world. In being conceptualized as active and intense processing of the mediated environment, involvement fosters Spatial Presence as processing the content of the media stimuli strengthens the assumption that the mediated spatial environment is the primary ego-reference frame.

Based on the proposed definitions of Spatial Presence, and involvement we now report the development of the scales measuring these constructs.

4. Development of scales measuring Spatial Presence, and involvement

A new Presence questionnaire has been developed [31] that targets both modelled dimensions of Spatial Presence
4.1. Basic principles in item formulation

As items should be theoretically derived and valid, i.e. they should be based on a solid theoretical definition and appear consistent with the theoretical domain of the construct [33], we started the development of the Spatial Presence and involvement scales from the proposed definitions. In doing so, regarding Spatial Presence we took into account the already existing wording of the items from different presence questionnaires. These items were reviewed and updated according to our definition of Spatial Presence, i.e. referring to the dimensions self-location and possible actions.

According to recommendations of different authors dealing with the development of scales in general, we took care of a couple of methodological guidelines in formulating single items:

- each item should express only one idea [34]
- lengthy items should be avoided [32]
- items that are not clear, not concise, ambiguous, not concrete should be avoided [34]
- use of negatives to reverse the wording of an item should be avoided [34]

Thus, we tried to make sure that the items were as easy to understand as possible in favour of the scales’ reliability. Moreover, as there is no common notion for all possible media the scales might be applied to, we decided to insert a placeholder for the type of media stimulus in each single item that had to be replaced by the medium tested.

4.2. Method

In order to ensure the applicability of the Spatial Presence and involvement scales to different media settings, the item pool derived from the definitions of the constructs was tested with four types of media. For each kind of medium, a typical content was selected:

- Linear text: extract out of „The pillars of the earth“ by Ken Follett
- Film: sequence from „Das Boot – Director’s Cut“
- Hypertext: „The Art of Singing“ (2-D virtual academy of song)
- Virtual environment: Musée d’Orsy in Paris (exhibition of art of the 19th century)

As the scales were intended to be applicable to international samples, the questionnaire was developed in English and tested in both Europe and the U.S. Participants were recruited at international schools and universities in Porto (Portugal), Helsinki (Finland), and Los Angeles (USA). The investigation was indicated as a study relating to the field of media psychology dealing with how people experience the different media used, thus guaranteeing the participant’s ignorance of the actual purpose. As the scales should be applicable for adults only, participants had to be at least 15 years old.

Experiments with the text and film stimuli in L.A. were conducted in group sessions with 6-7 participants per session. Hypertext (Helsinki) and VR (Porto) were tested in single sessions due to technical restrictions. Subjects spent 10 minutes time receiving their media stimulus.

At each location a dual-task paradigm was implemented to validate the sensitivity of the scales. Half of the participants were distracted several times during media usage and had to perform a secondary task, the other participants were not distracted. Participants were randomly assigned to either the distracted or the non-distracted condition.

The basic idea of this experimental manipulation was to produce different intensities of Spatial Presence experiences systematically in order to test the scales’ sensitivity for this variation. If the scales respond to the variation, this would be an empirical demonstration of their validity. According to MEC’s model of Spatial Presence, attention allocation is the origin of the process that leads to the experience of Presence. The employed secondary task (random number generation) was intended to reduce subjects’ attention for the medium they were exposed to and thus “slow down” or “interrupt” the formation of Spatial Presence repeatedly by affecting the starting point of this process. It was therefore considered as a strong, theory-based tool to manipulate the subjects’ experience of Spatial Presence. As cognitive involvement is presumed to be affected by situational factors, differences in involvement were also assumed.

The questionnaire started with welcoming the participants, instructions how to fill out the questionnaire, and assuring anonymity. Then the initial item pool was assigned. Items were presented in randomly mixed order. Finally, participants had to answer some sociodemographical questions (age, gender, education) and were asked for their English skills, if they were non-native speakers.

5. Results

Altogether, 291 participants took part in the investigation. One subject had to be excluded from further analysis due to too many missing data. 80 participants read the linear text, 81 in each case watched the film or attended the hypertext, and 49 tested the virtual environment. The participants’ mean age was 21.4 years (SD=5.2), ranging from 15 to 54 years. The samples for hypertext and virtual envi-
5.1. Basic principles for item analysis and item selection

Examination of data showed that item distributions were very similar across the four locations. Thus, to form a larger sample, reliability analyses were performed on the scales using the combined data from all four experiments. We used an iterative approach to eliminate items for the final version of the three scales. The analysis was intended to result in a final 8-item version of each scale. Although shorter versions (with 6 or 4 items per scale) have been developed as well, throughout this paper we will concentrate on the 8-item version only, for this version is recommended for future research.

In a first step, principal component analyses (PCA) were conducted to test the scales’ reliability and validity. The involvement scale was analysed separately, the two Spatial Presence scales were analysed conjointly using PCA with varimax rotation. This is due to theoretical considerations as Spatial Presence is considered to exist of the dimensions self-location and possible actions. In case of homogeneity, a one-factor-solution for involvement and a two-factor-structure for Spatial Presence should emerge. Factor loadings were then used as first indicators for fits of single items. Items with small factor loadings or double loadings (regarding the scales self-location and possible actions) were considered for exclusion.

Furthermore, difficulties of all items were computed. For dichotomous (e.g., right/wrong) items, the difficulty is the proportion of right answers, ranging from 0 to 1. In our case (Likert-scales from 1-5), it reflects the item’s arithmetic mean, i.e. a difficulty of 0 corresponds with a mean of 1 and a difficulty of 1 is equivalent to a mean of 5. Item difficulties should range from .20 to .80 and ideally come close to .50 [37]. We used variability of item difficulties within this range as one criterion for excluding and retaining items.

Because internal consistency and homogeneity of the final scales were a major concern of this study, we ranked each scale’s items considering the following (corresponding) measures:

- Cronbach’s alpha if item deleted indicates to what extent dropping one item would elevate or decrease Cronbach’s alpha for the remaining scale. Cronbach’s alpha of the resulting scale should be at least .70 [34].
- Item-remainder coefficients (corrected item-total correlation; item-scale correlation) are computed as correlation of each item and the sum of remaining items. This value should be as high as possible, but not fall below .30, as a rule of thumb [37]. Correlations with other scales should be lower than item-total correlations. These correlations were computed tentatively by using the original scales with 12 items.
- Item homogeneity (average inter-item correlation) of an item should also be high, indicating that these items assess similar information. However, moderate homogeneity coefficients comply with higher validity [37].

Preliminary analysis showed that all scales were highly internally consistent even in the original version with 12 items. However, if the statistical criteria mentioned above were applied, merely very similar worded items were selected. Although redundancy can be useful in terms of repeated measures [32], it is possible to derive a scale with high internal consistency by writing the same items in different ways [33]. Therefore, during each step of item selection, two items that normally could have been dropped were retained, if they accounted for more variability in the scale’s wording. This procedure only marginally decreased internal consistencies.

5.2. Reliability and validity of Spatial Presence and involvement scales

5.2.1. Reliability

The reliability of the three scales was tested by computing PCA (see section 5.1.) and computing Cronbach’s Alpha for each scale. Having conducted a PCA with varimax rotation by entering all items of the two Spatial Presence scales, the screeplot suggested two components explaining 35.8 % (self-location) and 23.6 % of the variance (possible actions). The factor solution gives support for two separate subscales, as most of Self Location items clearly loaded on the first component, and most Possible Action items constituted component two3. Alternative analyses were performed (principal axis analyses, oblique rotations), but did not reveal any significant difference in the ranking of the items.

PCA for involvement items initially suggested a three factor solution, explaining 52.8 % of variance. However, this solution did not allow for a reasonable interpretation. Screeplot of eigenvalues (3.89, 1.36, 1.10, .93, .85, .70, .66, .59, .53, .50) suggested two components providing 63.8 % and Porto (49.0 %) on the one hand, and L.A. on the other hand (text: 88.5 %, film: 84.0 %).

Almost three-fourths of participants were female, with female proportion significantly differing between Helsinki (63.8 %) and Porto (49.0 %) on the one hand, and L.A. on the other hand (text: 88.5 %, film: 84.0 %).

<table>
<thead>
<tr>
<th>Applied medium</th>
<th>Statistical values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear text</td>
<td>n 80  M 19.7  SD 3.9</td>
</tr>
<tr>
<td>Film</td>
<td>81  19.5  1.4</td>
</tr>
<tr>
<td>Hypertext</td>
<td>81  24.4  3.7</td>
</tr>
<tr>
<td>VE</td>
<td>49  22.9  9.5</td>
</tr>
<tr>
<td>all 4 media</td>
<td>290  21.4  5.2</td>
</tr>
</tbody>
</table>

Table 1 Age of participants by medium tested

PRESENCE 2004
.61, .59, .56, .43, .32) suggested the appropriateness of a single factor solution, explaining 32.4% of variance, which pointed at the unidimensionality of the scale.

The initial scales yielded high internal consistencies which only marginally decreased after dropping items for the 8-item scale (see table 2).

<table>
<thead>
<tr>
<th>Scale</th>
<th>Statistical values</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cronbach’s α</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Spatial Presence Self Location</td>
<td>.94</td>
<td>2.38</td>
<td>.91</td>
</tr>
<tr>
<td>Spatial Presence Possible Actions</td>
<td>.88</td>
<td>2.32</td>
<td>.81</td>
</tr>
<tr>
<td>Cognitive Involvement</td>
<td>.78</td>
<td>2.85</td>
<td>.76</td>
</tr>
</tbody>
</table>

Table 2 Statistical values of the 8-item-scales

5.2.2. Validity

A core aim of scale development is the establishment of a valid method of measurement [36]. Validity indicates whether a measure properly captures the meaning of the concept or construct it represents [37]. In contrast to reliability, validity is much more difficult to establish with certainty. Usually, it is distinguished between face or content validity (does the measure appear to measure what is aimed to measure), criterion validity (predictive and concurrent validity, i.e. does the measure allow to predict a future event or is it associated with another indicator that has already been shown to be valid), and construct validity (convergent and discriminative validity, i.e. to what degree does a scale measure a theoretical construct or trait). Although face validity offers a basic level of judgement [37], it is no objective specific value [37]. Nonetheless, we ensured face validity in theoretically deriving the items from the definitions of each construct and at the same time reverting to already existing items for each construct. Criterion validity, however, could be assessed in testing the scales’ sensitivity for the experimental manipulation, i.e. the secondary task participants had to fulfil when using the media environment. For if the scales are sensitive to the distraction, they can be used to measure different intensities of user presence in mediated environments. In G. Riva, F. Davide & W. A. IJsselsteijn (Eds.), Being there: Concepts, effects, and measurement of user presence in synthetic environments (pp. 3-16). Amsterdam: IOS Press.

<table>
<thead>
<tr>
<th>Scale</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Not distracted</td>
<td></td>
<td></td>
</tr>
<tr>
<td>n = 146</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distracted</td>
<td></td>
<td></td>
</tr>
<tr>
<td>n = 144</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-location**</td>
<td>2.55</td>
<td>.89</td>
</tr>
<tr>
<td>Possible actions*</td>
<td>2.44</td>
<td>.75</td>
</tr>
<tr>
<td>Involvement**</td>
<td>2.97</td>
<td>.71</td>
</tr>
</tbody>
</table>

Table 3 Scale means of distracted and not distracted participants

Convergent and discriminative validity, i.e. to what degree all scales did respond to the distraction, they can be used to measure different intensities of user presence in synthetic environments, too: The two Spatial Presence scales are highly intercorrelated ($r = .76, p < .01$), the involvement scale correlates significantly positive with the self-location scale ($r = .38, p < .01$) and the possible actions scale ($r = .45, p < .01$). These findings are in line with theoretical assumptions, as involvement is substantial connected with Spatial Presence, but at the same time is distinct from it.

6. Conclusions

Altogether, Spatial Presence and involvement have been demonstrated to be theoretically and empirically differentiable. All scales developed were sensitive for distraction during exposure, thus indicating criterion validity. Construct validity has been supported by inter-scale correlations. Moreover, the three scales yielded very satisfactory reliability estimates. In contrast to former conceptualizations of presence and involvement the distinction of the different concepts and the scales developed allow to investigate the formation of Spatial Presence experiences theoretically based and in more detail than former research. Although the MEC model here concentrates on user factors, media factors can be integrated as well.

Acknowledgements

The work presented here was conducted within the Information Society Technologies program of the European Commission (EC), project no. IST-2001-37661, and has been granted by the European Commission and the Federal Office for Education and Science of Switzerland. We thankfully acknowledge the support of the European Commission and the Federal Office for Education and Science of Switzerland.

References


Temporal and Spatial Variations in Presence: A Qualitative Analysis

Maia Garau, Hila Ritter Widenfeld, Angus Antley, Doron Friedman, Andrea Brogni, Mel Slater
Department of Computer Science, University College London (UCL), London, UK
{M.Garau | H.RitterWidenfeld | A.Antley | D.Friedman | A.Brogni | M.Slater}@cs.ucl.ac.uk

Abstract

This paper presents the qualitative findings from an experiment designed to investigate breaks in presence. Participants spent approximately 5 minutes in an immersive CAVE™-like system depicting a virtual bar with five virtual characters. On four occasions the projections were made to go white to trigger clearly identifiable “anomalies” in the audiovisual experience. Participants’ autonomic responses were measured throughout to investigate possible physiological correlates of these experienced anomalies.

Our analysis of the interviews with participants suggests that these anomalies were subjectively experienced as breaks in presence. The findings also reveal that breaks in presence have multiple causes and can range in intensity, resulting in varying recovery times. In addition, presence can vary in intensity within the same space, suggesting that presence in an immersive VE can fluctuate both temporally and spatially.

1. Introduction

In this paper we present a qualitative analysis of interviews from a 30-person experiment designed to investigate breaks in presence (BIPs). During the course of a highly immersive experience in a virtual environment (VE), participants experienced four deliberately triggered “anomalies” designed to remind them that they were not in the virtual scene depicted in the Cave¹, but in a laboratory conducting an experiment. The purpose of triggering these anomalies was to investigate whether these clear interruptions in the mediated experience could be associated with an identifiable physiological signature. A method for using GSR data to successfully predict when these anomalies occur will be described elsewhere.

Presence is of interest to practitioners in a number of fields including engineering, computer science, psychology, cognitive science, communication and philosophy [5, 10], as well as telecommunication and teleoperation [8]. Research has been driven both by theoretical and practical concerns [10], since a heightened sense of presence is considered essential for effective psychotherapy [7], for performance in training simulations [2], and for a wide variety of other VE applications.

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.

The debate concerning the definition and determinants of presence is accompanied by open questions concerning its measurement. A number of measurement approaches have been proposed. These can be classified according to how the measurement is taken (whether during or after the experience), and the type of data collected (whether subjective or objective). Increasingly, researchers are investigating ways of combining multiple measures.

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. The purpose of the qualitative analysis presented in this paper was to shed light on how participants subjectively characterized these anomalies, and specifically whether they experienced them as BIPs. If so, our findings could inform further explorations of non-intrusive ways of identifying BIPs during the mediated experience.

In the following section we introduce related work on presence measurement, referring specifically to the rationale for studying BIPs. Next we describe our experimental procedure and the agenda for the semi-structured interviews. We then discuss the qualitative method used for our analysis, and present our findings. We conclude with a discussion of our results and propose directions for future research.

2. Related work

Presence research has relied extensively on subjective reporting, using post-experience questionnaires such as the SUS [16] and PQ [20] to assess participants’ sense of “being there” in a mediated environment. However, questionnaires present a number of drawbacks in that inaccurate recall and demand characteristics can distort results. They also capture post-hoc rationalisations, as opposed to the experience itself.

It has been argued that rather than being a stable constant throughout the mediated experience, presence may vary over time [3, 8, 13]. Slater and Steed explored a “breaks in presence” (BIPs) approach, asking participants to signal each time their sense of presence in the VE was interrupted by a sudden awareness of their physical surroundings [15]. They report a strong positive correlation between questionnaire-based presence and presence as estimated by the number of reported BIPs.

Alternative approaches have been explored to study temporal variations without requiring participants’ conscious attention. Freeman et al. investigated the use of behavioural measures, studying postural shifts in response...
to motion stimuli [6]. Meehan et al. investigated the use of physiological measures including skin temperature, heart rate and GSR to measure objective responses to a virtual “pit” room containing a steep drop-off to the floor below [11]. The advantage of both these approaches is that they potentially offer a graded measure of the objective response. However, they both require specific stimuli to capture the responses, and are therefore limited by content-dependency.

Slater, Brogni and Steed sought to address this limitation by exploring the hypothesis that breaks in presence are associated with an observable physiological signature [14]. Their findings reveal a close association between reported BIPs and increases in GSR and heart rate and therefore have implications for the non-intrusive observation of responses to a wide variety of VEs. However, they caution that the physiological changes may be at least partially caused by the act of signalling the BIP.

The experiment reported in this paper expands on previous BIP studies by disambiguating the physiological responses from the signalling of BIPs; participants in the main condition were not briefed on “transitions to real” and were not asked to report them. By inducing “whiteout” anomalies in the experience, we also sought to match physiological findings with clearly identifiable anchor points in participants’ experience. Our goal was to compare any physiological findings with participants’ subjective accounts of the whiteouts, and thus to investigate whether lower-level autonomic responses and higher-level cognitive responses presented a coherent or contradictory message.

Spagnolli and Gamberini [18] describe a study exploring participants’ responses to technical breakdowns in the course of immersive interaction in a virtual library. Findings from their interaction analysis suggest that technical anomalies do not automatically translate into a state of “emersion”, but rather lead users to logically and actively incorporate the anomaly within the immersive experience. In our analysis we sought to further extend their research by comparing the effect of the brutal whiteouts with content- and apparatus-related anomalies, with a view to exploring the complex ways in which presence is enhanced and undermined by a variety of factors.

3. Experimental procedure

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 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 (EKG) 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. 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 in the Cave](image)

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. Users were wearing wireless trackers. The application was written on top of DIVE².

4. Semi-structured interview

One of the reasons for gathering physiological data is to shed light on participants’ involuntary responses during

² www.sics.se/dive
the experience itself. However, in this research we were also interested in understanding how participants themselves viewed their experience of interacting in the VE. For this reason, at the end of each experimental session we conducted an in-depth semi-structured interview on various aspects of the experience, in particular the causes and extent of any anomalies experienced, as well as responses to the virtual characters. A total of 30 participants were interviewed, but 27 were kept in the data pool because the audio quality on three interviews made them unsuitable for transcription.

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. Interview agendas are designed in advance to identify logically ordered themes [17]. The interview agenda contained “open” questions designed to avoid “yes/no” answers. We also avoided asking leading questions or using jargon. We deliberately avoided using the word “presence”, referring to BIPs as “transitions to real”, and to the deliberately induced anomalies as “whiteouts”.

We began with general questions, asking participants to describe the overall experience in the bar, and to highlight any factors that were surprising or that violated their prior expectations. We then asked them about their sense of being in the bar, and whether (and how) this might have changed over time. After this, we focused specifically on “transitions to real”.

In addition, we experimented with the use of visual graphs to help participants describe their presence experience. They were asked to draw a line representing the extent to which they felt they were in the bar versus in the laboratory over time. The use of these graphs helped to focus the discussion of why and how their sense of presence may have fluctuated during the experience.

5. Data analysis

The interviews were taped and then transcribed verbatim. The transcripts were analysed by combining two methods of qualitative analysis: content analysis [19], and thematic analysis [9].

First, content analysis was used to locate themes in the transcripts that related to our research questions; as mentioned, themes of interest included responses to the avatars, and the subjective experience of BIPs. As a “system of observation and empirical verification” [12], content analysis provides a research method that attempts to assess texts objectively. Its value is that it moves beyond subjective interpretation because the analyst develops categories before searching for them in the data [9]. The text is then ordered into manageable content categories by coding words or phrases related to the research questions. Next, each content category is quantified by counting the number of times it appears in the data.

Next, thematic analysis was used to provide a more in-depth view of the data. Where content analysis looks for preconceived themes in the data, thematic analysis searches for additional ideas that are not linked to the initial research questions [9]. The combination of these two methods allowed us to classify preconceived themes, as well as themes that emerged from the data itself. Throughout the analysis, an additional researcher checked the results against the data to provide credibility checks [1].

6. Findings

This section describes our analytical findings. First, we address the overall sense of presence, beginning with participants' response to the immediate questions. Next, we focus on the theme of temporal variations in presence, relating it directly to our research questions concerning the subjective experience of BIPs. Finally, we discuss a theme that emerged from the thematic analysis of the interview transcripts, concerning spatial fluctuations in presence. The findings are illustrated by direct quotes from the interviews with participants; participants are identified by number and gender.

6.1 Overall sense of presence

Analysis of the immediate questions showed that the majority of participants experienced a sense of being in, and responding to, the bar more than fifty percent of the time. The results are illustrated in Figure 2:

![Figure 2: Responses to the immediate questions](image)

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 rationalisations. Afterwards, they were able to expand on their answers in the semi-structured interviews.
6.1.1 The sense of "being in" the bar

In terms of the sense of “being there”, some participants expressed a sense of feeling drawn into the bar environment and forgetting about the spatial limitations of the Cave: “I did get the impression of being in a bar. I was quite surprised to the extent the bar extended out into the space beyond the wall. It felt like I should have been able to touch the bar” (P4 male). This expectation of being able to touch the objects resulted in a sense of surprise at feeling the physical boundaries of the Cave wall: “I was going to see... where, um, where I could put my hands, but then, obviously, it was on the wall and I realised I was just about to walk into it and I thought ‘woof’” (P12 female). This experience of touching the wall while expecting to reach out for virtual objects in the VE leads, in this instance, to a sudden awareness of the physical reality of being in the Cave, as opposed to the virtual bar.

Participants were asked whether they considered the virtual bar a place they visited, or images that they saw. The majority reported a sense of being in a place: “It’s like a place I went to because I won’t think of, ‘Oh, you remember that hologram’ or whatever he was. I’ll be like, ‘remember the barman?’ instead of ‘remember that image’... Definitely a place. Also, because it was so different from the space and it was definitely somewhere I went” (P21 female). The sense of being in a surrounding space populated by people contributed to the sense of being in a place: “Yeah, I felt more in a bar. Very much. Because the whole scene was, it was 3D, so I really felt that I was inside the bar and watching all those people speak and behave” (P9 female).

In addition to the visuals, many also said that the audio aspect of the experience added to their sense of being in the bar. In particular, they mentioned the music and the characters’ chatter as contributing factors: “I think I felt like being in a real bar. I think perhaps the music helps. And the fact that the people were talking. I felt that it was the environment of a real bar” (P2 male). This sense of being in the bar was not described as stable or constant, but was buoyed up by moments when the audio made the space come to life, for example when characters spoke: “The music helped a lot. It was moments I felt I was in a bar: like when people were talking” (P8 female).

6.1.2 The sense of "responding to" the bar

The combined audio and visual experience offered some participants a spatial sense of being in a virtual bar. This led them, in some instances, to respond to the bar as if it were real, for instance by instinctively trying to reach out and touch virtual objects. In addition to inanimate objects, some participants reported automatically responding to the virtual characters in social ways. One example was the attempt to engage in verbal interaction: “Rationally of course you know that it’s unreal because it’s an experiment, but it’s more of instinct, because once you are in a 3D thing, the music is there, and the people are there, they’re talking, and I said ‘hello’” (P3 male).

Participants often expressed surprise at their own social responses to the characters: “The man that was in the right side... They smile to me. I smiled back. It was like my reaction to a real situation. I am surprised because my response was as if they were real people. At the beginning, I didn’t expect to treat them as real people” (P2 male).

In addition to verbal interaction, some participants said they had tried to engage with the characters by waving at them: “I talked to them... I said ‘hello’ and I moved my hand in a queer way” (P20 male). Some also attempted to make physical contact: “The one in the shorts, the blue one in front of me. I touched him like this, on the shoulder. I wouldn’t put him on the shoulder, I would just go like this, Just like a real person” (P27 male). Interestingly, this touch is qualified as a socially acceptable form of touch, rather than an invasive or overly familiar gesture.

The virtual characters in the bar had fairly limited behaviours; for example, they did not drink out of their glasses, dance or move around the room, rather they stood in the same spot and made relatively muted gestures while speaking. In spite of this, many participants reacted emotionally to their body language and behaviour. Particular characters were often singled out for mention, such as the barman: “This barman that I would keep turning to look at... that look was very real. So I wanted to look at him...with him I could actually feel like ‘Oh my God, there’s somebody staring at me.’ The barman did not talk to me. I felt uncomfortable, like in real life, like when you know someone is staring at you and somebody doesn’t say anything to you” (P1 female). In this case, the relatively limited animation produced a realistic and powerful social response. Despite the fact that the barman did not look or behave in a highly realistic way, he was able to produce a sense of social discomfort simply by engaging in eye contact.

This sense of mutual gaze, combined with a purely accidental coincidence in animation, produced in some participants the sense of postural congruence, and made them wonder whether the characters were watching and imitating them: “I was trying to find out if the guy that was standing up was trying to maybe mimic me or not because he was, like, I was crossing my hands and he was doing the same at some point. And then I had the feeling he was looking at me, so then I tried to move from one side to the other to see if he was following me and he wasn’t, he wasn’t really following me” (P20 male). This example illustrates the fluctuating nature of the experience: at certain times, specific behaviours would coincide with expectations, causing participants to engage with the characters. However, lack of consistency ultimately undermined the illusion, making for a fluctuating sense of belief in the characters as sentient social entities.

There is some evidence that responses were partially shaped by participants’ individual characteristics. For example, one shy participant reported a significant sense of discomfort in the virtual bar, explaining that his response was comparable to what it would have been in an equivalent real-world situation: “I behaved reasonably as I’d behave in a real bar. Usually I do nothing really. I don’t particularly like bars. I think bars are, like, nervous social situations, because it is a situation where you are supposed to bond, impress other people, so I don’t...
particularly like those situations. I think I felt nervous before I entered the space. It had nothing to do with the virtual reality. It had something to do with the subject choice, or the object choice for the bar. Whereas I probably would have felt less nervous if it was like... I don't know... some less social situation” (P6 male).

However, several people behaved in a more open and 'daring' way compared to how they would usually behave in real life: "I was behaving like in a real bar, with maybe a little bit more staring, and a little bit more daring” (P1 female). Some usually shy people reported interacting with the avatars in a way that they would not ordinarily interact with real people: "I don't usually talk to a lot of people in the... in normal bars, but this time I felt like replying to them" (P26 male).

This section focused on participants’ overall sense of presence in the virtual bar, expressed as their sense of "being in" the bar and thinking of it as a place visited rather than as images seen. It also discussed some automatic behaviours reported by participants, that shed light on their spontaneous responses to the space and to the characters in the bar. The following section addresses factors that contributed to fluctuations in presence throughout the experience.

6.2 Temporal variations in presence

This section begins by summarising the presence graphs participants drew, depicting their sense of presence over time. It then describes various causes for BIPs, beginning with the induced whiteouts and continuing with factors relating to the apparatus and the virtual characters. It concludes by discussing how varying recovery times shed light on the varying intensity of experienced BIPs.

6.2.1 Graphs

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 3:

![Presence graph illustrating BIPs](image)

**Figure 3: Presence graph illustrating BIPs (P8 female)**

The graphs took one of four main patterns:

1. **High initial presence**: an initially low sense of presence, increasing towards the middle, and decreasing towards the end;
2. **Strong in the middle**: an initially low sense of presence, increasing towards the middle, and decreasing towards the end;
3. **Strong in the end**: a low initial sense of presence, steadily increasing until the end of the experience.

The graphs illustrate that the subjective experience of presence in the bar varied significantly between participants. For some, a high initial sense of presence gradually diminished due to insufficient stimuli in the VE. Others reported the opposite, explaining that it took them some time to habituate to the experience and become involved in it. Overwhelmingly, regardless of the overall shape of their presence graph, participants described an experience interspersed with interruptions (see Figure 3). The most obvious of these were the induced whiteouts.

6.2.2 Whiteout anomalies as BIPs

In the semi-structured interview, participants were asked about the induced anomalies. Specifically, they were asked to describe how many times the walls of the Cave had gone blank, what their response had been, and whether their reaction had been the same each time, or whether it had changed. Although there were four whiteouts, not all participants were accurate in their recall: “It happened three times. I think... The first time, it was like ‘Oh’... You know, it was like waking up, and the second time, it was like ‘Oh, it's happened again’” (P1 female). This statement illustrates the fact that the first whiteout appears to have had the strongest effect for the majority of participants. The first occurrence represented a sudden and surprising event in the experience, which participants sometimes attempted to explain to themselves in terms of a technical malfunction: “The first time I thought, like, a wire had gone loose” (P3 male).

However, after additional whiteouts, they often sought a plausible explanation for their repeated occurrence: “I didn’t know if the whiteouts were triggered or anything. I assumed that it was loading the next bit of the program, or something like that, or, just, like, a blip in the, I don't know... As it stopped it was just like a temporary jolt from your surroundings” (P10 male).

The sensation was described as similar to waking up from a dream. Although unclear in their cause and meaning, these induced anomalies had the effect of breaking participants’ sense of presence in the bar by reminding them of the apparatus and the laboratory. This was particularly the case after two or more whiteouts: “The second time I was like, ‘Oh no, they were doing it deliberately, to make me feel that this is artificial. That you’re still in the lab’” (P3 male).

The purpose of inducing the whiteouts was to generate clearly identifiable anomalies in the experience, in order to link any patterns in the physiological data with precise anchor points in the experience that participants could subjectively describe. The interviews reveal clearly that the induced anomalies were experienced as breaks in presence. However, additional causes of BIPs were also reported, including environmental factors relating to the technical apparatus.
6.2.3 Environmental factors as BIPs

Several “environmental factors” relating to the apparatus used in the experiment contributed to breaks in presence. Participants found the 3D stereoscopic glasses uncomfortable, and were aware of not wanting to damage what they suspected was fragile equipment: “Maybe the sensation of this thing on the glasses, because I’m not very comfortable. The worry that I would step on the cable and break your equipment” (P25 female).

The VE was deliberately designed to be approximately the same size as the Cave, leaving participants free to walk around the bar without needing to use a 3D mouse. The objects and characters in the environment were also spatially arranged such that all elements of interest were located along the back and side walls of the Cave. However, participants sometimes turned to face the back open wall, seeing the laboratory: “So when I turned back then I saw the curtain and all, I saw you guys on the computer, again, and then it was back to the lab feeling again” (P17 male). In addition to seeing the laboratory and the experimenters, other participants looked up and reported feeling shocked at seeing the projectors on the ceiling.

Certain aspects of the VE itself also undermined the sense of presence. One participant cited the inability to touch the virtual objects in the Cave: “Trying to touch something. If I try and touch the beer, I just think, ‘Yeah, this is virtual reality.’ It’s just when you’re looking that everything seems real” (P14 male). In addition to the lack of haptic interaction, certain visual elements also detracted from participants’ sense of presence, in particular the fact that not all visual objects appeared to be equally convincing in the bar: “The door behind both the barman… It was just, like the complete switching off” (P1 female). In comparison, the BIPs caused by the apparatus and the VE, a number of BIPs were caused by the appearance and behaviour of the virtual characters, as discussed below.

6.2.4 Virtual characters as BIPs

Just as the door stood out as an object in the environment that did not “flow” with the rest of the space, often specific virtual characters were singled out as less convincing than others. In one case, the female character on the left was described as undermining presence: “The girl in the corner because she, I think where she was in the corner it was kind of shadowed, so she didn’t look as real as the others, they didn’t look as real, but, how do you explain? She kind of didn’t look convincing, I suppose. So, when she said things and I looked at her, that, well, kind of reminded me that I was back in the experiment again” (P12 female).

In addition to appearance, the behaviour of the avatars sometimes had a significant impact: “Whenever I would look at the two who were just standing there, they seemed a little unreal… Because there wasn’t any movement and their movements, as opposed to the others, were a little more jerky, if you will. They were not very smooth. Their body language was unreal, it was inhuman. It was like, a reminder, ‘okay, you’re not in the bar’” (P1 female).

Both the characters’ appearance and behaviour served to undermine their role as social entities. Once belief in the bar as a social space was broken, it appears to have been irreparable: “When I realised that I can’t interact with the people, I think that I was in an experiment and that I want just to look around, I want just to grab, or feel the things, the objects” (P2 male). What is interesting is that once the belief in the characters was undermined, participants stopped treating the bar as a social environment and began exploring it as if they were alone, uninhibited by the presence of others.

6.2.5 Recovering from BIPs

Participants were able to recover their sense of presence after some BIPs. In many cases, recovery was apparently rapid: “I experienced a change very briefly when there was a break in the signal. There I did get that feeling, but it passed quickly. As soon as the signal came back on, I felt that I was back in the bar. It was pretty much complete and immediate” (P4 male). However, recovery became more difficult with the each successive whiteout: “It just got longer after the second and third break. You were just sort of, like, ‘Oh, okay, it's back again’ then, you know, back again, back again, let’s try to get back (laughs) again. Yeah, so it sort of lengthens after the second and third time” (P27 male).

In some cases, recovery was significantly longer: “About ten or twenty seconds. It wasn’t immediate. I turned back to look at you all” (P26 male). The act of turning back to look at the laboratory served to reinforce the BIPs. Also, more intense BIPs required active effort on the part of participants in order to recover a sense of presence in the bar: “Well to get back into it, that was almost like a positive… like an effort. Like, ‘Oh, okay. Now it’s back. Now what are they saying.’ It was kind of like that now. So it was like, it went off, and then I was like, ‘Okay…and then came back on so I had to focus on something in the bar to bring it back to life. It was an effort” (P1 female).

The analysis points to a range of intensity of BIPs, and a resulting range of recovery times. The BIPs caused by the characters resulted in relatively rapid recoveries: “A few seconds, maybe like two, three seconds. It wasn’t like immediately that ‘Okay, I’m involved in the conversation again’” (P1 female). In comparison, the BIPs caused by the whiteouts were generally more intense: “It’s possible to compare but at different levels. The lights going off were stronger feeling” (P1 female).

The whiteouts also had a stronger effect than environmental factors: “And my hand going straight through the bar. Or trying to touch something. Halfway down, maybe. It’s the light that takes you all the way down, like the complete switching off” (P14 male).

Participants experienced a longer recovery after whiteouts than character-related BIPs. The act of suddenly hitting the physical Cave boundaries had a similar effect to
the whiteouts and resulted, for some, in an even stronger BIP: “When comparing the flash versus the hitting the wall, I guess, probably bumping into the wall was more, sort of, a sharp reminder” (P18 male).

This section has addressed various causes of BIPs, and has presented evidence suggesting that BIPs range in intensity and recovery time. This supports the notion that rather than being a stable response, presence may vary through the course of the mediated experience. In the following section we present findings suggesting that presence also varies spatially within the same VE.

6.3 Spatial variations in presence

As discussed above, participants sought to maintain a sense of presence in the VE. This desire to avoid disruptions in presence also expressed itself in terms of where they chose to go in the environment. They tended to avoid spending time on the left side of the room, near the seated couple: “I didn’t seem to spend that much time on the left of the bar. Those people were further away. I couldn’t see them as well and I was a bit confused about that and the bar just kept getting me. The guys sitting down, I almost didn’t notice them because they were away, because there was that distance” (P21 female).

The couple on the left was located just beyond the boundaries of the Cave wall, so it was not possible to physically approach them as closely as the couple standing by the bar. For some, the seated couple also appeared more socially distant: “The sitting couple were very into their own conversation, and did not want me to join. Standing couple…. they tried to interact, smiled, tried to get me into the conversation” (P2 male).

Similarly, the bartender was often singled out as a more sympathetic and engaging character: “The bartender, he didn’t say much until the last part when he said you should order something (laughs), but he was smiling so I found that I was actually looking at him a little bit more than I looked at the rest…. He doesn’t speak, I think that’s something that’s quite (laughs)… It’s just something you remember because everyone is always talking, talking, talking and he doesn’t speak, but he always looks at you. He smiles sometimes so when he spoke at the last part, it was a nice change, I was like, ‘Whoaaa, he actually speaks.’ So I sort of, like, retained the best memory of him” (P27 male).

Participants generally gravitated towards the right side of the room partly for social reasons, because the characters appeared visually brighter and clearer, and more approachable. They also avoided areas where the ambient lighting made the Cave boundaries more evident, reminding them of a physical reality separate from the virtual bar. This, combined with the fact that participants put in effort to recover from BIPs, suggests that they sought to remain present by gravitating towards those parts of the VE that helped them remain present. This desire to remain present

7. Discussion

The primary goal of the qualitative analysis presented in this paper was to establish how participants subjectively experienced the whiteout anomalies. The analysis revealed that they did indeed perceive them as breaks in presence, likening the experience to a feeling of “waking up” or a “shock”: Experimental minders observing the participants’ behaviour in the Cave noted that the first whiteout was often accompanied by “startle” behaviours, with participants suddenly standing still and sometimes expressing verbal surprise. In many cases, physical responses to successive whiteouts were less pronounced. This observation tallies with our preliminary physiological findings indicating that the first whiteout resulted in a more pronounced increase in GSR. It is also consistent with participants’ explanation that they experienced a strong reaction to the first whiteout, but less of a surprise after subsequent whiteouts. With regard to the whiteout-related BIPs, the physiological and subjective data therefore appear to present a cohesive picture.

An additional goal of the analysis was to learn more about how the whiteouts, and other possible causes of BIPs, affected subjective presence over time. Participants were asked to draw graphs describing their feeling of presence over time, during their experience in the virtual environment. Although the graphs could not be quantified or directly compared, they proved to be a useful tool in focusing the interviews, and gaining a better understanding of the different ways people experience temporal variations in presence.

The analysis identifies a range of factors contributing to BIPs. These include the apparatus, the limited sensory modality of the VE (specifically the lack of haptics), insufficient consistency in the level of visual realism of the environment, and aspects of the appearance and behaviour of the characters. Our findings suggest that BIPs can have different intensities, resulting in varying recovery times. The majority of participants were able to recover more quickly from environment- and avatar-related BIPs than from the whiteouts. Also, BIPs appear in some cases to have had a cumulative effect, so that recovery time increased with subsequent BIPs, requiring greater effort on the part of the participant in order to feel present again.

One surprising finding was the notion that presence varies in intensity within the same space. The interviews revealed that participants had a pronounced preference for specific areas within the VE that they perceived to be more presence-inducing. They sought out those areas where characters appeared visually brighter and clearer, and more approachable. They also avoided areas where the ambient lighting made the Cave boundaries more evident, reminding them of a physical reality separate from the virtual bar. This, combined with the fact that participants put in effort to recover from BIPs, suggests that they sought to remain present by gravitating towards those parts of the VE that helped them remain present. This desire to remain present
is consistent with Spagnolli and Gamberini’s finding [18] that participants experiencing a technical anomaly sought to address it within the logic of the VE, rather than acknowledge it as a BIP.

In summary, our findings support the view [3, 8, 13] that presence is not a stable response. Our content analysis offered insights into how presence varies temporally as a result of apparatus, content and other factors. The thematic analysis allowed us to explore new themes that emerged from the data, and highlighted the notion that presence can also vary spatially within the same environment.

8. Conclusions and future work

This paper presented the qualitative findings from an experiment designed to investigate breaks in presence (BIPs). We sought to investigate presence as a multi-level phenomenon encompassing both involuntary autonomic responses and subjective perceptions. By triggering clearly identifiable whiteout anomalies in the experience, we were able to directly anchor participants’ subjective accounts of breaks in presence to specific points of the experience. Our preliminary physiological findings link the whiteouts to an increase in GSR. The fact that our qualitative analysis also qualifies the whiteouts as subjectively experienced BIPs is encouraging, and suggests fruitful avenues for further research into the use of physiological measures to study the temporal fluctuations in presence during any mediated experience.

Our findings offer insights into the subtle ways presence can be undermined, linking different causal factors with BIPs of varying intensities. The analysis also suggests that in addition to varying in time, presence can vary spatially within the same environment, and that participants actively gravitate towards those areas that are more presence-inducing. In future we plan to conduct focused studies with a smaller number of participants, with a view to making detailed cross-comparisons between autonomic, behavioural and subjective responses, to explore the overall picture they paint of presence.

Acknowledgements

This research was supported by the EU FET Program, PRESENCE IST-2001-37927. The EPSRC-funded project EQUATOR (GR/N15986/01) offered generous support with running the experiment. We also thank Anthony Steed for his precious help with DIVE, and David Swapp for his invaluable Cave support.

References

Heart-Rate Variability and Event-Related ECG in Virtual Environments

Guger C.*, Edlinger G.*, Leeb R.+, Pfurtscheller G.+, Antley, A.#, Garau, M.#, Brogni A.#, Friedman D.#, Slater M.#

*Guger Technologies OEG, Herbersteinstrasse 60, A-8020 Graz, Austria, www.gtec.at
+Institute of Human-Computer Interfaces, Graz University of Technology, Austria.
#Department of Computer Science, University College London, United Kingdom
guger@gtec.at, edlinger@gtec.at, m.slater@cs.ucl.ac.uk

Abstract

Beside standard questionnaires physiological measures can be used to describe the state of Presence in a virtual environment. A total of 21 participants explored a virtual bar in a CAVE like system. The experiment was divided into a baseline-, training- and experimental phase. During the experimental phase breaks in presence (BIPS) in form of whiteouts of the virtual environment scenario were induced. The heart rate variability and the event-related ECG were calculated from the acquired ECG data. The study showed that the heart rate variability can be used as parameter that reflects the physiological state of the participant. The event-related ECG showed the effect of the induced BIPS and of speaking avatars on the subjects.

Keywords--- break in presence, Virtual Reality, parasympathetic system, sympathetic system

1. Introduction

An important aspect of a virtual environment is to induce a state of Presence. Presence may be thought of as the observer’s subjective sensation of “being there” in the virtual environment [3], although there are many other definitions, but this is not the point of this paper. A strategy for measurement of presence was proposed by Slater and Steed [7]. This work is premised on the idea of eliciting moments in time when ‘breaks in presence’ (BIPs) occur. A BIP is any perceived phenomenon during the virtual environment (VE) exposure that launches the participant into awareness of the real-world setting of the experience, and therefore breaks their presence in the VE. Examples include gross events such as bumping into a ‘Cave’ wall, getting wrapped in cables, through to more subtle effects such as revelations that come from seeing a tree as a texture map rather than a solid object.

The measurement of presence can be done e.g. with subjective measures such as questionnaires and objective measures. Objective measures include the monitoring of the Electrocardiogram (ECG), respiration, galvanic skin response (GSR) and also the Electroencephalogram (EEG). Galvanic skin response was used successfully to describe the relation between physiological response and virtual height exposure [5]. In the same study consistent changes in heart-rate and skin temperature were not observed. Changes in GSR signal were also useful to describe the anxiety during a virtual flight [11], whereas heart-rate, respiration and peripheral skin temperature did not show a correlation with the anxiety level. Changes in heart-rate and GSR signal had also a high level of correlation with Presence questionnaire scores and indicated the degree of presence [12].

Recently, it was been shown that there is a very likely correlation in heart-rate and GSR changes associated with reporting of a BIP [8]. However, beside the GSR changes other parameters extracted from physiological signals can be used to describe subject specific reactions to the experimental setup. Especially from the ECG different parameters can be extracted. Beside the HR, the heart-rate variability (HRV) can be used to describe the physiological behavior of the participant. The event-related heart rate response is useful to describe the reaction of the subject to an event (e.g. BIP).

The variability of HR is also influenced by the autonomous nervous system (ANS) activity. Statistical measures in time and frequency domains are used for the quantification of the HRV. Recent studies show that the parasympathetic and sympathetic nervous activities influence the HRV at different frequencies [10].

In general there are several effects that change the HRV:

1. respiratory sinus arrhythmia (RSA) mediated by respiration. This activity is responsible for changes of the heart-rate in 2-5 seconds intervals and is controlled
by parasympathetic activity. The sympathetic system is too slow to influence this frequency band;
2. the blood pressure regulation contributes to HRV in 10 second rhythms;
3. changes with a periodic length above 20 seconds are mediated by the sympathetic system;
4. changes in the minutes and hours range are influenced by the neurohumoral oscillations in the circulating blood, by circadian rhythms or rapid eye movement phases during sleep.

Hence different methods able to delineate the different influences of the ANS are therefore necessary.

Low levels of sympathetic activity result in slow oscillations of sympathetic activity entrained to the vasomotor oscillations. However, as the level of sympathetic activity increases, these oscillations are damped and the fluctuations disappear. Under intense sympathetic drive, the heart-rate becomes metronomic in its regularity. The coupling between sinus node oscillation and sympathetic oscillations drops down to zero [1, 4].

The parasympathetic system also influences the sinus node. Respiratory oscillations affect both the sympathetic and parasympathetic nervous system. However, because of the slow response time of the sympathetic system, these rapid oscillations are mediated purely by the parasympathetic system. The parasympathetic oscillations have typically the same frequency as the respiration frequency.

It is generally assumed that high HRV correlates with good health and low HRV with cardiovascular disease. Other reasons for a decrease in HRV are mental stress, depression and exercise. Low frequency components (LF, 0.1 Hz) and high frequency components (HF, 0.15-0.4 Hz) indicate mental stress when the LF component is increased and the HF component is decreased. During dynamic exercise the heart rate changes but the HF component does not change significantly.

3. Experiment

A total of 21 subjects explored a virtual bar in the breaks in presence (BIPs) experiment conducted in a four-walled Cave-like system. This is an approximately 3 meter cubed area, with projection screens on the floor and three walls (but not the ceiling). Each experimental session was divided into three phases:
1. a baseline phase during which subjects were standing in the Cave with no images displayed;
2. a training phase during which subjects were given a simple navigation task to help them feeling comfortable moving around the VE;
3. an experiment phase during which the subjects were in the virtual bar.

During the experiment phase a scenario of a ‘bar’ was displayed to the subjects. There were 5 virtual characters in the bar, a barman and two couples, one pair standing near the bar, and the others sitting across the room. These virtual characters (2 women and 2 men) would be ‘aware’ of the location of the participant and would often address remarks towards him/her. During the experience two songs played in the background one after the other, and in addition there was background chatter as might be heard in a real bar. The entire experience lasted approximately 5 minutes. The system was implemented using DIVE [2] and the network interface between the various components achieved using VRPN. The experiment was approved by the Ethics Committee.

During the experimental phase avatars were speaking to the subjects and artificial breaks in presence (BIPs) occurred. About 30 sentences were uttered by the virtual characters, and BIPs were induced four times by creating a visual “whiteout” in the Cave. These ‘whiteouts’ lasted for 2 seconds each, distributed approximately uniformly in time throughout the bar experience.

3. Methods

The ECG was acquired as standard Einthoven I derivation (sampling frequency: 256 Hz) and the analysis was performed with the g.BSanalyze biosignal analysis software package (g.tec – Guger Technologies OEG, Graz, Austria).

The first step in ECG analysis is to detect QRS (ventricular contraction) complexes in the ECG time series. The QRS complexes determine the distance in time from one heart contraction to the next one (RR interval). The term “NN interval” is used in the literature to indicate that only normal-to-normal beat distances are used for the calculations (non-normal beats like extra systoles are excluded). The QRS complexes in the ECG data were detected automatically based on a modified Pan-Tompkins algorithm [6]. Then a visual inspection of the detected QRS complexes was performed to guarantee high data quality.

3.1 Heart-rate variability

Changes in RR-intervals are referred to as HRV and can be described in time and frequency domains.

The following are most important time domain measures:

MeanRR - mean RR interval [ms]
SDNN - standard deviation of NN intervals [ms]
MaxRR - maximum RR interval [ms]
MinRR - minimum RR interval [ms]
**MinMaxRR** - difference between MaxRR and MinRR [ms]

**MeanHR** - mean heart rate [bpm]

**SDHR** - standard deviation of the heart-rate [bpm]

The segmented measures divide the recorded ECG signal into equally long segments to calculate:

**SDANN** - standard deviation of the average NN interval calculated over short periods

**SDNNindex** - mean of e.g. 1 min standard deviation of NN intervals calculated over total recording length

The following measures yield differences between adjacent intervals:

**SDSD** - standard deviation of successive NN differences [ms]

**RMSSD** - square root of the mean squared difference of successive NN intervals [ms]

**NN50** - number of intervals of successive NN intervals greater than 50 ms

**PNN50** - NN50 divided by the total number of NN intervals

Frequency domain measures provide information on how power is distributed as a function of frequency:

RR time series were resampled with a frequency of 2 Hz. Then the power spectrum of the resampled time series were estimated with the Burg method [9] of order 15. The RR sequence was detrended and a Hanning window was applied prior to the spectrum estimation. The FFT length was 128 with an overlap of 64 and a sampling frequency of 2 Hz. Three main spectral components were distinguished:

(i) very low frequency (VLF): <0.04 Hz
(ii) low frequency (LF): 0.04 – 0.15 Hz and
(iii) high frequency (HF): 0.15 – 0.4 Hz.

The unit of these parameters is ms². To express LF and HF in normalized units, each parameter is divided by the total power minus the VLF component. This minimizes the effect of the total power on LF and HF. The LF/HF ratio describes the balanced behavior of both components.

**3.2 Event-related ECG (ER-ECG)**

A specific number of RR intervals before the event and after the event (BIP, avatar speaking) were averaged. The difference shows the event-related change in HR. This procedure was repeated for all speaking events and separately for all BIPS. An important parameter is the number of RR intervals (n) used for the averaging. Therefore, the ER-ECG was calculated for n=2 up to n=30.

### 4. Results

Table 1 shows the changes of HRV parameters in time domain for the two phases: (i) training and (ii) experimental phase. A sign test for paired samples was applied between the parameters of the training and experimental phase. The 4th column gives the corresponding p-values. If the p-value is not displayed then the changes were not significant.

The analysis shows that there is a significant difference between the training and experimental phase in terms of the heart-rate (MeanHR): 93.41 bpm versus 88.33 bpm (p=0.0072) and also in the heart rate variability (HRV) parameters. RMSSD, SDSD, PNN50 and the SDNNindex are smaller in the training phase than in the experimental phase.

<table>
<thead>
<tr>
<th>HRV</th>
<th>training phase</th>
<th>experimental phase</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>MeanRR [ms]</td>
<td>654.71</td>
<td>693.53</td>
<td>0.0072</td>
</tr>
<tr>
<td>SDNN [ms]</td>
<td>42.39</td>
<td>41.62</td>
<td></td>
</tr>
<tr>
<td>MaxRR [ms]</td>
<td>775.95</td>
<td>807.47</td>
<td></td>
</tr>
<tr>
<td>MinRR [ms]</td>
<td>573.47</td>
<td>606.02</td>
<td>0.0266</td>
</tr>
<tr>
<td>MinMaxRRDiff</td>
<td>202.48</td>
<td>201.45</td>
<td></td>
</tr>
<tr>
<td>MeanHR [bpm]</td>
<td>93.41</td>
<td>88.33</td>
<td>0.0072</td>
</tr>
<tr>
<td>SDHR [bpm]</td>
<td>5.84</td>
<td>5.22</td>
<td></td>
</tr>
<tr>
<td>RMSSD [ms]</td>
<td>22.51</td>
<td>27.08</td>
<td>0.0266</td>
</tr>
<tr>
<td>SDSD [ms]</td>
<td>22.61</td>
<td>27.21</td>
<td>0.0266</td>
</tr>
<tr>
<td>NN50 [nr]</td>
<td>5.42</td>
<td>8.33</td>
<td></td>
</tr>
<tr>
<td>PNN50 [%]</td>
<td>4.94</td>
<td>8.26</td>
<td>0.0118</td>
</tr>
<tr>
<td>HRVindex [nr]</td>
<td>8.57</td>
<td>8.90</td>
<td></td>
</tr>
<tr>
<td>SDANN [ms]</td>
<td>16.92</td>
<td>16.69</td>
<td></td>
</tr>
<tr>
<td>SDNNindex [ms]</td>
<td>37.46</td>
<td>40.86</td>
<td>0.0266</td>
</tr>
</tbody>
</table>

**Table 1: HRV time domain parameters**

Figures 1 and 2 show the HRV frequency analysis results for one subject. During the training phase the LF component is increased to about 180 ms² and the HF component is decreased to about 5 ms². In the experimental phase LF is about 118 ms² and HF is 18 ms².

The mean values for all subjects are represented in Table 2. Between training and experimental phase the normalized LF component (LFnorm) decreases from 74.5 to 64.9, the normalized HF component (HFnorm) increases from 15.1 to 22.3 and the LF/HF ratio decreases from 6.9 to 4.0. All three parameters have a p-value of 0.0266.
components in the HRV map around 0.35 Hz (HF component) and the sympathetic system yields frequency components around 0.1 Hz (LF component).

Basically the LF component is dominant throughout the whole experiment. The HF component in contrast varies between the baseline-, training- and experimental phase. The arrows indicate the changes between the different experimental phases. It can be seen that the activated HF component from the baseline phase becomes immediately smaller after the change to the training phase, but the amplitude increases towards the end of the training phase. With the change to the experimental phase the HF component disappears again and comes back after around 30 seconds. The HF component is present throughout the rest of the experimental phase. It is interesting that during the BIPs the HF component does not disappear.

### Table 2: HRV frequency domain parameters

<table>
<thead>
<tr>
<th></th>
<th>training Phase</th>
<th>experimental phase</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP [ms²]</td>
<td>1560.2</td>
<td>1262.3</td>
<td></td>
</tr>
<tr>
<td>ULF [ms²]</td>
<td>52.8</td>
<td>19.7</td>
<td></td>
</tr>
<tr>
<td>VLF [ms²]</td>
<td>202.8</td>
<td>86.0</td>
<td></td>
</tr>
<tr>
<td>LF [ms²]</td>
<td>1003.0</td>
<td>786.1</td>
<td></td>
</tr>
<tr>
<td>HF [ms²]</td>
<td>215.9</td>
<td>276.4</td>
<td></td>
</tr>
<tr>
<td>LFnorm [n.u.]</td>
<td>74.5</td>
<td>64.9</td>
<td>0.0266</td>
</tr>
<tr>
<td>HFnorm [n.u.]</td>
<td>15.1</td>
<td>22.3</td>
<td>0.0266</td>
</tr>
<tr>
<td>LF/HF [1]</td>
<td>6.9</td>
<td>4.0</td>
<td>0.0266</td>
</tr>
</tbody>
</table>

The change of the HR that was initiated by the speaking avatars was analyzed with the ER-ECG analysis. Therefore,
the mean HR immediately before and after each speaking event was calculated. This was done for each speaking event and the results were averaged to show the HR changes for each subject. For the speaker ER-ECG, the most significant change was found for n=3. Figure 4 displays the ER-ECG change of the RR intervals in %. The mean HR increase for all subjects was 0.7 bpm with a p-value of 0.04.

The opposite was the case for the BIPS. The HR decreased by 2.3 bpm because of the BIP (p=0.007). The most significant change was found with n=5. Figure 5 shows again the ER-ECG of the RR intervals in %.

5. Discussion

The time and frequency parameters display significant changes of HR and HRV between the training and experimental phase. The heart-rate decrease of 5.08 bpm could have 2 reasons:

1. during the training phase the subjects were specifically instructed to move around the VE to practice navigation, whereas in the experimental phase they were free to actively move around the bar or stay still if they preferred;
2. the subjects were more relaxed in the experimental phase when they were in the bar environment.

The HRV displays a high variability in the experimental bar phase. The time domain measures increased: RMSSD by 4.57 ms, SDSD by 4.6 ms, pNN50 by 3.32 % and the SDNNindex by 3.4 ms. In the case of frequency domain measures the LFnorm component decreased by 9.6, the HFnorm component increased by 7.2 and the LF/HF ratio is reduced by 2.9. The time domain difference measures RMSSD, SDSD and pNN50 correlate normally to the HF component in the frequency domain and are also in line in our experiment. RMSSD, SDSD, pNN50 and HFnorm are all increased. The time domain measures SDNN/SDHR correlate with the frequency domain measure TP. All these measures did not show a significant change.

It is well known that during dynamic exercise the HR is increased, but the HF component does not change significantly. In the experiment presented here the heart-rate changed and the HF component changed. Therefore it can be argued that the change was not initiated by dynamic exercise. Furthermore, an increased LF component and a decreased HF component normally indicate mental stress. This is precisely what happened in the training phases in our experiment, where subjects where more stressed than in the experimental phase. There are several possible explanations: the novelty of the experience of being in a VE, the unfamiliar equipment (including stereoscopic goggles and 3D mouse), and the novelty of entering the bar environment. Conversely, the positive relaxing effect of being in the bar could have caused the changes in HRV and HR.

The most surprising result is the increase of the HR when an avatar speaks to the subjects. In two subjects the ER-ECG of RR intervals was decreased by almost 3 % because of the speaking avatar. In the case of the BIP it is interesting that the HR was decreased because of the BIP. This can be explained by the sudden “whiteout” event, which surprised the subjects and changed their respiration rhythm.

Conclusions

A standard Einthoven I ECG derivation can be used to calculate the HRV and event-related ECG to describe the physiological state of participants in VR environments.
HRV is useful to compare different experimental states which last over several minutes. In contrast the event-related ECG can be calculated of intervals of some seconds and reflects changes of HR from a baseline interval to an active interval.

Acknowledgments

This project is funded by the European Union project PRESENCIA, IST-2001-37927.

References

Internal Consistency and Reliability of the Networked Minds Social Presence Measure

Chad Harms, Frank Biocca
Iowa State University, Michigan State University
Harms@iastate.edu, Biocca@msu.edu

Abstract
This study sought to develop and test a measure of social presence. The networked minds battery is proposed for a broad self-report measure of social presence. An experiment was conducted to test the internal consistency and criterion validity of the six constructs as determined by theory, specifically the ability of the measure to distinguish levels of social presence between (1) face-to-face interaction and mediated interaction, and (2) different levels of mediated interaction. The confirmatory factor analysis supports a model based upon a structure six distinct factors. In criterion validity tests the measure was generally sensitive to predicted differences between face-to-face and mediated interaction. On the other hand the measure was less sensitive to differences among low affordance and high affordance media, although the differences suggesting that text rated higher on perceived message and emotional understanding may provide some insight into the communication effectiveness of print media.

1. Introduction
Social presence was originally defined [2] as “The degree of salience of the other person in the interaction and the consequent salience of the interpersonal relationships,” (p. 65) and measured individuals’ perceptions of particular media. More recent scholarship of social presence has keyed in on the distinction of, “the social presence afforded by the [medium]” (p.73) and has measured the perception of the other with whom one is interacting [1].
Social presence for the purpose of the current research is defined in the following way: Social presence in a mutual interaction with a perceived entity refers to the degree of initial awareness, allocated attention, the capacity for both content and affective comprehension, and the capacity for both affective and behavioral interdependence with said entity.
The medium is a filter diminishing or magnifying the mix of sensorimotor, cognitive, and affective cues used to model of the other.

2. Dimensions of social presence
An analysis of the social presence literature has led us to conceptualize social presence as conceived of six sub-dimensions [1,4]. These dimensions affect the degree to which the user feels that the mediated other is accessible and responsive and include such constructs dealing with sensory accessibility (i.e., co-presence), psychological involvement (attentional allocation, perceived message understanding, perceived affective understanding), and behavioral interaction (perceived affective interdependence, and perceived behavioral interdependence). We briefly describe these below.
2.1 Co-presence

Co-presence is the degree to which the observer believes he/she is not alone and secluded, their level of peripheral or focal awareness of the other, and their sense of the degree to which the other is peripherally or focally aware of them.

2.2 Attentional allocation

Attentional allocation addresses the amount of attention the user allocates to and receives from an interactant.

2.3 Perceived message understanding

Perceived message understanding is the ability of the user to understand the message being received from the interactant as well as their perception of the interactant’s level of message understanding.

2.4 Perceived affective understanding

Perceived affective understanding is the user’s ability to understand an interactant’s emotional and attitudinal states as well as their perception of the interactant’s ability to understand the user’s emotional and attitudinal states.

2.5 Perceived affective interdependence

Perceived affective interdependence is the extent to which the user’s emotional and attitudinal state affects and is affected by the emotional and attitudinal states of the interactant.

2.6 Perceived behavioral interdependence

Perceived behavioral interdependence is the extent to which a user’s behavior affects and is affected by the interactant’s behavior.

3. Scale construction

Initially, three categories of social presence research were identified [1]. First, co-presence research dealt with the degree to which the observer believes he/she is not alone and secluded, their level of peripheral or focal awareness of the other, and their sense of the degree to which the other is peripherally or focally aware of them. Next, psychological involvement research identified the degree to which the observer allocates focal attention to the other, empathically senses or responds to the emotional states of the other, and believes that he/she has insight into the intentions, motivation, and thoughts of the other. Finally, behavioral interaction is the degree to which the observer believes his/her actions are interdependent, connected to, or responsive to the other and that the other’s perceived responsiveness are interdependent, connected to, or responsive to the observer’s actions. From these categorizations of social presence research, the six distinct dimensions of social presence identified above were established.

An initial pool of eighty-eight items was created. The items were created to reflect the identified dimensions. Some items were based on existing measures or were modified to meet the criteria for cross media generalization identified by [1]. As each item characterized a statement about the nature of the mediated social interaction, a Likert scale format would be used to measure each item.

The items were analyzed for their translation validity [3], specifically the face validity and content validity. Items were determined as to how well they captured the underlying structure and scope of the conceptualization and dimensions of social presence. A set of 5 researchers in social presence reviewed the initial item pool and specifically evaluated items that did not meet the criteria for cross media generalization. Nineteen items deemed problematic due to redundancy across items and confusing wording were removed. The sixty-nine item scale was tested in a pilot study [4] using 76 participants. Although the results were inconclusive, analysis identified certain items as poor indicators and exit interviews suggested that additional items were problematic due to wording that caused confusion. This information was used to finalize 50 items. In order to measure symmetry, each of the 50 items were reflected to measure the observer’s perception of the other’s response. The final result was a 100 item pool.

4. Validation study

This study was designed as an initial validation of the networked minds social presence scale. It used a between subjects experimental design in which participant were randomly assigned into one of three conditions: (1) face-to-face interaction, (2) mediated interaction via text-based low affordance media (3) mediated interaction via video-conferencing high affordance media.

4.1 Participants

240 students enrolled in a communication course at a large Mid-western university participated in this study for extra-credit.

4.2 Apparatus

This study used two sets of networked desktop PC computers supporting either low affordance text based or high affordance audio/video based interaction. Participants used one computer to interact with a confederate on a second computer located in a remote site. Each was isolated in order to eliminate distractions.
from the interaction. Face-to-face interactions were conducted in a separate room.

During the text-based low affordance media condition participants interacted with the confederates using AOL Instant Messenger. Over 95% of the students had previous experience with this application. Those who had never used IM were given instructions for the application.

In the video-conferencing high affordance media condition participants used Microsoft NetMeeting. The majority of the students had not used this application before but made easy use of the application once introduced. Two web-cams were used by the participant and the confederate along with two microphones and two headphones.

4.3 Measure

The initial Networked Minds Social Presence Inventory was made up of one hundred items. These items reflected the six hypothesized dimensions as well as self-report items. Participants completed the questionnaire including the measure online on a separate computer.

4.4 Procedure

Participants interacted with another student for approximately 5 minutes in a simply “get-to-know” interaction to determine their partner’s major, how their partner likes school, and what their partner does for fun in his/her free time. This was done both to give similar structure to the interaction.

Participants were randomly assigned to one of three conditions: (1) face-to-face, (2) text-based low affordance media (3) video-conferencing high affordance media. In the video conferencing condition care was taken to ensure that the participant could see and hear the confederate. This required slight adjustments to the web-cam. Once the interaction had started a timer was also started.

After five minutes the participants were told to wrap up their conversations. Participants were then moved to another computer to answer the Networked Minds Social Presence Inventory of items. Participants were provided instructions both on the questionnaire itself and by the investigating researcher.

5. Results

5.1 Confirmatory Factor Analysis

We used confirmatory factor analysis to test whether the factor structure of the Networked Minds Social Presence scale was consistent with the dimensional structure suggested by the theoretical analysis of the social presence construct. Because of the theoretical model involving several dimensions (latent variables) confirmatory factor analysis was more appropriate for effectively testing the parameters of the measurement model [5]. Confirmatory factor analysis, though not a sufficient test for construct validation, surpasses exploratory factor analysis, which often produces fewer factors than there are underlying variables in the data and disguises errors for bad items (p. 273).

5.2 CFA results

CFA was used to test hypothesis one that social presence would form six separate factors. Specifically, this study used four criteria to determine the quality and dimensionality of the social presence scale: face validity, reliability, internal consistency, and parallelism. Of the 100 items tested, 64 items were deleted in total to acquire an optimally sized scale. Items were removed due to low reliability, poor pair matching, and confusing wording as mentioned by participants in exit interviews. Support was found for hypothesis one in that social presence was found to form six separate factors based on the literature. After deleting problematic items and items with low reliabilities, 36 items were retained. These passed tests of internal consistency producing errors no greater than 0.18 and parallelism producing only 6 errors greater than .20 and none greater than 0.24. No trends were evident in the error matrix. Tables 2-7 provide all scale items retained for this study, their factor loading and descriptive statistics. A valid set of indicators was obtained for all six factors of social presence.

5.3 Reliability of the sub-scales

Chronbach Alpha tests indicated that the subscales identified by the factor analysis were internally consistent. Alpha reliability for all scales ranged at satisfactory levels between .81 and .87. See Table 3 for individual reliability scores, means, and scale items. All scale means are on a scale from one to seven.

5.4 Criterion-Related Validation Test

First, an analysis of variance test was conducted across confederates to ensure that there was no significant difference between any confederates. No significant difference was found across confederates on any of the six factors. Also, no trends among confederates were evident.

An analysis of variance was used to test the first criterion-related validity test of predictive validity; the ability of the measure to distinguish between levels of social presence experienced between unmediated face-to-face social interactions and mediated social interactions. The comparison of face-to-face interactions and mediated interactions across the six dimensions of social presence showed partial support for the sensitivity of the measure to predicted media differences. [see Table 1]

Another analysis of variance was conducted for the second test of criterion-related validity test concurrent validity, specifically the ability of the measure to distinguish between social presence experiences in low
affordance textual media and high affordance audio/video media. In the factor Perceived Behavioral Interdependence, text or the low affordance media (M=4.46, SD=0.84) was greater than the audio/video condition or high affordance medium (M=4.08, SD=.99), F(1, 158) = 2.61, p < .01, eta^2 = .14. Perceived Message Understanding also resulted in text (M=4.82, SD=0.88) scoring higher than audio/video (M=4.42, SD=0.97) for mediated interactions F(1, 158)= 2.74, p < .01, eta^2 = .06. [see Table 2]

6. Discussion

The confirmatory factor analysis support a model of social presence based up a factor structure made up of six distinct factors. The scale was consistent with the structure suggested by social presence theory. A strength of the assessment measure used in this study is that the items were grounded in prior social presence literature and research. The factor structure supported the construct validity of the Networked Minds Social Presence measure.

Each factor or subscale appeared to internally consistent as confirmed by Cronbach Alpha scores consistently greater than .80 across all factors. Continued validity tests will be necessary to further confirm this structure of the construct.

The current study supported two criterion-related validity tests of the measure. Consistent with predictions from theory, the social presence measure was able to distinguish between social presence experience of face-to-face interaction and mediated interaction. This pattern was found for four of the factors or subscales of the measure. Two of the factor subscales, perceived message understanding and perceived emotional interdependence yield null results, although the differences between conditions were in the direction predicted. In general this result was supportive of the measures ability to distinguish between levels of mediated and face-to-face levels of social presence.

On the other hand, the measure failed in the criterion-related validity test involving the ability of the measure to distinguish between two levels of mediated interaction involving different levels of social cues. Here the overall test of the measure indicated no differences between experiences in these two media. An analysis of the two subscales or factors, perceived message understanding and perceived emotional understanding, indicated that the low affordance textual medium provided greater social presence than the high affordance medium. This result is contrary to what would be predicted from theory and, furthermore, is the reverse result. This suggests that the measure may not be sensitive enough to detect differences in social presence across different media, and/or that differences predicted by theory either do not exist or are in a different direction than predicted.

Interpretations of the findings were inconsistent with theoretical predictions for the criterion tests.

7. Future directions

Research was recently completed comparing the relationships of interactants and media type. The focus of this research is how interaction, at a perceptual level, is influenced by various mediated channels and the interpersonal relationship between interactants. The design was a 2 X 4 independent groups experimental design where relationship (friend/stranger) is crossed with interaction conditions. Participants were randomly assigned to one of four media conditions: (1) face-to-face, (2) audio/video, (3) audio-only and (4) text. Results are forthcoming.[6]

Acknowledgements

Presence 2004 is organized by the International Society for Presence Research (ISPR) and supported by the European Union’s FET Presence Research Initiative.

References

Table 1

<table>
<thead>
<tr>
<th>Construct and Media</th>
<th>Mean</th>
<th>SD</th>
<th>t</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>Co-Presence</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Face-to-Face</td>
<td>5.11</td>
<td>.66</td>
<td>5.68</td>
<td>.01</td>
</tr>
<tr>
<td>Mediated</td>
<td>4.51</td>
<td>.24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attention Allocation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Face-to-Face</td>
<td>4.93</td>
<td>.22</td>
<td>3.95</td>
<td>.01</td>
</tr>
<tr>
<td>Mediated</td>
<td>4.44</td>
<td>1.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peer Message Understanding</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Face-to-Face</td>
<td>4.41</td>
<td>.26</td>
<td>1.14</td>
<td>ns</td>
</tr>
<tr>
<td>Mediated</td>
<td>4.27</td>
<td>.94</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peer Emotional Understanding</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Face-to-Face</td>
<td>5.08</td>
<td>.72</td>
<td>3.78</td>
<td>.01</td>
</tr>
<tr>
<td>Mediated</td>
<td>4.62</td>
<td>.94</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peer Behavioral Interdependence</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Face-to-Face</td>
<td>4.25</td>
<td>.93</td>
<td>5.34</td>
<td>.01</td>
</tr>
<tr>
<td>Mediated</td>
<td>3.96</td>
<td>1.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peer Emotional Interdependence</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Face-to-Face</td>
<td>3.79</td>
<td>1.01</td>
<td>1.72</td>
<td>ns</td>
</tr>
<tr>
<td>Mediated</td>
<td>3.54</td>
<td>1.08</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2

<table>
<thead>
<tr>
<th>Construct</th>
<th>Condition</th>
<th>Mean</th>
<th>SD</th>
<th>t</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>Co-Presence</td>
<td>Test</td>
<td>4.30</td>
<td>.38</td>
<td>-.20</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>Video</td>
<td>4.53</td>
<td>.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attention Allocation</td>
<td>Test</td>
<td>4.61</td>
<td>.27</td>
<td>1.29</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>Video</td>
<td>4.39</td>
<td>1.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peer Message Understanding</td>
<td>Test</td>
<td>4.46</td>
<td>.24</td>
<td>2.61</td>
<td>.01</td>
</tr>
<tr>
<td></td>
<td>Video</td>
<td>4.38</td>
<td>.59</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peer Emotional Understanding</td>
<td>Test</td>
<td>4.52</td>
<td>.32</td>
<td>2.74</td>
<td>.01</td>
</tr>
<tr>
<td></td>
<td>Video</td>
<td>4.42</td>
<td>.97</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peer Behavioral Interdependence</td>
<td>Test</td>
<td>3.45</td>
<td>1.18</td>
<td>-.14</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>Video</td>
<td>3.47</td>
<td>1.12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peer Emotional Interdependence</td>
<td>Test</td>
<td>3.50</td>
<td>1.05</td>
<td>-.52</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>Video</td>
<td>3.59</td>
<td>1.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Factor Items</td>
<td>Factor Loading</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>---------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Co-presence (M = 4.72, SD = 0.83) α = .84</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. I noticed (my partner).</td>
<td>.76</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. (My partner) noticed me.</td>
<td>.75</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. (My partner’s) presence was obvious to me.</td>
<td>.65</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. My presence was obvious to (my partner).</td>
<td>.64</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. (My partner) caught my attention.</td>
<td>.62</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. I caught (my partner’s) attention.</td>
<td>.64</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Attentional Allocation (M = 4.58, SD = 1.00) α = .81</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. I was easily distracted from (my partner) when other things were going on.</td>
<td>.71</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. (My partner) was easily distracted from me when other things were going on.</td>
<td>.61</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. I remained focused on (my partner) throughout our interaction.</td>
<td>.67</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. (My partner) remained focused on me throughout our interaction.</td>
<td>.63</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. (My partner) did not receive my full attention.</td>
<td>.58</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. I did not receive (my partner’s) full attention.</td>
<td>.69</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Perceived Message Understanding (M = 4.78, SD = 0.90) α = .87</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. My thoughts were clear to (my partner).</td>
<td>.52</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. (My partner’s) thoughts were clear to me.</td>
<td>.77</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. It was easy to understand (my partner).</td>
<td>.81</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. (My partner) found it easy to understand me.</td>
<td>.80</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17. Understanding (my partner) was difficult.</td>
<td>.71</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18. (My partner) had difficulty understanding me.</td>
<td>.73</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Perceived Affective Understanding (M = 3.72, SD = 1.14) α = .86</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19. I could tell how (my partner) felt.</td>
<td>.79</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20. (My partner) could tell how I felt.</td>
<td>.70</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21. (My partner’s) emotions were not clear to me.</td>
<td>.72</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22. My emotions were not clear to (my partner).</td>
<td>.69</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23. I could describe (my partner’s) feelings accurately.</td>
<td>.72</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24. (My partner) could describe my feelings accurately.</td>
<td>.68</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Perceived Emotional Interdependence (M = 3.62, SD = 1.06) α = .85</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25. I was sometimes influenced by (my partner’s) moods.</td>
<td>.81</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26. (My partner) was sometimes influenced by my moods.</td>
<td>.69</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27. (My partner’s) feelings influenced the mood of our interaction.</td>
<td>.73</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28. My feelings influenced the mood of our interaction.</td>
<td>.64</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29. (My partner’s) attitudes influenced how I felt.</td>
<td>.78</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30. My attitudes influenced how (my partner) felt.</td>
<td>.53</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Perceived Behavioral Interdependence (M = 4.32, SD = 0.91) α = .82</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31. My behavior was often in direct response to (my partner’s) behavior.</td>
<td>.58</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32. The behavior of (my partner) was often in direct response to my behavior.</td>
<td>.74</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33. I reciprocated (my partner’s) actions.</td>
<td>.71</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34. (My partner) reciprocated my actions.</td>
<td>.55</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35. (My partner’s) behavior was closely tied to my behavior.</td>
<td>.70</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36. My behavior was closely tied to (my partner’s) behavior.</td>
<td>.65</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Perception of Self-motion and Presence in Auditory Virtual Environments

Pontus Larsson¹, Daniel Västfjäll¹,², Mendel Kleiner¹,³

¹Department of Applied Acoustics, Chalmers University of Technology, Göteborg, Sweden.
²Department of Psychology, Göteborg University, Göteborg, Sweden
³Program in Architectural Acoustics, Rensselaer Polytechnic Institute, Troy, NY, USA

{pontus.larsson@ta.chalmers.se, daniel.vastfjall@psy.gu.se, kleinm2@rpi.edu}

Abstract

Apart from inducing a high sense of presence, creating a sensation of self-motion is often a key issue of many Virtual Reality applications. However, self-motion perception (vection) has been primarily investigated for visual stimuli. This study explored the possibility of inducing vection by realistic auditory stimuli. Furthermore, influence of various audio design parameters on auditory-induced vection was studied. The results suggest that sound source characteristic is a primary determinant of auditory-induced vection, especially for an environment with only a single sound source. However, it was also found that the type of sound source may play less of a role when the environment contains multiple sound sources. Auditory-induced vection is also depending on whether or not room acoustic cues are included in the simulation. It is however likely that it is the interaction between type of sound source and the environment that is important.

Keywords--- Self-motion perception, Vection, Auditory cues, Ecological acoustics, Binaural reproduction, Auralization, Presence.

1. Introduction

In many Virtual Reality (VR) applications creating a compelling sensation of self-motion is central. A large body of research reports that illusory self-motion can be elicited by visual stimuli. These illusory sensations of self-motion, often called vection, can be induced by large visual stimuli that move in a uniform manner. Observers then perceive as moving in the opposite direction to the visual stimulus [1].

However, in the real world, several other types of cues, primarily vestibular, auditory and somatosensory sensory signals, provide important self-motion information [2]. VR fun-rides, flight simulators and other types of motion simulators often utilize vestibular cues to enhance the sensation of self-motion.

Nonetheless, creating vestibular cues by means of e.g. motion platforms is most often technically complex, expensive and requires safety measures to be taken in order not to harm the user. It therefore seems more attractive to optimize other cues, such as auditory, to enhance the sensation of self-motion in VR-simulations.

However, research on illusory self-motion induced by sound-fields received little attention until recently [2-5]. One of the few studies available was performed by Lackner [2] where subjects were exposed to both real (loudspeaker array presentation) and virtual (dichotically presented) rotating sound fields. In this study it was found that subjects experienced self-rotation in both real and virtual conditions but that the real sound field was significantly more effective than the virtual one. Furthermore, when the contours of the experimental room were visible to the subject, auditory stimulation did not elicit illusory self-rotation.

The present study investigates some of the parameters of a rotating, virtual sound field, which may affect the illusion of self-rotation. The experiment resembles the study performed by Lackner [2] but our main goal here is to investigate how self-rotation is elicited by realistic stimuli such as those created for use in Virtual Environments and computer games.

2. Hypotheses

Based on the ideas of ecological acoustics [6,7], we believe that the character of the sound source is a relevant parameter that needs to be considered when studying auditory-induced illusory self-motion. Including easily recognizable sound sources in an acoustic motion simulation simply helps us resolving the conflict “I am moving” versus “The source is moving”.

For example, a rotating sound field containing sounds which immediately can be identified as being created by immovable objects, such as church bells or fountains, tells us “I am moving” while sound sources that are easily characterized as moving (such as sounds of driving cars, bicycles etc.) signal “The source is moving”. Synthetic sounds or, in general, sounds that do not provide any information on what type of source emitted them, are unable to provide such motion identification information.

Instead, the auditory system has to rely totally on binaural cues or room acoustic cues in trying to resolve the motion identification conflict.

Moreover, we hypothesize that the intensity of auditory-induced vection is contingent also on the number
of concurrent sound sources present in a rotating auditory stimulus, given that all sources move at the same angular speed. This is because we simply have more cues available when resolving the self vs. object motion conflict.

Furthermore, the current experiment also investigates the influence of acoustic rendering quality on auditory-induced vection. A recent finding from research in the visual domain is that spatial presence is a possibly vection-mediating factor [1]. That is, if we have the feeling that we are in a particular spatial context, we can also be more easily convinced that we are actually moving. An intriguing question is thus whether adding room acoustic simulation (auralization) can facilitate auditory-induced vection since it has been shown that post-experimental subjective ratings of presence is contingent on the quality of auralization [8,9].

It may also be the case that if room acoustic cues are included in the simulation, these cues directly facilitate the resolution of the object/self-motion conflict. From a physical perspective, it is clear that for some types of environments, such as asymmetrically shaped rooms, there is indeed a measurable difference in the listener’s position between rotating the listener and moving the sound source (see Figures 1 and 2).

An informal listening session on simulations of these two situations (listener rotation and source movement) confirms that they are indeed perceptually different, however further investigations have to be undertaken in order to reveal whether this difference has any effect on vection responses.

In sum, the hypotheses for the current experiment can be stated as;

H1: Rotating auditory stimuli consisting of one or more sound sources will elicit a stronger, more compelling sensation of self-rotation if the sound source(s) can be identified as being still compared to if the sound source(s) can be identified as being moving or if it is unidentifiable (artificial).

H2: A rotating auditory stimulus consisting of several concurrent sound sources will elicit a stronger, more compelling sensation of self-rotation compared to if the stimulus includes only one sound source.

H3: An acoustic simulation of listener rotation where realistic room acoustics cues are included will give rise to a stronger sense of presence and self-rotation compared to a simulation where only direct sound is rendered.

3. Method

The stimuli used to test the hypotheses presented above were all binaural simulations of a virtual listener standing in one place and rotating a certain number of laps. Sound sources that were included in the simulation were never actually moving; only the character of the sound source was varied as discussed below.

Stimuli were rendered offline in CATT-Acoustic v 8 [10] by using the “Walkthrough Convolver”. The parameters varied were:

1) Auralization quality (Marketplace or anechoic rendering).
2) Number of concurrent sound sources (1 or 3),
3) Type of input source sound (still, moving or artificial)
4) Turn velocity (20, 40 or 60 degrees per second)
5) Turn direction (left or right)

The acoustic model used to render the stimuli in this experiment is shown in Figure 3. The size of the model is approximately (W x L x H) 40x70x16 m. The main reason for choosing this model and not an indoor environment was that sound sources such as buses and fountains could be naturally associated with this type of outdoor context. Furthermore, a highly detailed visual model of this environment already exists which allows for performing similar experiments in auditory-visual conditions. Two different types of auralizations were created; one were realistic absorption and diffusion values were assigned to the model’s surfaces (RT= 1.2s @ 1 kHz) and one where only the direct sound (source-receiver) was included, i.e. an anechoic auralization.

By using the Walkthrough Convolver, various listener rotations could be simulated. The velocity of these listener rotations followed the following profile: 1) Stationary listener for 3s 2) Acceleration to maximum velocity (20, 40 or 60 degrees / second) for 3s 3) Constant rotation speed for 60s 4) Deceleration for 6s. Both left and right turns were simulated.

3.1 Apparatus

The experiment was conducted in a semi-anechoic room. Stimuli were presented with Beyerdynamic DT-990Pro circumaural headphones driven by a NAD Amplifier, model 3020.

Given the results in Lackner [2], some special measures were taken in order to achieve auditory-induced vection and to make the experience more convincing. First of all, a special seating arrangement, shown in Figure 4, was used. The arrangement consists of an ordinary office chair mounted on an electrically controllable turntable placed on a wooden base plate. The purpose of using this type of seat was to make the participant believe that rotational movements actually occurred during the experiment (although they in fact did not). Furthermore, the arrangement also prevented the participants from having any contact with footrests or the floor. In addition, four loudspeakers visible to the participant as he/she entered the test room, were placed around the experimental chair. Finally, the participant was also blindfolded during each trial.

3.2 Measures

As the subject heard the sound, he/or she was asked to report when vection was perceived simply by saying the
direction (left/right). The experiment leader then noted the
time and direction. After each trial, subjects were also
asked to fill in a single-page questionnaire containing the
following items:
1) Intensity of vection (0-100)
2) Compellingness of vection (0-10)
3) Sensation of source vs. self motion (-5-5)
4) Localization of sounds (0-6)
5) Sensation of sounds coming from different
directions (0-6)
6) Envelopment of sound (0-6)
7) Realism (0-6)
8) Magnitude estimation, presence (0-100)

3.3 Participants procedure and design

Twenty-six participants (13 female) with a mean age of
24 (SD 3.7) participated. They participated on two separate
occasions approximately two weeks apart.

A within-subjects design was used where Type of
sound source (3: moving, still, artificial), Velocities (3: 20,
40, 60 deg/sec), and Acoustic rendering: (2: marketplace-
anechoic) was varied for both single and multiple sound
sources. For each type of sound source three different
sounds were used. To avoid exposing participants to all
combinations, this variable was a between-group variable.
Also, the single and multiple sound sources were separated
and tested on different occasions. Thus, each participant
was exposed to 18 combinations on each occasion.
Different randomized orders of presentation were used for
each participant. However, all participants were exposed to
the single sound sources on the first occasion.

Upon arrival to the laboratory, the experiment leader
thoroughly instructed participants on the use of equipment
and scales. After a period of relaxation, the participant was
seated in the chair and the task (binary vection) and
questionnaire was then introduced. Following this, the
participant was blindfolded and a series of two test sounds
were replayed. After the test sounds, the participant was
again reminded about the procedure and the main
experimental task started. After completing all ratings,
participants were debriefed, thanked and paid for their
participation.

4. Results

The data was analyzed separately for single and
multiple sound sources, respectively. Initial data screening
included tests of possible effects of the between group
variable sound (three different types of sounds for different
participants) for each type of sound source, tests of effects
direction (left or right), as well as tests of demographic
factors (age, gender). No systematic effects of any of these
variables were observed and therefore they were discarded
in subsequent analyses. Also the three rating scales
concerning sound characteristics were mainly included as
controls and are not directly relevant for testing the
hypotheses. The only systematic effects on these scales
were, as expected, that the marketplace environment was
more immersive and that sounds could be better localized
than in the anechoic environment. For this reason, no
further data are presented related to these three ratings
scales. However, it should be noted that the fact that
participants rated the marketplace sounds as being more
easily localizable is somewhat surprising. A possible
explanation to this effect is that the higher realism and
increased externalization in the marketplace conditions
made participants believe that they could localize sounds
more easily, although they in fact could not.

4.1 Single sound sources

Binary vection and vection onset time. The number of
participants experiencing vection as indicated by the binary
vection measure is shown in Figure 5. Overall, vection was
relatively low (the range was 6-13 [23-50%] of the
participants), but it can be seen that the still sound sources
as expected induced higher vection than both moving and
artificial. This pattern was also more pronounced for the
marketplace environment than the anechoic. No systematic
effects of velocity were evident, why the data was collapsed
over these conditions. A McNemar-test on these data
showed that a significant higher number of vection
responses (36) was obtained in the still, than the moving
(21) or artificial (27) conditions for the marketplace
environment. This effect was however not obtained for the
anechoic environments where the number of responses in
the still sound source condition was similar (29) as the
moving (22) and artificial (24) conditions.

![Figure 5 Frequency of binary vection.](image)

Visual inspection of the vection onset-time suggests
that overall onset time is shorter for the anechoic than the
marketplace environment (Figure 6). Second, the artificial
sound sources resulted in longer vection onset time as
compared to still and moving sound sources for both types
of renderings. It should however be noted that because
these means are based on six to thirteen observations, no
inferential parametric statistical tests can be performed and
these results should be interpreted with caution.
Rating scales. Next, the rating scales that participants filled out after each sound presentation were analyzed. 3 (sound source) x 3 (velocity) x 2 (rendering) within-subjects ANOVAs were performed on each of the rating scales. Greenhouse-Geisser correction was used to correct for unequal variances. A significance level of p<.05 was adopted as a criterion for the inferential statistics.

Vection intensity and Convincingness of vection. The means for the intensity ratings are shown in Figure 7. The ANOVA for intensity yielded a significant main effect of sound source (F(2,25) = 5.66, p<.001) where the still sound sources were significantly higher (M = 36.3) than the moving (M = 20.3) and artificial (M = 20.1). Neither the main effects of velocity and rendering, nor any interactions reached significance (F>1).

Similar results were obtained for the convincingness ratings. A main effect of sound source (F (2,25) = 7.12, p<.001) where the still sound sources were significantly higher (M = 5.1) than the moving (M = 3.4) and artificial (M = 3.4). A main effect of rendering was also found (F(2,25) = 4.76, p<.01, where the rendered marketplace was perceived as more convincing (M = 4.80), than the anechoic (M = 3.11). As may be seen in Figure 8 there was also a significant interaction between sound source and rendering (F(2, 38) = 3.90, p<.05), where the still sound source condition were more convincing in the marketplace than the anechoic environment. No other effects reached significance.

4.2 Multiple sound sources

Binary vection and vection onset time. The number of participants experiencing vection as indicated by the binary ego-motion measure is shown in Figure 9. Overall, vection was higher than for single sound sources, but still far from all participants experienced vection (the range was 7-17 [28-66%] of the participants). Figure 9 indicates that the stationary sound sources as expected induced higher vection than both moving and artificial. This pattern was also more pronounced for the marketplace environment than the anechoic. No systematic effects of velocity were evident, why the data was collapsed over these conditions. McNemar-tests on these data showed that a significant higher number of vection responses (50) was obtained in the still, than the moving (39) or artificial (36) conditions for the marketplace environment. This pattern was however not as pronounced for the anechoic environments where the number of responses in the still sound source condition was similar (42) as the moving (35) and artificial (33) conditions.
Visual inspection of the vection onset-time suggests that overall onset time in the anechoic and the marketplace environment are similar (Figure 10). There is however a trend across conditions that vection onset time is shorter for the 60 °/s velocity (as compared to 20 °/s and 40 °/s).

Vection intensity and Convincingness of vection. The ANOVA for intensity yielded a significant main effect of sound source (F(2, 25) = 12.15, p<.001) where the still sound sources were significantly higher (M = 39.2) than the moving (M = 30.6) and artificial (M = 29.7). A main effect of velocity was also found (F(2,25) = 7.74, p<.01) where the 60 °/s velocity induced higher presence (M = 38.5) than both 20 °/s (M = 30.5) and 40 °/s (M = 32.0). No other effects reached significance.

Presence and realism. For the magnitude estimation presence scale the three main effects all reached significance while no interaction effect was significant. A main effect on sound sources (F(2, 38) = 5.07, p<.01) showed that the still and moving sounds induced higher presence (M = 63.1 and 62.7, respectively) than the artificial (M = 52.3). For velocity, (F(2, 38) = 3.55, p<.05) the 60 °/s velocity induced higher presence (M = 62.2) than 20 °/s and 40 °/s velocities (M = 58.8 and 56.2, respectively). Finally, the effect on rendering F(1, 38) = 4.01, p<.01) showed that the marketplace environment gave rise to higher presence (M = 62.4) than the anechoic (M = 59.1).

For the realism scale a main effect of sound source was found F(2, 38) = 16.37, p<.001) where the artificial sound received lower ratings (M = 3.17) than did still and moving (M= 4.0 and 4.2, respectively). A main effect of rendering was also found F(1, 38) = 6.44, p<.01), where the marketplace environment (M = 4.4) was rated as being more realistic than the anechoic (M = 3.5). No other effects were significant.

4.3 Comparison of Single and Multiple sound sources

Binary vection. Overall, the frequency of “yes” responses on the binary vection measure increased with app. 20% when comparing single to multiple sound sources. Collapsed across rendering conditions the McNemar tests showed that the increase was significant (p<.05) for all cases.
5. Discussion

Overall, support was found for the hypothesis that still sound sources are more instrumental in inducing vection than both moving and artificial sound sources. Measures of vection and intensity/convincingness showed exactly this for both single and multiple sound sources. The effect was however more pronounced for single sound sources.

Second, some support was found for the notion that a realistically rendered environment may increase perception of self-motion. For single sound sources the marketplace resulted in slightly more vection responses and higher ratings of convincingness. While the effect of rendering on the binary vection response was replicated for multiple sound sources, no effects were obtained on intensity or convincingness. However, as expected, the realistically rendered environment received higher presence ratings for both single and multiple sources.

Third, in line with our hypothesis, multiple sound sources induced significantly more vection responses than the single sound source condition.

Finally, and somewhat unexpected, we found that velocity influenced vection for multiple sound sources. More specifically, the faster velocity simulations (60 °/s) seemed to induce more vection as measured on the binary measure and on the rating scales. Even though this effect was not predicted, it mimics results on visual vection [1]. The reason why the velocity effect was not found for single sound sources is unclear. However, this might again indicate that single sound source environments provide unstable reference frames, which are unsuitable for inducing vection.

In summary, the present results suggest that the type of sound source is a primary determinant of auditory-induced vection, especially for an environment where there is only a single sound source. The present findings however suggest that the type of sound source may play less of a role when the environment contains multiple sound sources. Auditory-induced vection is also depending on the rendering of the environment. It is however likely that it is the interaction between type of sound source and the environment that is important. Our results suggests that the rendering mainly affected ratings of vection for the still sound sources while it had little effect for ratings of the moving and artificial sound sources.

The finding that auditory-induced vection is higher for multiple sound sources and that rendering in those cases become less important is a result directly applicable to the development of a perceptually driven ego-motion simulator; in scenes with multiple sound sources, simple room acoustic simulation could be used with the computational effort instead being allocated to realistic rendering of sound source movements.

The present experiments have concerned rotational movement. However, we believe that the ideas of ecological acoustics can be employed in the case of linear vection as well, which is something that will be investigated in future experiments. Furthermore, the possibility of using other measures of auditory vection, such as nystagmus [2] and postural responses [11], will be explored in future work. Measuring motion after-effects and motion sickness may also provide further insights in the area of auditory-induced vection.

Acknowledgements

The work presented in this paper was supported by the EU FET Presence Research Initiative project POEMS (Perceptually Oriented Ego-Motion Simulation), IST-2001-39223.

References

Objective Presence Measures through Electric Brain Activity

Tassos A. Mikropoulos1, Evangelos Tzimas2, Georgios El. Dimou3
1Associate Professor, The Educational Approaches to Virtual Realities Technologies Laboratory, Department of Education, University of Ioannina, Ioannina, 45110, Hellas
2Ph.D. student, Educational Technology Unit, University of Ioannina, Ioannina, 45110, Hellas
3Professor, Educational Technology Unit, University of Ioannina, Ioannina, 45110, Hellas
{amikrop@cc.uoi.gr, me00620@cc.uoi.gr, gdimou@cc.uoi.gr}

Abstract
This article presents results on objective measures concerning the sense of presence in virtual environments. Electric brain activity of seven students was recorded and analyzed during their interaction with four different versions of the same virtual environment, a representation of a room. Our aim was to detect if some sensory, realism and distraction factors that are related with presence result on differences in electric brain activity, indicating attentional activity and visual awareness as a virtual environment is enriched with textures and objects. Our results show both an alpha band decrease and a gamma band increase with the virtual environment enrichment with textures and objects. These are indications that factors related with the sense of presence in virtual environments affect electric brain activity. Electric brain activity could be considered as an objective measure of presence in virtual environments.

1. Introduction

Presence is the main attribute, the defining experience for virtual reality (VR), appearing in almost all the definitions researched by specialists coming from different fields.

From a philosophical perspective, researchers such as Heim [1], Zhai [2] and Lévy [3] refer to Heidegger’s work “Being and Time” for whom presence is synonymous with being and is a function of temporality.

In the sociological context, Riva and Mantovani [4], Heeter [5], Lombard and Ditton [6], as well as Schuemie et al. [7] proposed a cultural concept for presence.

From a psychological perspective, presence is the defining experience for VR [8]. Biocca stated that presence oscillates among the three poles of physical, virtual and imaginal environments [9] and Lombard defined presence as a psychological state or subjective perception in experiences generated by human-made technologies [10].

In the research on presence in virtual environments (VEs), three important topics are under consideration: 1. Theories and taxonomies 2. Attributes and factors that contribute to a sense of presence 3. Methodologies for presence estimation and measurement.

Concerning the theories for presence, Schuemie et al. referred to several models [7]. Concerning these models, we believe that presence as non-mediation, exclusive presence and estimation theory could be considered as theoretical models for presence. With regard to presence taxonomies, Heeter has proposed three dimensions of presence: personal presence, social presence and environmental presence [5].

Concerning the attributes of presence, Witmer and Singer began from the necessary conditions for presence, arguing that it depends on the ability to focus on one meaningful, coherent VE set [11]. Regarding the factors that contribute to presence, Witmer and Singer considered a number of factors that influence presence, namely control factors, sensory, distraction and realism factors [11].

Independently of the methodology for presence measurement, one could formulate presence as the sum constructed out of the above factors, as proposed by Slater [12].

Concerning presence measurements, many researchers categorized them as subjective and objective measures. Subjective measures have been based on questionnaires, developed by many groups, and are the most commonly used. Slater et al. developed and tested their questionnaire, which pays much attention to the subjects’ sense of “being there” [13, 14]. Witmer and Singer constructed the Presence Questionnaire (PQ) based on their approach of involvement and immersion [11]. Schubert et al. developed the Igroup Presence Questionnaire (IPQ) based on the two
showing that electroencephalography (EEG) can be used to probe the relation of hemispheric functioning to both emotion and cognition [24, 25]. Lombard and Ditton also created a cross media presence questionnaire [18].

Concerning objective measures, Schuemie et al. categorized them in behavioral and physiological measures [7]. Behavioral measures follow Slater’s approach corresponding to observable responses to various stimuli measured for example by log files [13]. Physiological measures concern measurements using skin conductance, skin temperature and heart rate giving some positive results, although these types of measures are probably related to arousal and not directly to presence [19]. Electric brain activity might give some positive results on presence in VEs, as our first measurements on cognitive processing that takes place in VEs have shown [20].

Presence, as a key feature of VEs, plays an important role in all the domains of VR applications. Education is a domain where VR seems to be a powerful and promising tool. Pedagogical models have been proposed for the exploitation of VR in the educational process giving emphasis on constructivism [21]. Educational VEs have been developed and evaluated in various disciplines [22].

We believe that presence is one of the main attributes of VR that differentiates it from other Information Technologies giving learners an active role, which is one of the most important characteristics in the teaching and learning processes. This is in agreement with Slater et al.’s statement that, if humans are required to perform tasks within VEs, then surely it would beneficial for them to feel present in the environment in which the task was taking place [23].

The goal of this article is exploratory. We record the users’ electric brain activity in virtual environments in order to investigate the sense of presence using objective measures. The research is in accordance with findings showing that electroencephalography (EEG) can be used to probe the relation of hemispheric functioning to both emotion and cognition [24, 25].

2. Previous Research

Only a few studies report results on brain activity in virtual worlds that may lead to conclusions concerning presence.

Maguire et al. applied Positron Emission Tomography (PET) to measure regional cerebral blood flow changes while 11 normal subjects explored and learned in two virtual environments, one containing salient objects and textures and the other one was empty [26]. The findings showed that learning in both cases activated a network of bilateral occipital, medial parietal, and occipitotemporal regions. The first environment resulted in increased activity in the right parahippocampal gyrus, while the region was not activated in the empty environment. The authors suggested that these findings contribute to the encoding of object location into virtual environments. We believe that Maguire et al.’s data are related to factors that contribute to a sense of presence. Steuer’s vividness [8] and Witmer and Singer’s sensory factors environmental richness and active search [11] are attributes that affect users’ explorations in the Maguire’s VEs.

Schier recorded electric brain activity during a driving simulation task [27]. Alpha activity (8 – 13Hz) decrease was interpreted as users’ more attentional activity. Greater alpha activity was consistent with fewer attentional resources required for an experienced driver. These data are related to Witmer and Singer’s sensory factor degree of movement perception, realism factors such as scene realism and information consistent with objective world, control factors such as the degree of control, the immediacy of control and anticipation of events, as well as distraction factor selective attention [11].

Mikropoulos reported electric brain activity during the navigation of 12 students in a VE [20]. The students showed more attentional activity in the virtual rather in the real environment and lower theta activity (4 – 8Hz) in the VE indicating less mental effort. These results are consistent with Witmer and Singer’s distraction factor selective attention.

Moore and Engel’s fMRI study recorded increased brain activity in the occipital complex during users’ interaction with VEs [28]. Their findings showed a neural activity increase with perceived volume in comparison with two dimensional shapes. We believe that these results are correlated to scene realism and information consistent with the objective world.

Bischof and Boulanger proposed theta activity as a measure for evaluating VEs by assessing the ease of navigation in virtual mazes [29]. These results are related with active search, degree of movement perception and physical environment modifiability, factors that affect presence.

Farrer et al. investigated the feeling of being causally involved in an action that is a constituent of the sense of the self [30]. They studied eight subjects’ degree of control of the movements of a virtual hand in four different conditions using PET. Their results showed that the less the subject felt in control of the movements of the virtual hand, the higher was the level of activation in the inferior part of the right parietal lobe, as well as a reverse covariation in the insula. The data are closely related to presence.
All the above references are indications that brain activity could be an objective measure for the sense of presence in virtual environments.

3. Methodology

3.1. Participants

The participants were seven (7) students (mean age 25). All students were right-handed, did not report any medical problems, neither had they used any medication, alcohol or drugs in the last 24 to 48 hours before the experimental procedure.

3.2. Procedure and materials

The virtual environment we designed was a representation of a real room. We have developed it using the SUPERSCAPE Virtual Reality Toolkit on a windows based personal computer. Four different versions of the VE were projected on VR glasses with a head tracker. Each student was sequentially interacted with:

1. an empty virtual room without textures (VE1)
2. the same empty virtual room with textures (VE2)
3. the same virtual room enriched with three solid objects without textures (VE3)
4. the same virtual room with three solid objects and textures (VE4).

Each student was progressively interacted with a more enriched environment involving more scene realism and virtual objects that could drive active search and selective attention. This was done because we would like to study some of the factors affecting presence, starting from simple, empty virtual environments. There was a rest period with closed eyes between each VE.

Our research axis was to detect alpha-band oscillations decrease and gamma-band increase as a result of attentional activity [20, 27] and visual awareness [31] coming from the virtual environment enrichment. The factors that contribute to the sense of presence are involved in the four virtual environments and are related with attentional activity and visual awareness are vividness, sensory factors such as the environment enrichment, realism factors such as the scene realism and information consistent with the real world, as well as distraction factors such as selective attention.

3.3. EEG recordings

Brain wave activity was recorded using a Micromed 98 system with 256Hz sampling rate. The digital EEG data acquisition system had a bandpass of 3.5 – 70Hz. EEG activity was monitored over 19 scalp locations, using the 10-20 Electrode Placement System. All leads were referenced to linked ear lobes and a ground electrode was applied to the forehead. Horizontal and vertical eye movements were recorded with four electrodes placed round the eyes, used for EOG rejection. Electrode impedance was maintained below 20KΩ. For each condition (eyes closed, eyes open, VE1, VE2, VE3, VE4) a fast Fourier transform (FFT) was performed on artefact-free data in order to derive estimates of absolute spectral power in different frequency bands. EMG artefacts were controlled. For each VE the artefact-free epoch was 2s. The bands of interest for the present work were alpha \((\alpha = 7.5 – 12.5Hz)\) and gamma \((\gamma = 30 – 70Hz)\). The signals presented in this work concern frontal (F3, F4), parietal (P3, P4, Pz) and occipital (O1, O2) lobes.

4. Results

Figure 1 shows alpha band activity for the four different versions of our virtual environment. The brain signal represents power density. By power density we mean the normalized % percentage of the power \((\mu V^2)\) that is distributed in a specific frequency band.
Alpha desynchronization is observed in frontal (F3, F4), parietal (P3, P4, Pz) and occipital (O1, O2) lobes. The students’ alpha waves decrease when the students navigate into a virtual room with textures. A further decrease is shown when the three objects without textures appear on stage, with the greater decrease observed in the enriched environment with textures and objects with textures. All signals for VE2 to VE4 are lower in the occipital lobe, since this is mainly responsible for visual stimuli. The results give evidence of attentional activity and visual awareness during the students’ interaction with the four virtual environments. Moreover, these results show a greater attentional activity and visual awareness as the virtual environments go from an empty room to the same empty room enriched with textures, to the room with solid
objects without textures, to the room with the objects and textures.

Figure 2 shows the mean gamma band activity (power density) for the four virtual environments and the same lobes.

As it is shown, gamma oscillations increase with the enrichment of the virtual environment. These results show a greater visual awareness as the students interact with VE2, VE4 and VE3. The relative decrease of the signal for VE3, is because of the absence of the solid three dimensional objects that appear in VE2 and VE4.
5. Conclusions

The goal of this article was to give data coming from objective measures concerning some factors that are related with the sense of presence in virtual environments. As a start, we allowed seven students to navigate in four different VEs without being able to manipulate and control virtual objects. Our aim was to detect if some sensory, realism and distraction factors result on differences in electric brain activity, indicating attentional activity and visual awareness as a virtual environment is enriched with textures and objects.

Our findings on alpha decrease show an increase in attentional activity and visual awareness with the VE enrichment. We believe that scene realism and information consistent with the real world that are involved in the environment richness and selective attention cause the observed alpha wave decrease in the frontal, parietal and occipital lobes. This conclusion is in coherence with Mikropoulos [20] and Schier’s [27] findings on electric brain activity in VEs. This conclusion is also in coherence with other brain measurements, such as Maguire’s concerning the encoding of object location into virtual environments [26]. The proposed relationship between presence and visual attention and awareness is based on the findings reported by Steuer [8] and Witmer and Singer [11].

Our findings on gamma increase come as a complement on alpha decrease. We believe that our results correspond with the visual awareness of form and color introduced by the appearance of the three virtual objects [31].

It seems that electric brain activity could be considered as an objective measure for presence that gives results on the various factors related with the sense of presence in virtual environments.

Our research is in progress. Future investigations involve the study of alpha desynchronization as a function of the time a participant stays in changing VEs, as well as the study of other EEG bands as a function of other presence factors.

References


A Comparison of Two Presence Measures Based on Experimental Results

Espen Nystad, Angelia Sebok
OECD Halden Reactor Project
{espen.nystad@hrp.no, angelia.sebok@hrp.no}

Abstract

This paper describes a study performed to evaluate two commonly used measures of presence. The study involved 24 participants using a VR-based procedural training application with different technology types. A variety of performance measures and background information were collected during the study. These variables were factors that are assumed to be related with sense of presence. The presence questionnaires of Witmer & Singer and Slater, Usoh & Steed were used to assess sense of presence in the virtual environment. Half of the subjects received the W&S questionnaire and half received the SUS questionnaire. Following the data collection, a variety of correlations were calculated and evaluated to provide insights into the two measures of presence. The questionnaire of Slater, Usoh and Steed gave results that were most consistent with the presence construct defined by the authors.

1 Introduction

The main advantage of Virtual Environments (VEs) is that they offer the user the opportunity to be present in another world. This means that training can be offered in an almost-real environment, safe from the hazards associated with the real world. For safety-critical industries, e.g., the nuclear industry, this is a tremendous benefit.

This feeling of being present in a simulated environment is referred to as sense of presence (SOP). This paper discusses two commonly-used presence metrics. A study conducted at the Virtual Reality Centre of the OECD Halden Reactor Project provides the basis for this discussion and offers further insights into these presence metrics. The study reported here was a part of a larger study investigating the use of different technology types for training in VEs [8].

When conducting research on presence, it is important to know what is really being measured. There is still no agreement on how to measure SOP. To reach an agreement on which measure to use, one must find a measure that demonstrably assesses the concept defined as presence. For a measure of presence to be valid, the responses of a subject group to the measure should be correlated with factors that are assumed to determine presence. For instance, if the level of realism in the VE is assumed to affect presence, then different levels of realism should result in different responses to the presence measure. This is called construct validity. Two ways of determining construct validity are by 1) considering how well the measure can reveal differences between groups, and 2) by observing correlations with important factors [2]. This study will identify potential factors that should be associated with an enhanced sense of presence, and see how the two presence measures relate to these factors. The presence measures studied are the Witmer and Singer Presence Questionnaire (PQ) [10, 17] and the Slater, Usoh Steed inventory (SUS) [12,13]. The experimental results will be presented in terms of these expectations and their implications will be discussed.

2 The presence concept

To be able to measure presence, it is necessary to have a clear concept of what presence is. Singer and Witmer [10] define presence as: 'the subjective experience of being in one place when one is physically in another.' and in virtual environments as 'experiencing the computer generated environment rather than the actual physical locale'. Steuer [15] similarly defines telepresence as 'the extent to which one feels present in the mediated environment, rather than the immediate physical environment', and Sheridan defines virtual presence as 'feeling like you are present in the environment generated by the computer' [9].

Presence has been identified as a key component in the definition of virtual reality [15,20]. Steuer [15] argues for using presence as the defining trait of VR to make the definition independent of the technological implementation of the VR system.

In the present study we distinguish between presence and immersion. Immersion is seen as a characteristic of the VR technology and describes the degree to which the technology engrosses the user in the VE; the degree to which the VE surrounds the user and the degree to which the technology gives a realistic or high-fidelity representation of the VE. By realistic / high-fidelity representation is implied a representation that is close to the way we perceive the real world, e.g. with stereo (rather than mono), real time shadows (rather than no shadows) or high resolution (rather than low). This definition of immersion is in line with Slater [11].
3 Measuring presence

A common way to measure presence is by using questionnaires because they are fairly simple to administer and analyse. The SUS and PQ questionnaires have been chosen for two reasons. Both are frequently used measures of presence, and both consist of a relatively small number of questions, which make them convenient to use. The PQ consists of 32 questions. Of these, 19 have been through a process of empirical testing by reliability and cluster analyses. These 19 questions were used in the present study. Three subscales have been determined from cluster analysis. They address issues related to the level of involvement and control, the naturalness of the interaction, and the quality of the interface. The questions were derived from a literature review of factors assumed to influence presence. The factors were categorised as control, sensory, realism and distraction factors. The PQ questions are in the form of a Lickert-scale with seven response choices (e.g., ranging from “not at all” to “completely”).

Whereas the PQ measure is constructed from factors assumed to influence presence, the SUS measure is more directly related to the feeling of being in the virtual environment. There are different versions of this questionnaire; one with three questions and one with six questions. In the present study the one with six questions was used. The questionnaire includes three aspects: the sense of being in the VE, the extent to which the VE becomes more dominant than the real world, and the extent to which the VE is remembered as a place the user visited rather than images the user has seen. The SUS questions are in the form of a Lickert-scale with seven response choices.

Slater [11] has criticized the PQ measure for confounding the assessment of the user's personal characteristics and assessment of the immersiveness of the system. The questionnaire asks about the user's subjective response to various system factors. The answers will therefore vary according to individual differences of the users, not only according to variations in system characteristics. Slater is also critical of the way the questionnaire is based on assumptions about which factors should influence presence. He argues that the questionnaire "cannot be used to study the factors, according to W&S, that influence presence. The presence score is constructed out of those factors. It is their sum." [11].

Both the PQ and the SUS measures have been used extensively to study the effects of various factors on SOP. Youngblut and Perrin [19] did a review of roughly thirty studies where the PQ or SUS measures were used. They looked at factors that were related with each of the questionnaires, and reported for which of these factors there were consistent results across two or more studies. They found that PQ was positively related with display field of view (FOV), while for head tracking, task-related experience and gender, the results consistently showed no relationship with the PQ. The SUS questionnaire was positively related with immersive tendencies. Youngblut and Perrin concluded that the data did not give an indication about the validity of the measures, or whether they measured the same construct.

Some previous studies have compared the two presence measures. Youngblut and Huie [18] studied learning of procedures using either immersive or desktop VE technologies. There was no difference between the immersive and desktop VEs in SOP, experienced during training for any of the measures. Performance in the transfer of training test was positively correlated with the SUS measure, but not with the PQ. Youngblut and Perrin [19] used the two questionnaires in a study of VR maintenance training. They found a partial negative correlation between PQ and the Immersive Tendencies Questionnaire (ITQ, see section 4.3.2), and a partial positive correlation between SUS and ITQ. Both measures were negatively correlated with number of errors on a transfer of training task.

Usoh et al. [16] used the two questionnaires to compare a real office environment with a virtual one. The PQ was not able to distinguish between the two environments, while the SUS showed a marginal difference.

4 Factors associated with presence

A number of factors are believed to influence SOP. In this section, we describe the factors that were examined in this study and how they are assumed to relate to SOP. The factors are divided into system factors, performance and personal factors. The system factors are immersion/realism and usability, the personal factors are familiarity with the environment and immersive tendencies. A valid measure of presence should be sensitive to variations in these factors.

4.1 System factors

4.1.1 Immersion In general, SOP is expected to be greater in more immersive, realistic systems. Sheridan suggest that SOP is determined by the extent of sensory information, i.e. the fidelity or richness of the VE [9]. Similarly, Steuer describe sensory depth (the resolution or quality of the sensory information) as one factor influencing SOP [15], and Zeltzer claim that SOP "provides a rough, lumped measure of the number and fidelity of available input and output channels" [20]. Wittmer and Singer claim that SOP is related to scene realism, with FOV and dimensionality being factors that govern scene realism. They also suggest that a display that isolates the user from the physical environment may give a higher SOP [17]. Display systems with stereoscopic view can be expected to have associated higher SOP than monoscopic view displays because they provide more sensory information and realism. Large-screen, wide FOV displays are expected to be associated with higher SOP than standard desktop displays. HMDs with orientation tracking should make the user feel more immersed in the VE and more isolated from the physical environment.
surroundings, and therefore give a higher SOP than displays without orientation tracking or displays that isolate the user to a lesser degree.

4.1.2 Usability System usability is expected to be positively correlated with SOP. Particularly, SOP is assumed to be related to the control capabilities, e.g. the ability to modify viewpoint [9], response time [15], and the extent to which mapping between controls and effects in the VE are natural and predictable [15]. VEIs in which users can easily implement control actions lead to users feeling present. If the control mechanisms for moving about and interacting with the VE are intuitive and easy to use, they become more or less transparent and the user can focus on the environment itself. Conversely, if users must struggle to figure out how to take control actions, must wait for feedback, are unable to navigate easily, they will not be convinced that they are in the environment. One way to measure usability is by questionnaires. In this study Brooks’ usability test was used [1].

4.2 Performance

SOP is generally assumed to be associated with improved performance [9, 15, 20]. According to Stanney et al. [14], the claim that there should be a relation between presence and performance has face validity and is supported by perceptual and cognitive theories. While Draper, Kaber, and Usher [3] state that the evidence linking presence and performance is scarce, they identify a number of sources that claim that presence is associated with enhanced performance. However, the nature of the relationship is not clear. Draper, Kaber and Usher note that many of the factors commonly thought to determine presence, are also determinants of performance. Although this makes it difficult to discover if there is a causal relationship between presence and performance, it supports the assumption that there is a positive correlation between the two.

4.3 Personal factors

4.3.1 Familiarity with environment User familiarity with the corresponding real-world environment is expected to be associated with SOP in a VE. Witmer and Singer predict that SOP should increase as the VE becomes more meaningful to the user [17]. If users are familiar with the corresponding real-world environment, and if the VR model is sufficiently realistic, familiarity might be expected to be positively correlated with SOP. Users recognize the environment and it becomes meaningful to them. Users who have never seen the actual physical environment might be less likely to be convinced by a VR model.

4.3.2 Immersive tendencies Steuer predict that individual factors will influence SOP [15]. Slater, Usoh and Steed assumed that individual factors influencing SOP are the representation system used in a given context (visual, auditory or kinaesthetic) and the dominant point of view of information processing (internal or external) [12]. Witmer and Singer have constructed a questionnaire the Immersive Tendencies Questionnaire, based on the assumption that some people have a higher tendency to feel involved in a VE than others [10,17]. The ITQ tries to assess this tendency. If presence is in fact related to an individual’s ability to be drawn in, then immersive tendencies can be expected to be positively correlated with SOP.

4.4 Summary of expected results

Presence should be positively correlated with:
1. Level of immersion
2. Usability
3. Performance
4. Familiarity with the environment.
5. Immersive tendencies

5 Method

5.1 Participants

24 employees of the OECD Halden Reactor Project volunteered for the study. These consisted of 22 males and 2 females, with ages ranging from 25 to 61 years. Gender was not assumed to influence presence. Therefore no attempt was made to balance gender in the experimental design.

Participants were from three different work groups. These groups represented subjects with high, medium and low familiarity with the real-world environment depicted in the VE and the maintenance task performed in the study.

5.2 Experimental design

The experimental task was to learn a multiple-step maintenance procedure in a nuclear reactor setting. The procedure was divided into three parts, and all subjects learned all three parts of the procedure. Four different technology types with different levels of immersion were evaluated: desktop display, desktop display with stereo view, large screen with stereo view, and HMD with orientation tracking. See Table 1 for details of each technology type. Each subject used three of the four technology types, and each technology type was used by the same number of subjects. The order of technology types was balanced across subjects. SOP was assessed for each part of the procedure. Half of the subjects were given the PQ measure and the other half the SUS measure. All subjects answered the presence questionnaires after each of the three procedure training sessions. There were 36 responses for each of the two questionnaires. Since all participants did not use the same technology types, the design was treated as a between-subjects design instead of a repeated-measures design when analysing the data.
5.3 Independent variables

Subject group and technology type were the two independent variables in the study. The three subject groups varied in the level of familiarity with the environment and task. Four different technology types was used (Table 1). Given our definition of immersion and the discussion of immersion factors in section 4.1.1, the technology types can be ordered on a scale of increasing immersion primarily by increasing FOV. Where the FOV is identical, as is the case with the two desktop displays, the stereo display is more immersive than the mono display. The HMD has a FOV that lies between the desktop and large screen displays, and it has head-tracking. So it is more immersive than the desktop mono display. On the other hand, the resolution is lower. It is therefore difficult to say how this compares with the large screen and desktop stereo displays.

### Table 1 Technology types evaluated in the experiment

<table>
<thead>
<tr>
<th>Technology type</th>
<th>Approx. Physical FOV</th>
<th>Reso- lution</th>
<th>Depth effect</th>
<th>Head orient. tracking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Desktop mono</td>
<td>30° horiz. 20° vert.</td>
<td>1024x768</td>
<td>Mono</td>
<td>No</td>
</tr>
<tr>
<td>Desktop stereo</td>
<td>30° horiz. 20° vert.</td>
<td>1024x768</td>
<td>Stereo</td>
<td>No</td>
</tr>
<tr>
<td>Large screen</td>
<td>90° horiz. 45° vert.</td>
<td>2048x768</td>
<td>Stereo</td>
<td>No</td>
</tr>
<tr>
<td>HMD</td>
<td>55° horiz. 45° vert.</td>
<td>800x600</td>
<td>Mono</td>
<td>Yes</td>
</tr>
</tbody>
</table>

5.4 Dependent variables

The main variable of interest was SOP, measured with the PQ and SUS questionnaires. Several measures were collected to reveal potential relationship with SOP. These include objective measures and questionnaires.

#### 5.3.1 Usability

A 10-item questionnaire, Brooks' usability questionnaire, was used to evaluate system usability [1]. It included statements about system ease-of-use, learnability and complexity. The items used a five-point Likert scale, and the subject was asked to indicate the amount of agreement with these statements, from “Strongly disagree” to “Strongly agree”.

#### 5.3.2 Performance

Two types of performance were assessed in the study. Firstly, the number of incorrect actions taken by the participant during the training sessions were recorded in the data logs as errors. Secondly, transfer of training was assessed by having the subjects recall the procedure steps and the tools used while looking at pictures from the real reactor hall. The number of tools incorrectly remembered or forgotten were counted. The number of procedure steps omitted or incorrectly remembered were also counted.

The fewer errors the subject made, the better the performance. It is therefore expected that errors are negatively correlated with presence.

5.3.3 Familiarity with environment

Participants were asked how many times they had been in the real reactor hall which was represented in the VE. The answers were recorded in the following categories: 1-5 times, 6-25 times, 26-100 times, 101-300 times and more than 300 times.

5.3.4 Immersive tendencies

A questionnaire was given to all participants to assess their immersive tendencies, or their ability to get drawn into external events. The Witmer and Singer ITQ was used [10,17]. It has 18 questions with a seven-point Likert scale. The inventory consist of the subscales focus, involvement and games.

5.5 VR application

Three procedural training programs were developed to assess procedural learning. These training programs are based on a control-station change-out procedure that is used at the Halden boiling-water reactor. The procedures are small portions (8-12 actions) of the entire control-station change-out procedure.

A training application, VRTexp, was developed specifically for this experiment based on an existing training toolkit [7]. VRTexp was used for showing the procedures in a VR model of the reactor hall, and provided a geometric FOV of 45 degrees (figure 1). This application was built using an extended version of X3D, which uses Java 3D to visualize VRML models and allows the application to be run on a variety of VR technology types, including stereoscopic displays.

Users navigated and looked around in the VE with a mouse. In the HMD condition, users were able to change direction of gaze by moving the head instead of using the mouse.

5.6 Experimental Procedure

Each participant attended the study for four sessions held on separate days. Each session lasted from 30 minutes to 1 hour, depending on the material to be covered and the subject’s personal working speed. The order of session presentation was identical for all subjects. The first procedure was learned in the first session. A retention test was given during the second session. The second and third procedures were learned in the third and fourth sessions.

For all VR sessions, participants arrived at the VR lab and were shown the equipment they would be using for that day’s session. They received instructions on the use of the equipment, navigating in the VR model, and interacting with objects in the VE. Participants worked with a practice scenario for 10 minutes to familiarize themselves with selecting objects and navigating in the VE. This familiarization phase was offered for all technology types, to ensure that subjects had equal exposure to VR models in all technology types before beginning the training.
Following the familiarization phase, subjects began the procedure training. Subjects experienced 5 repetitions of each procedure. The first two repetitions were passive viewings; in the final three, the subjects were actively performing the procedures. The experimenter first gave the subjects verbal instructions on how the procedure should be performed. The subject then watched two virtual “videos” (i.e., animation sequences of the actions to be performed) of the procedure. In the first viewing, the subject controlled the rate at which the steps were presented. During the second viewing, the subject watched a video of the entire procedure. This played at a pre-determined rate that the subject could not influence. In both videos, text was presented at the bottom of the display to describe the step that was about to occur or actually occurring.

Following these first two viewings, the subject was then given a short (1-2 minutes) set of instructions for actively performing the procedure. Then the subject began doing the procedure him/herself. Here, the text to describe the steps was not presented. Subjects did have access to this text (by using a Help command). Further, if the subject was unable to remember the step, s/he could instruct the application to execute the task.

When a subject performed a procedure in the VE, s/he did so by selecting (i.e., clicking on) objects - generally the tool that would be used to perform the task. These were associated with an animation sequence. The subject selected a tool and the appropriate action occurred in the VE. If the subject selected an incorrect object, feedback would be given in the text field: “Please select another object.”

After each session, the subjects filled out the presence and usability questionnaires. After the first session, they also filled out a background questionnaire including environment familiarity. After the second session they filled out the ITQ.

5.7 Data handling and analyses

Upon completion of the data collection, analyses were conducted. Data were obtained from computer logs and questionnaires. The responses to each presence measure were summed across questions to get the total presence score or subscale scores. The scores were then standardised.

In the data analyses, Spearman's rank order correlation was used to calculate correlations between the presence measures and each of the hypothesized influencing factors. ANOVA analyses were used to look for differences in presence between the three subject groups and between the four technology types. A significance level of 0.05 was used.

6 Results

The following tables show correlations between reported sense of presence and various factors believed to be related to presence. For the SUS measure, the correlations with the total (sum) score is presented in the tables. For the PQ measure, correlation with the total score and each of the subscales involved/control (IC), natural (N) and interface quality (IQ) are presented.

6.1 System factors

6.1.1 Presence and level of immersion Analysis of variance was performed for the SUS and PQ measures respectively to see if there were any differences in presence for the four technology types. Neither of the measures showed any significant differences.

6.1.2 Presence and usability Usability, as measured by Brooks' usability test [1], was significantly correlated with the PQ total score and the involved/control and interface quality subscales (Table 2). It was not correlated with the SUS measure.

6.2 Performance

Correlation between presence and errors during the active training is shown in Table 3. There was a significant negative correlation with the SUS measure only in the third active repetition. There was no significant correlation between the PQ measure and errors during training.

The correlation between presence and retention errors on the memory test one day after learning the procedure are shown in Table 4. Separate correlations were calculated for tool errors and procedure errors. Tool errors were negatively correlated with the SUS measure and positively correlated with the PQ interface quality scale.

<table>
<thead>
<tr>
<th>SUS</th>
<th>IC</th>
<th>PQ</th>
<th>N</th>
<th>IQ</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>not sig.</td>
<td>R=0.39,</td>
<td>not sig.</td>
<td>R=0.69,</td>
<td>R=0.46,</td>
<td></td>
</tr>
<tr>
<td>p=0.02</td>
<td>p&lt;0.001</td>
<td>p&lt;0.005</td>
<td>p=0.005</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2 Correlations between presence and usability
### Table 3 Correlations between presence and errors during training

<table>
<thead>
<tr>
<th></th>
<th>SUS</th>
<th>PQ</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Session 3:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R = -0.38, p=0.02</td>
<td>not sig.</td>
<td>not sig.</td>
</tr>
</tbody>
</table>

Table 4 Correlations between presence and retention errors

<table>
<thead>
<tr>
<th></th>
<th>SUS</th>
<th>PQ</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Equipm. errors</strong></td>
<td>R=-0.87, p=0.0002</td>
<td>not sig.</td>
</tr>
<tr>
<td><strong>Procedure errors</strong></td>
<td>not sig.</td>
<td>not sig.</td>
</tr>
</tbody>
</table>

Table 5 Correlations between presence and familiarity with the environment

<table>
<thead>
<tr>
<th></th>
<th>SUS</th>
<th>PQ</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Focus</strong></td>
<td>not sig.</td>
<td>not sig.</td>
</tr>
<tr>
<td><strong>Involv</strong></td>
<td>R=0.67, p=0.0001</td>
<td>not sig.</td>
</tr>
<tr>
<td><strong>Game</strong></td>
<td>not sig.</td>
<td>not sig.</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>R=0.40, p=0.02</td>
<td>not sig.</td>
</tr>
</tbody>
</table>

Table 6 Correlations between presence and immersive tendencies

<table>
<thead>
<tr>
<th></th>
<th>SUS</th>
<th>PQ</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>System factors</strong></td>
<td>Level of immersion</td>
<td>positive</td>
</tr>
<tr>
<td><strong>Usability</strong></td>
<td>positive</td>
<td>none</td>
</tr>
<tr>
<td><strong>Performance</strong></td>
<td>Performance (Errors during training)</td>
<td>negative</td>
</tr>
<tr>
<td><strong>Performance (Retention errors)</strong></td>
<td>negative</td>
<td>negative **</td>
</tr>
<tr>
<td><strong>Personal factors</strong></td>
<td>Familiarity with environment</td>
<td>positive</td>
</tr>
<tr>
<td><strong>Immersive tendencies</strong></td>
<td>positive</td>
<td>positive *</td>
</tr>
</tbody>
</table>

### 6.3 Personal factors

#### 6.3.1 Presence and familiarity with environment

The number of times the subject had been in the real reactor hall was positively correlated with SUS, but not with PQ (Table 5).

The three subject groups were also investigated to see if there was a difference in presence between them. The three groups differed in their familiarity with the environment and also in familiarity with the procedure learned in the study. There was a significant interaction effect between subject group and presence measure, (F(2, 48)=16.13, p< 0.0001), see Figure 2. Tukey HSD post hoc-test showed a significant difference between the presence measures for both the high familiarity (p=0.01) and low familiarity subjects (p=0.001). The high familiarity subjects rated presence lower with the PQ measure than the SUS measure, while the opposite was true for the low familiarity subjects.

### 6.3.2 Presence and immersive tendencies

The SUS measure was positively correlated with the total immersive tendencies score, and specifically with the involvement subscale. The PQ natural subscale was negatively correlated with the ITQ focus subscale and the ITQ total score (Table 6).

### 6.4 Summary of presence findings

Table 7 summarizes the results of the study described in the previous sections. The expected relationship between presence and each of the factors is also given in the table.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Expected relationship</th>
<th>SUS</th>
<th>PQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>System factors</td>
<td>Level of immersion</td>
<td>positive</td>
<td>none</td>
</tr>
<tr>
<td><strong>Usability</strong></td>
<td>positive</td>
<td>none</td>
<td>positive **</td>
</tr>
<tr>
<td><strong>Performance</strong></td>
<td>Performance (Errors during training)</td>
<td>negative</td>
<td>negative **</td>
</tr>
<tr>
<td><strong>Performance (Retention errors)</strong></td>
<td>negative</td>
<td>negative **</td>
<td></td>
</tr>
<tr>
<td><strong>Personal factors</strong></td>
<td>Familiarity with environment</td>
<td>positive</td>
<td>positive **</td>
</tr>
<tr>
<td><strong>Immersive tendencies</strong></td>
<td>positive</td>
<td>positive *</td>
<td>negative (partly)</td>
</tr>
</tbody>
</table>

*: p<0.05  
**: p<0.01 

partly: correlation with one or more of the subscales of the PQ; but not with the total score.

Table 7 Summary of results
7 Discussion

In this study, we investigated a number of factors thought to be associated with sense of presence. We have evaluated the relationship between these factors and both the SUS and PQ presence scores. Each of the individual results will be discussed and the overall picture will be summarized.

7.1 System factors

The level of immersion varied in the form of screen size, stereoscopic or monoscopic presentation and head-tracking. There was no difference in reported sense of presence among the four technology types. This was true for both the SUS and PQ presence measures. The result is in accordance with the lack of difference between immersive and desktop technologies found previously for both measures [18] and the lack of a relationship with head-trackung found for the PQ [19]. Immersion is believed to be an important factor to lead the user to feel present in the VE. The measures may be lacking in sensitivity for this factor. Another reason may be that the technologies are too different. Usoh et al. [16] suggest that comparisons across different virtual environments, e.g. comparing desktop to immersive technologies may not be valid using subjective presence measures. However, other presence questionnaires have been found to be sensitive to immersion factors [4, 6].

The other VE-related factor thought to be associated with sense of presence is rated system usability. If the system is perceived as easy to use, it should contribute to (or at least not interfere with) the sense of being in the environment. The results indicate that only the PQ measure was correlated with subjective ratings of usability. The correlation was positive. The PQ questionnaire subscales that were positively correlated with usability were Interface quality and Involved / control. Since these two subscales include questions which deal directly with the quality of the interaction, they are expected to correlate positively with usability. The SUS questionnaire, on the other hand, does not specifically address quality-of-interaction issues and did not correlate with rated system usability.

7.2 Performance

SUS was negatively correlated with errors during learning and retention errors, as expected. This is consistent with the relationship found between presence and transfer of training by Youngblut and Perrin [19], and Youngblut and Huie [18]. It is not possible to say something definite about the relationship between presence and performance in this study, but it may be that familiarity with the environment and task is an underlying factor influencing both presence and performance. The PQ interface quality scale was partly positively correlated with retention errors. This is opposite to the expected relationship. The explanation for this may be connected with the negative relation between PQ and familiarity indicated in figure 1, although there is not a statistical significant relationship. If familiarity with the environment is in fact influencing both presence and performance, then the negative correlation between presence as measured by the PQ and performance could be mainly due to variation in the familiarity factor. In other words, performance may be more related to familiarity than to the PQ measure. PQ looks to be negatively related with familiarity and, as a consequence of this, positively related with performance errors.

7.3 Personal factors

The results showed a strong correlation between the SUS measure and familiarity with the environment. This finding is consistent with the relationship found between chess experience and sense of presence by Hoffman et al.[5]. That study used two questions from the SUS measure and two other similar questions. Both results show that presence as measured by the SUS questionnaire is related to the meaningfulness of the VE, and that presence can be enhanced when the content of the VE matches previous experience. The prior experience with the environment probably makes the user associate the virtual environment with the real one, and memories of the real environment add to the experience to make it more like the real thing. Meaningfulness of experience is one of the factors that Witmer and Singer predict will influence presence [17]. But no statistical significant relationship was found in our study between the PQ measure and familiarity with the environment. In this respect, the two questionnaires seem to measure different aspects of the presence construct. A lack of realism in interaction and selection of objects may underlie this difference. The interaction in performing the procedure tasks may have felt less natural for the familiar subjects, since they knew how the tasks are performed in reality. In the PQ measure, this could result in a low score on interaction quality and naturalness. The unfamiliar subjects may have been less critical to the way the task was performed, and rated these scales higher. The questions in the SUS measure are more about the feeling of being in the VE than about the tasks performed. So the familiar subjects may have felt more ‘there’ than the unfamiliar subjects. This points to a possibly important distinction between the two measures. The PQ seems to be oriented around the tasks performed in the VE, while the SUS is more oriented around the place itself.

Another personal factor believed to be associated with sense of presence is immersive tendencies. The SUS measure was correlated with the Immersive Tendencies Questionnaire, and this relationship was accounted for mainly by the involvement scale of the ITQ. This scale refers to the ability to get deeply involved in an activity or a stimulus, e.g. books, television or movies. The PQ measure was partly negatively correlated with the ITQ, for the natural subscale. The expected relationship was not found. This may be because, as Slater claims [11],
the PQ confounds individual differences and assessment of variations in system characteristics.

7.4 Summary and conclusion

For the five factors where a significant positive or negative relation was expected (counting the two performance measures as one factor), the PQ measure gave expected results for one factor: usability. The SUS measure gave expected results for three factors: performance, familiarity with the environment, and immersive tendencies. Both measures were, contrary to expectation, unrelated with level of immersion. There were no factors for which the measures gave similar, expected results.

Our findings indicate that the SUS and PQ questionnaires are tapping into two different aspects of presence. It seems that responses to the SUS measure are more dependent on personal factors than factors related to technology and the VR system. The SUS questionnaire indicates a sensitivity to user familiarity with the environment, which provides a strong indication of its real-world usefulness. SUS is also sensitive to performance. Some authors see the link between presence and performance as the main reason for studying the presence phenomenon [3]. This is a good argument for using the SUS measure. However, the possibilities for and quality of interaction is commonly seen as a determinant of presence, and a good presence measure should be sensitive for these factors. The SUS did not show a significant correlation with these factors as indicated by the usability questionnaire. This may be because the SUS questions ask about the feeling of being in the place, and not specifically about interaction with the environment.

In this study, the PQ was sensitive to usability. Youngblut and Perrin [18] reported that the PQ was consistently related with FOV. The PQ therefore seems to be more related to technology or interaction factors.

The SUS questionnaire fits better with the definition of presence – a sense of being in the virtual environment – and the expectations we had identified based on that concept. The PQ seems to assess a too narrow slice of the presence concept. This study suggests that a good deal of research is still needed to develop a valid measure of presence.

8 References


The Experiential Dimensions of Two Different Digital Games

Jari Takatalo*, Jukka Häkkinen#, Jeppe Komulainen*, Heikki Särkelä*, Göte Nyman*
*Psychology of the Virtual –Group, Department of Psychology, University of Helsinki
#Audio-Visual Systems Laboratory, Nokia Research Center
{jari.takatalo@helsinki.fi, jukka.hakkinen@nokia.com, jeppe.komulainen@helsinki.fi, heikki.sarkela@helsinki.fi, gote.nyman@helsinki.fi}

Abstract

A framework is introduced for measuring user experience in virtual environments (VEs). It has been developed in various VEs (e.g. CAVEtm) and applied here to two different digital games played with PC and two different displays. The framework integrates basic psychological constructs considered to be essential in creation of human experience in virtual environments. Included are perceptual-attentive, cognitive-emotional and motivational constructs. These constructs have previously been used in various presence and flow studies. In this study following four dimensions represent these constructs: Physical presence, Emotional involvement, Situational involvement and Performance competence. The results show how these dimensions vary across two different types of computer games played in two different displays. The role of the presence as a part of the human experience in VEs is also considered.

Keywords---experience, presence, flow, entertainment

1. Introduction

1.1. Experience

As we perceive the world around us we attend to features and events that interest us. These features make us think and consider our future actions [1]. They activate past memories in which the current situation is referred and reconsidered. Current features and events as well as past memories evoke emotional responses in our bodies, which are felt as different feelings. Feelings support our rational thinking process [2]. All perceptions, thoughts, memories and feelings that enter our awareness are shaping our experiences. The dynamical process of experiencing begins from the perception of an environmental feature or event. It evolves and continues as we change our focus and act upon our experiences [3].

Dynamical process depicted above is based on the trilogy-of-mind set of cognition, emotion and motivation [4]. This set forms our awareness and mind in consciousness and co-operates with perception, attention and memory in developing the subjective experience [3]. Understanding the psychological dynamics of the experience helps to consider the meaning and value as well as the quality and intensity of that experience to the person in particular situation [5]. Because of the complex nature of the human experience also the method to measure it should be multidimensional. The field of digital game research lacks this sort of an approach.

1.2. Presence-flow -framework

We have studied subjective experience in different virtual environments (VEs). Based on our previous work in high-tech CAVE™ Experimental Virtual Environment (EVE), a multi-dimensional framework of the human experience was developed [6].

Presence, i.e., the sense of being in a VE is a special psychological feature of the VEs [7] and thus the Big-Three [8] components (sense of space, feelings of realness and attention to the VE) of physical presence [9, 10] were included into our framework. The presence components were thought to cover perceptual and attentional aspects of the mediated experience.

To cover cognitive, emotional and motivational aspects in our framework we included constructs used in studies measuring the concept of optimal experience, i.e., flow [11]. Flow has been studied in various non-mediated and mediated environments, e.g., in WWW [12].

In the theory of flow, the development of an experience is based on the cognitive evaluation between users perceived skills and opportunities and challenges provided by the current situation [11]. The eight-channel flow model [13] integrates basic emotional components of arousal, valence and control [14] into this cognitive evaluation process. In the model these components are
considered outcomes of different skill-challenge situations, e.g., arousal is an outcome of a high challenge and moderate skill situation. The theory of flow [11] also covers motivation, i.e., personal relevance and interest of the user to the current situation.

Thus the framework was named as Presence-Flow – framework (PFF). Its three Varimax-rotated dimensions were composed of 13 scales measuring different aspects of the human experience in VEs [6]. Also other authors [15, 16] have acknowledged the need for such expanded frameworks to measure human experience in VEs.

1.3. Presence-involvement-flow -framework

In our current studies PFF has been developed to cover the special needs possessed by PC based gameworlds [17]. These gameworlds are technologically less advanced than the previously studied EVE, but the content they provide can be much richer.

In addition to presence and flow scales we included measures that were considered important in digital game context. Included were social as well as role building and drama/plot aspects in user engagement to the technology. Also various emotionally charged feelings considering the gaming event were included. These scales have been examined and factored in a larger sample collected from the WWW [17].

Based on these two data (n=68 [6] and n=164 [17]) an explorative four-dimensional framework was developed. The Presence-involvement-flow -framework (PIFF) measures Physical presence (and engagement), Emotional involvement, Situational involvement and Performance competence (and competing) in VEs (Figure 1).

2. Methods

2.1. Participants

80 participants were tested. They were mainly university students from the faculty of behavioral sciences and the department of computer science. There were 40 males (50 %) and 40 females (50 %). The mean age of the participants was 24.7 years.

The participants were selected by applying the background questionnaire. Based on the background questionnaire answers we excluded participants who did not like driving games, did not have any computer game playing experience or who reported playing computer games for six hours or more every day.

2.2. Technology

All experimental groups used the same computer (Pentium 4 CPU at 3.00 GHz – Total memory 512 MB DDR-SDRAM). The Display adapter used was Sapphire ATI Radeon 9600 - 256MB (8 x AGP) and Sound card Realetek AC97 Audio. There were two different display conditions 1) Olympus Eye-Trek FMD-700 near-eye display and 2) a 21 inch Sony Trinitron GDM-F520 CRT monitor from the viewing distance of 1 meter.

2.3. Task & procedure

Four different test groups were formed (20 participants each) and a 2x2 test design was used. Two of the groups played Need for Speed Underground, which is a 1st person 3D – driving game with lots of camera movement, horizontal changes and intensive flux. Microsoft sidewinder Gamepad was used to play NFS UG. Other two groups played Slicks n’ Slide 1.30d

![Figure 1 PIFF - structure](image-url)
which is a 3rd person, 2D – driving game with no camera movement and otherwise static environment. The participants used keyboard to play Slicks n’ Slide. One of both NFS and Slicks groups used Olympus Eye-Trek near-eye display. The remaining two of both NFS and Slicks groups used Sony Trinitron CRT monitor.

The participants were instructed to proceed in his/her own pace and not to ask instructions during the game play, if possible. However, they were assisted if insurmountable problems (i.e. technical or otherwise immediate) occurred. The task lasted for 40 minutes after which the subjects filled the EVEQ- questionnaire.

2.4. Scales used in EVEQ

To measure user experiences an EVEQ – questionnaire was used. EVEQ is a developmental tool, which includes 146 items, which are mainly collected from the previous presence and flow studies. These items can be further summed into 23 scales to measure different experiential constructs (e.g., presence and flow components).

Each of these 23 scales includes 4-10 items. The scales have been factored individually in two different studies (n=68 [6] and n=164 [17]) to ensure both the one-dimensionality of each scale and the fitting of the items into a scale. In this study the items are summed according to these previous studies and the scores of the summed scales are compared. Next we shortly describe the content of the 23 scales that form the four main dimensions. Also the Cronbach’s α of the four dimensions in this study are presented. To read more about the scales and items forming the scales the reader is referred to Takatalo [18].

1. Physical presence (α=.93)
   Action (Objects and things could almost touch me, game induced real motion feelings)
   Attention (Concentration on the game instead of the real world, time distortion)
   Real (Gameworld was natural, live and vivid)
   Spatial (Spatial awareness of a place, being part of the gameworld)
   Being there (Visited a virtual place instead of being in a lab, journey to the gameworld)
   Drama (Perception of a plot that affected behavior)
   Enclosed (How much did the media support the different senses e.g. vision, hearing)

2. Emotional involvement (α=.89)
   Mediari richness (e.g., how warm, close and sensitive the media was experienced)
   Valence (Was the experience negative or positive)
   Pleasant (Enjoyment, fun, time flew)
   Impressed (Strong experience and excitement)
   Involvement of the played game (The game was important, meaningful, interesting and appealing)

Playful (Free, flexible, natural, live)
Innovative (Creative, innovative, learning new skills)

3. Situational involvement (α=.80)
   Bored (Playing the game evoked boredom and anxiety)
   Challenge (Perceived challenges provided by the game)
   Arousal (Level of arousal evoked by the situation)
   Interaction SMR (Evaluation of the interaction speed, mapping and range)
   Involvement of the test situation (The test was important, meaningful, interesting, appealing and fun)

4. Performance competence (α=.62)
   Social presence (Acting/ competing with others, other actors were aware of the user)
   Skill (Experienced skill to play the game)
   Control (Sense of control over situation)
   Exploration (Ability to explore the environmental limits of the gameworld)

3. Results

3.1. The difference between the two games

The differences between the experimental conditions were measured with a one-way ANOVA. As it is shown in Figure 2, there was a significant difference in the Physical presence scores between the Need for speed (NFS) and Slicks players (F(1,78) = 19.75, p < .001). NFS provided more real motion feelings; it was experienced more natural, live and vivid. It was able to provide users more spatial sense of space and place in which the action took place. It enclosed users better to the visual and auditive environment. Its plot was stronger thus supporting better users role build-up. NFS also supported users more in suppressing the surrounding environment. To summarize, NFS provided more complete transfer to the game world.

NFS was also emotionally more involving than Slicks (F(1,78) = 14.05, p < .001). It was experienced warmer and more sensitive as well as more pleasant than Slicks. Playing NFS was more impressing and exciting. It interested the users more than Slicks. It was also more fun and appealing. Playing NFS felt more free, flexible, innovative and creative as compared to Slicks, which was felt more constrained and monotonous. Thus, NFS provided emotionally more intensive playing experience.

The playing situation of the Slicks was less involving than that of NSF (F(1,78) = 10.82, p < .01). Slicks was experienced more boring and frustrating. NFS was more arousing and it provided more interaction speed. Its interaction was also better mapped to the real world action and the range of its interaction was wider than that provided by the Slicks. Interaction technique
used in NFS gave the players more realistic feeling and they felt that the game responded more naturally and intuitively to their actions (1st person). Slicks was less involving and motivating, thus decreasing the experienced meaning of the situation.

NFS also evoked more social feelings towards the other drivers (competitors) and feelings of other drivers’ awareness of the user. It also offered more chances to explore the boundaries of the gameworld. However, the two games did not differentiate significantly in the Performance competence –dimension (F(1,78) = 3.84, p = .054).

3.2. The effect of the VR –display

Next the two games were compared within the display conditions (CRT – VR). Within CRT –condition the players of the NFS experienced more Physical presence (F(1,38) = 9.62, p < .01) and Emotional involvement (F(1,38) = 4.61, p < .05) than Slicks players. The two games did not differentiate either in Situational involvement or Performance competence.

When the two games were compared in VR display –condition NFS players experienced more Physical presence (F(1,38) = 10.54, p < .01), Emotional involvement NSF (F(1,38) = 9.42, p < .01) and **Situational involvement** (F(1,38) = 10.05, p < .01) than Slicks players. The two games did not differentiate in Performance competence –dimension in VR –condition.

Within the Situational involvement dimension NFS played with VR was experienced more arousing (F(1,38) = 7.20, p < .05), interactive (F(1,38) = 8.06, p < .01), the test situation more involving (F(1,38) = 4.08, p < .05) and the game less boring that the Slicks (F(1,38) = 11.34, p < .01). In CRT condition NFS was only more arousing than the Slicks (F(1,38) = 4.36, p < .05).

4. Conclusion

This paper presents an explorative multidimensional method (Presence-involvement-flow –framework) to evaluate user experience in VEs and its application to measure and profile user experiences in four different experimental conditions. This method has been developed and tested in our previous studies [6, 17].

The results showed that four dimensions (Physical presence, Emotional involvement, Situational involvement and Performance competence) depicted and discriminated user experiences well in different experimental conditions.

The difference between the two games studied is obvious. Need for Speed is a fast paced three-dimensional 1st person racing game and Slicks is a simple two-dimensional 3rd person racing game that resembles simple electronic or mobile games. However, the experiential differences between the games are more complex.

Our results show the psychological profile of the driving game genre. They also depict the differences in this profile between two types of racing games within the genre. NFS evoked motivationally, cognitively as well as emotionally more intensive experience. The gameworld created by NFS seemed qualitatively more rich and positive and gave more pleasant and interesting experience. Also, the computer generated competitors and social scenes created an impression of a real social competition. All in all NFS meant more to user than simpler Slicks. Competence wise the two games did not differentiate, which can be explained by the fact that they both are easy to learn and the participants of the test were carefully selected trough background questionnaire and they were equally experienced rally game players.
4.1. The role of the presence in user experience

Presence is an important feature in measuring the human experience in VEs such as gameworlds. However, besides presence also other measures should be considered to reach the holistic human experience [16].

It is obvious that presence should be higher in 1st person games than in 3rd person games. But by simply measuring the presence we would not have recognized the effect of the VR –display to the gaming experience. As compared to the CRT –condition VR –condition caused the difference between the two games also in Situational involvement. This is an important dimension, because it measures perceived and evaluated environmental opportunities. As we are able to control these, we are able to increase, e.g., the quality and meaning of the experience.

The heightened Situational involvement in VR –condition can be partly explained by gender differences. But it is out of the scope of this paper to further analyze the effects of the background variables to these results.

4.2. Future

The developmental process of the PIFF continues. Although, its current structure profiles quite well the experience gained from the various VEs, the biggest challenge is to establish the structure and content of the framework. Also the background of the user should be studied more carefully to understand the experience a situation evokes. People see same situations in different ways and to increase knowledge of this evaluation process increases also knowledge of their PIFF –profiles.

Acknowledgements

This study was conducted in co-operation with Psychology of Virtual group and Nokia Research Center. It was partly funded by the Finnish Cultural Foundation.

References

Abstract

Web-enabled cellphone video recording devices, such as the Nokia 3650, provide a site for creating and broadcasting rhetorical discourses adapted to the time and place constraints they present. Our project explores the rhetorical constraints of time and place as they relate to the production of persuasive multimedia messages with video cellphones. We employ rhetorical strategies of enargeia to meet the time and place constraints posed by current video cellphone technology and its media of audio-visual and text-based communication. Among other presence-inducing rhetorical techniques, we employ audio-visual repetition, association, rhetorical questioning, and key dialectical terms to persuasively display the conflicts over conflict—the operative positions advanced in arguments for and against war.

Keywords—cellphone videos, persuasion, rhetorical presence, war.

1. Introduction

Rhetoric has always borne its interests to the site of particular cases—artfully employing its means of persuasion to influence the judgments and actions of intended auditors. Accordingly, rhetoric’s effective engagement is always negotiated anew as its aims are constrained by circumstances constituting what are called “rhetorical situations.” Rhetorical situations require phronesis or practical wisdom—the apt adaptation of communication to the complex of autonomous people, measured times, and cultural contexts. It is through the strategic adaptation and mediation of symbols that rhetoric coactively induces cooperation without coercion—that rhetoric achieves its persuasive aims in the “realm of the probable and contingent” at particular times and places.

Web-enabled cellphone video recording devices, such as the Nokia 3650, provide a site for creating and broadcasting rhetorical discourses adapted to the time and place constraints they present. Most video-enabled cellphones allow up to 10 minutes of recording and playback (in Europe and Asia video cell phones have greater functionality than in the United States because of early service provider decisions to standardize methods of data transfer thus allowing more effort to be directed toward functions within a particular phone model). The videos may immediately be sent over the internet as multimedia messages. In turn, the videos can be received and viewed by video-enabled cellphones, handhelds, or by web- and video-enabled desktop and laptop computing devices.

On cellphones, the videoclips are displayed on screens ranging from 1 inch X 1 inch to 1 inch X 2.2 inches. On handheld devices (e.g., PDA’s, Tablets) the videoclips are displayed on a slightly larger screen. On desktop and laptop computers, the media player’s viewing window can be enlarged with considerable loss of quality to the video images. The sound quality of the original footage is fairly good at 16 bits mono with an 8000 Hz sample rate. On a cellphone, sound is mediated through the speakerphone function and the volume is adjustable as it is on all computer-based media players.

2. Vidblinks and Rhetorical Presence

Our project explores the rhetorical constraints of time and place as they relate to the production of persuasive multimedia messages with video cellphones. Specifically, we focus on the production of multimedia messages that persuasively display attitudes toward war.

Rhetorical presence may constitute a fundamental rationale of a given discourse’s persuasive appeal. In the history of rhetoric, presence has been associated with visualization through stylistic devices of ekphrasis or enargeia. Enargeia is understood as a quality of vividness that strategically operates under the maxim that “seeing is believing” and the cultured assumption that vivid discourse can induce a sense of “being there” that provides the discourse’s creator and receiver with a sense of witnessing and experiencing the discourse’s claim. In his Orator’s Education (6.2), Quintilian refers to this quality of experience as phantasia—a sort of “functional” hallucination—prescribing that speakers mentally engage a visual representation of their subject at the time of a given speech’s utterance as an inducement to deliver the speech in an emotional tone commensurate with the speech’s intention. For example, mentally “seeing” a husband and wife stranded in the ocean with sharks encircling them during the course of speaking of their plight will induce an emotional tone in the speech that evidences the speaker’s emotional connection to the subject. Similarly, the vivid depiction and emotional speaking of the couple’s plight is designed to evoke an emotional tone in the listener.
commensurate with “being there” closing the experiential gap between simply “hearing” the report of their plight and eye-witnessing it.

The strategic manipulation of strategies of enargeia provides an orientation toward a given discourse that aligns its emphases with its intention to induce a given judgment. That is, any given “observation” of any given event is perspectival—it is derived from a standpoint that is usually underwritten by values or cultural and other screens that call attention to features of phenomena and, therefore, conceal, or put out of awareness, other features of the same phenomena. Accordingly, two people can witness the ‘same’ event and judge it differently under different hierarchies of interest that highlight and downplay aspects of the phenomenon under view.

As far as rhetorical narratives intend to constrain listeners’ judgments of given events, actions, artifacts, etc. in accord with the rhetor’s intention, strategies of enargeia project orders of emphasis that are designed to preclude conflicted ways of seeing among a group of observers. Where audiences are symbolically/rhetorically induced to cooperate in seeing similarly they are able to coordinate their judgments, arrive at collective decisions, and undertake conjoint actions with a perception of deliberate, democratically-formed consensus. This tactic of concealing and revealing is also metaphorically understood as framing and reframing, where perceptions are induced and actions motivated by linguistically or visually contextualizing or re-contextualizing phenomena under view. In visual media, verbal narrative may frame what is seen by providing linguistic prompts that pattern interpretations. For example, the same painting with different titles may induce different judgments of the painting’s significance, different interpretations of the artist’s motives, and different interpretations of the painting’s meaning. With regard to a literal frame, different textures, widths, colors and materials may also induce different experiences of the differently framed painting’s significance—even to the extent that some ways of framing are prescribed for some ways of painting. For example, expressionist painting is typically framed with gilded material. This operation may also be understood by the turn of phrase punning on the homophones “gilt” and “guilt”: “Gilt by association.” In this view a given phenomenon is framed, identified, or associated with some other phenomenon thereby blending their meanings. This is especially effective in visual rhetoric where, for example, people who are seen together may be considered related in some way—either by guilt or gilt where the character of the dominant part of the pair may “rub off” on the other. Of course, for better or worse, this strategy of physical positioning resonates with the cultural maxim “Birds of a feather flock together.”

Our project queries the rhetorical limits of web-enabled video cellphone media, under the constraints of time and place, for creating rhetorical presence and inducing judgments of war’s efficacy. Using Toshiba VM 4050 video cellphones and cellphone service provided through Sprint PCS Vision with average data transfer rates between 50 kbps-70 kbps (Sprint/Qualcom 3G CDMA 1x), we produced 43 “vidblinks.” A vidblink is the video counterpart of a sound bite—a brief, vivid, memorable, and rhetorically effective message. Operating under the time constraint afforded by the Toshiba VM 4050, each vidblink is 15-seconds long. The vidblinks display value-judgments of war’s efficacy. Some of the vidblinks engage war’s efficacy-in-general. Other vidblinks indirectly address current conflicts.

Each vidblink consists of a visual image and an oral narrative. They employ rhetorical strategies of enargeia to meet the time and place constraints posed by current video cellphone technology and its media of audio-visual and text-based communication. Among others, the vidblinks employ narrative repetition; visual and oral association, antitheses, metaphor and synecdoche; rhetorical questioning; and a variety of camera-movement techniques to persuasively display the conflicts over war—the operative positions advanced in arguments over war’s efficacy. These stylistic strategies enable the creation of rhetorical presence within a relatively brief time-period; within the size- and place-constraints of the cellphone’s screen; and the recording, editing, broadcasting, and playback capabilities of the device.

Most of the vidblinks draw their visual substance from objects “found” in the local environment—gasoline cans, flowerpots, baseball diamonds, playgrounds, chessboards, ceiling fans, and road signs. Using the readily observable environment as their visual scene and context, each vidblink’s narrative rhetorically frames its objects as relevant to contemplating war’s aspects in different ways. Additional presence is created through the combination of visual commonplaces and uncommon rhetorically framed narratives that bring to awareness “meanings” embedded in the everyday environment that may transform their commonplace observation into sites of contemplation as taken-for-granted surroundings are imbued with meanings induced by the vidblinks’ narratives. For example, one vidblink portrays a gasoline container sitting on a wrought-iron patio table. It is narratively framed by a rhetorical question that metaphorically associates gasoline with blood: “How much blood does it take to fill your tank?”

Given the current war in Iraq, and the relevance of blood and gasoline to its prosecution, an association is made between a commodity and a conflict that may bring to awareness an economic rationale for the war that calls into question pronouncements characterizing the war as an operation in service of liberation and democratic social reform. All this happens in a 15-second time line, that, in its brevity and intensity as a re-contextualization of the commonplace, rhetorically induces presence and the possibility of remembrance of the message’s associative
import the next time the observer purchases gasoline, and so long as the war continues. This may, in turn, stimulate a repetitive recognition of a disjunction between the claims of the government and the realities of the war’s interests: humanitarian vs. commercial.

In addition, one inducement to attend to some of the vidblinks relates to the already-established symbolic significance of given props. For example, one vidblink is set on a baseball diamond. The camera moves around the bases and each base is characterized by a term related to war’s tragic losses. As the camera crosses home plate, the narrator asks “Isn’t it time we came home?” In this vidblink, there are kinetic associations (movement around the bases) as well as associations to the game of baseball and a play on the word “home” as the camera crosses “home” plate. In sum, given baseball’s commonplace characterization as “America’s pastime”, the baseball diamond is a site of strong cultural-symbolic significance to many “Americans.” Its association with war will probably be especially memorable, and possibly alienating, to baseball fans. As a potential site of alienation, the vidblink may prompt anger, thereby raising consciousness of the observer’s taken-for-granted position.

Figure 2 Come Home

Other vidblink props with generally established symbolic significance include the American flag, cemeteries, and playgrounds. So, in the found environment there may be objects and sites symbolically charged with political and social significance as well as objects and sites that are taken for granted in their own right as commonplaces of everyday life.

It is difficult to determine which set of props are most apt to be effectively reframed in the interests of our project. In the case of commonplace objects and sites, the rhetorical challenge is addressed by constructing narratives sufficient to symbolically charge them with unexpected emotional, political and social significance appropriate to the contemplation of war’s efficacy—as in the case of the gasoline container. In the case of the objects and sites with established symbolic significance, the rhetorical challenge is addressed by narratively downplaying their culturally established associations thereby reframing them as appropriate vehicles for contemplating war’s efficacy. In addition, as in the case of commonplace objects and sites, the narrative prompts to their contemplation may be unexpected drawing on the rhetorical appeal of surprise as an attention-getting device. Finally, and perhaps most importantly, all of the props are potentially “found” in the everyday environment and may therefore prompt recollection of given vidblinks and induce the contemplation of war’s efficacy as they are met in the course of everyday life.

3. Assessment

The 15-second format enables and drives the compact and potentially intense engagement of an audio-visual message (1) displayed by the cellphone’s small-screen, (2) heard through the cellphone’s speakerphone function, (3) wirelessly broadcasted and received as a multimedia message, and (4) apt to be successively rebroadcast by first-wave recipients to other recipients.

The wireless cellphone video medium affords the potential for the exponential chaining of single multimedia messages toward all web-based nodes of reception, constituting a cyber-grapevine and genealogy of meaning that may surpass the audience-in-view of the message’s creators, constituting the message anew as it may be taken up and rebroadcast world wide.

In addition to creating the vidblinks, we are currently in the process of broadcasting them via the internet to over 100 participants in our project. We have created a survey to (1) assess their influence on the recipients’ impressions of the conflicts over conflict, and (2) ascertain the scope of their rebroadcast. This feature of our project will enable us to understand the messages’ reception—how relatively brief small-screen audio-visual cellphone messages may effectively perform a presence-inducing persuasive function and circulate in a global network.

This aspect of the project addresses the recurrent objections raised in critics’ judgments of “sound bites” and their potential to over-simplify and subvert time-honored discursive practices characterized as central to appropriately engaging ideals of rationality connected to the formation of appropriate judgments. We are interested in the place of vidblinks in the provision of “good reasons” backing moral judgment. That is, in the history of rhetoric, specific concrete vivid images and examples (nonverbal or verbal/fictional or actual) have played a significant role in the inducement of cooperation—rhetoric’s chief persuasive aim. Operating under the rhetorical rationale of liveliness or vividness, and the idea that seeing is believing, schemes and tropes—especially metaphors (visual and verbal)—have been deployed to set striking images serving persuasive functions within the three traditional genres of rhetoric: deliberative (i.e. generally political and future oriented discourse), epideictic (i.e. generally value-educative and present-oriented discourse), and forensic (i.e. generally legal and past-oriented discourse).

We want to address our assessment of the vidblinks’ effectiveness and distribution in light of our rhetorical interests in the visual and oral “arguments” of style and delivery—of symbol choice and media of communication—operating within the constraints of time and space—as the vidblinks may be freely distributed across the world-wide-web. At the same time, however, we wish to assess the vidblinks’ broadcast and their likely drift to unintended auditors—where their generic intention as persuasive displays will probably be aborted as they are planted in argument fields unreceptive to their telos—to
their actualization as fruitful insights into war’s efficacy to be cultivated and harvested by their observers.

4. Conclusion

In sum, our project seeks to understand how the time- and place-compressed communication medium of cellphone video can be rhetorically managed to produce persuasive messages with political and social import potentially communicable throughout the world-wide web. That is, rhetoric has always been associated with persuasion, as far as it may vividly display—give presence to—the grounds of judgment in any given particular case. Our vidblinks follow in this ancient tradition, adapting it to new communication technology. In a way, our strategy of associating commonplaces of the environment with judgments of the efficacy of war potentially reverses the order of technologically mediated presence by imbuing the found environment with meanings that are derived from their mediated deployment as invitational props to narratives considering war’s efficacy. This reversal opens the prospect of the possibility of contemplating war’s efficacy in any landscape that comes before one’s view. It may induce a hermeneutic of everyday observation with ancient roots, stemming back to the belief in a divine logos permeating all phenomena with messages of spiritual import emanating from an omnipotent, omnipresent creator. In the case of our project, the vidblinks provide a pretext for contemplating their intention at the intersections of everyday life where their props are encountered and possibly recognized as sites of significance, bringing to mind the vidblinks’ brief narratives and possibly inducing the daily contemplation of war’s efficacy as one moves through environments containing the vidblinks’ visual props. For example, when one of the authors filled his lawnmower’s gas tank from a red plastic container, he was reminded of the “Blood and Gas” vidblink’s rhetorical question: “How much blood does it take to fill your tank?”

The interplay of mediated and unmediated presence may prospect toward inculcating a “way of seeing” that is mindful of the potential insights that can be gained from taking note of the world-as-given through a consciously affected interpretive frame—inducing a quality of experience that can be gauged by the quality of life it provides its bearers.\(^1\)

Acknowledgements

The authors gratefully acknowledge the support of The Center for Educational Technology, Middlebury, Vermont; Vassar College’s 2004 Summer Media Institute, its director and organizer, and the technical support staff located at Vassar College’s Media Cloister.

References


1 Our poster presentation at the International Workshop on Presence 2004 will consist of a screening of selected vidblinks and their discussion. In addition, we will discuss the concept of rhetorical presence, its application to the production of the vidblinks, and the technical and communicative prospects and limitations of the video cellphone medium for the production and dissemination of persuasive messages. The results of our survey will be forthcoming in a subsequent publication.
Do 3-D movies really reach out and grab you? The case of SpyKids 3-D

Cheryl Campanella Bracken, Matthew Lombard, Kimberly A. Neuendorf, James Denny & Michael Quillan
Cleveland State University, Temple University, Cleveland State University, Cleveland State University
{c.bracken@csuohio.edu, lombard@temple.edu, k.neuendorf@csuohio.edu, jtldenny@msn.com, bigtreesq@msn.com }

Abstract
This study examined the role of 3-D images in increasing the sense of presence experienced by movie goers. In an experiment manipulating the dimensionality seen by viewers (either 2-D or 3-D), the sense of presence experienced was measured in a post test questionnaire. The results demonstrate that at least for the particular film used (SpyKids 3D), the glasses necessary for viewing detracted from presence experiences.

1. Introduction
3-D movies are becoming increasing popular and financially successful [1, 2] for a variety of cinema formats including traditional film theaters, IMAX, and now home viewing. The selling point for nearly all of these formats is that the viewer will have a more immersive experience because they will feel more a part of the action. The web site for the film SpyKids 3D (www.spykids2.com) explains that 3D allows audience members to "not feel they are in the theater, you'll feel as if you are actually part of the action." In order for the viewers to experience a sense of presence, they must wear traditional plastic and paper 3-D glasses. This study was initiated by a curiosity about whether and to what extent 3-D films can and do increase audience members' presence sensations.

1.1 Presence
There has been an ongoing discussion about the nature of presence. Lee recently asserted that presence research needs to include all types of interaction (i.e., both mediated and non-mediated) [3], while David [4] argued that in order to better define the concept researchers need to limit the definition of presence to include only mediated experiences. Freeman also recently proposed limiting presence research and put forth the following definition: “A perceptual illusion of non-mediation yielding a subjective sensation of being there in a mediated environment” [5]. Freeman also states that across a variety of presence measures, three dimensions repeatedly appear. They are (1) sense of being physically located in environment (2) engagement/involvement/attention and (3) naturalness/reality/consistency with real world [5].

The authors agree with both David and Freeman in limiting the examination of presence to a mediated setting. However, we prefer a broader definition of presence not limited to mediated 'environments' and therefore for this study use the definition developed on the presence-l listserv [6]: “Presence (a shortened version of the term "telepresence") is a psychological state or subjective perception in which even though part or all of an individual's current experience is generated by and/or filtered through human-made technology, part or all of the individual's perception fails to accurately acknowledge the role of the technology in the experience” which is what Lombard and Ditton (1997) termed "an illusion of non-mediation." This definition is well suited to a wide range of media including virtual reality, computers, and the focus here, film and television.

Lombard and Ditton (1997) note six dimensions of presence in the literature and this study focuses on the following four: (1) presence as immersion, which refers to the extent of perceptual and psychological immersion in a mediated environment; (2) spatial presence, or presence as transportation such that the viewer is transported into the mediated environment (“You are There”), the mediated environment seems to intrude into the viewer’s environment (“It is Here”) or two users meet in a shared mediated space (“We are Together”); (3) social presence, the perception of access to and connection with people in a mediated experience, and (4) presence as realism, or the extent to which a medium can produce seemingly accurate representations of objects, events, and people. This study includes these three dimensions in its investigation of presence invoked in viewers when watching a 3-D movie, SpyKids 3D: Game Over.

1.2 Presence and 3-D images
While there are numerous studies on the impact of 3-D images with computers and virtual environments, only a few have explored the relationship between presence and 3-D television or film.

In an attempt to provide evidence for the claim that more sensory inputs will lead to stronger or more presence sensations [see 9], IJsselsteijn, de Ridder, Hamber, Bouwhuis, and Freeman [10] conducted an experiment to investigate the effect of viewing 3-D images on sensations of presence. They argued and found support for the idea that stereoscopic displays present the audience with more information and therefore should encourage sensations of presence.
Freeman and his associates have conducted several studies manipulating stereoscopic images. The overall results demonstrate that audiences report more and stronger presence responses when viewing content in the stereoscopic format. [11, 12]

1.3 Hypotheses

The current study attempts to build on the previous research findings by conducting a field experiment, where the participants view a film in a cinema-like setting. Additionally, unlike most other studies, the content seen is an actual Hollywood film which was a box-office success. Thus, based on previous finding we posit the following:

H1a: Participants who view the film clip in 3-D will report a stronger sense of immersion than participants who view the clip in 2-D.

H1b: Participants who view the film clip in 3-D will report a stronger sense of spatial presence than participants who view the clip in 2-D.

H1c: Participants who view the film clip in 3-D will report a stronger sense of social presence than participants who view the clip in 2-D.

H1d: Participants who view the film clip in 3-D will report a stronger sense of realism than participants who view the clip in 2-D.

2. Method

2.1 Participants

One-hundred forty-three students at an American Midwestern University (80 men, 77 women, average age 22.3, age range 18-55) participated in the experiment for which they received extra credit.

2.2 Design and Procedures

In a between participant design experiment, participants were randomly assigned to watch a 12 minute movie clip in either 3-D or 2-D. Viewing took place in sessions for each group in a large comfortable screening room/classroom, after which participants completed a short questionnaire.

2.3. Stimuli

Spykids 3-D: Game Over (2003, Dimension Films, rated PG) is the second sequel to Spykids and stars Daryl Sabara, Alexa Vega, Sylvester Stallone, Antonio Banderas, and George Clooney in a children's action/adventure story about the continuing adventures of spy siblings Juni and Carmen Cortez. The film was chosen both because of its box-office success and because the DVD version offered both 2-D and 3-D versions. The clip used was 12-minutes in duration and featured the identical race sequence in both versions. The only difference between the two conditions was that in the 3-D version the stimulus contained stereoscopic images and the participants wore traditional paper and plastic (one red lens and one blue lens) stereoscopic 3-D glasses.

The segment was selected because it contained point-of-view movement (POV). The scene included a “car” race in a video game environment, where the characters are transported to catch the villain “the toymaker”. The scene features the characters driving, crashing, and falling throughout the clip. In the 3-D version, the debris from the crashing vehicles seems to fly toward the audience. The clip ends at the conclusion of the race.

2.4. Measurement

The dependent measures included subjective measures of presence based on Lombard and Ditton's Presence Scale [15].

2.4.1 Immersion

Participants responded on a 7-point Likert scale (from very strongly disagree (1) to very strongly agree (7)) to six questions adapted from Lombard and Ditton (2001) to measure the extent to which television viewers feel a sense of involvement or engagement when they watch television programs and movies. The six questions include: How involving was the movie?, how exciting was the movie?, to what extent did you feel mentally immersed?, and how completely were your senses engaged? Cronbach’s alpha for the additive index was .74.

2.4.2 Spatial Presence

Participants responded on a 7-point Likert scale to four questions designed by Lombard and Ditton (2001) to measure the extent to which viewers experienced a sense of sharing a physical space. The four questions were: How much did it seem as if you and people you saw/heard were together in the same place?, how much did it seem as if you and people you saw/heard both left the places you were and went to a new place?, how much did it seem as if you could reach out and touch the objects or people you saw/heard?, and how often when an object seemed to be headed toward you did you want to move out of its way? Cronbach’s alpha for the additive index was .72.

2.4.3 Social Presence

2.4.3.1 Social presence: Active interpersonal

Participants responded on a scale from never (1) to always (7) to three questions designed by Lombard and Ditton (2001) to measure how much participants reacted to characters they saw/heard in the media environment. The three statements were: How often did you make a sound out loud in response to someone you saw/heard in the media environment?, how often did you smile in response to someone you saw/heard?, and how often did you want to or did you speak to a person you saw/heard? Cronbach’s alpha for the additive index was .84.
2.4.3.2 Social presence: Passive interpersonal

Participants responded on a scale from not well (1) to very well (7) to questions designed by Lombard and Ditton (2001) to measure the extent to which viewers could sense characters’ behaviors on television. The four questions were: During the media experience how well were you able to observe the facial expressions / changes in the tone of voice / the style of dress / the body language of the people you saw/heard? Cronbach’s alpha for the additive index was .78.

2.4.4 Realism

2.4.4.1 Realism: Social realism

Participants responded using a 7-point Likert scale (from very strongly disagree (1) to very strongly agree (7)) to three questions designed by Lombard and Ditton (2001) to measure the extent to which media users feel that presentations or experiences are an accurate portrayal of reality. The three statements were: The events I saw/heard would occur in the real world, the events I saw/heard could occur in the real world, and the way the events I saw/heard occurred is a lot like the way they occur in the real world. Cronbach’s alpha for the additive index was .87.

2.4.4.2 Realism: Perceptual realism

Participants responded on a scale from not at all (1) to very much (7) to questions designed by Lombard and Ditton (2001) to measure the extent to which television viewers feel a sense of realism when viewing television. The five questions were: How much did the things/people in the environment you saw/heard sound / look / smell / feel like they would if you experienced them directly, and how much did the heat/coolness (temperature) of the environment you saw/heard feel like they would if you experienced them directly. Cronbach’s alpha for the additive index was .85.

3. Analysis and Results

Each of the hypotheses were tested using independent samples t-tests with dimensionality (stereoscopic (3-D) versus non-stereoscopic (2-D)) as the independent variable and the various presence measures as the dependent variables.

The result for H1a predicting that participants who watched the 3-D version would experience a stronger sensation of presence was not supported. In fact, there was a significant relationship in the opposite direction with (t = -1.09, p = .05) with participants who saw the 2-D version reporting more immersion (M = 4.97; SD = 1.57) than those who viewed the 3-D version (M = 4.45; SD = 1.86).

The second hypothesis (H1b) was not supported; again the results were significant in the opposite direction (t = -2.49, p = .01), with participants who saw the 2-D version reporting more spatial presence (M = 3.45; SD = 1.41) than those who viewed the 3-D version (M = 2.93; SD = 1.05).

The hypothesis predicting more social presence (H1c) was not supported. Neither measure of social presence (active interpersonal and passive interpersonal) yielded a significant difference.

The hypothesis predicting participants will have a stronger sense of realism in the 3-D condition (H1d) was not supported. However, while there was no significant difference for social realism, there was some evidence that the participants who viewed the 2-D version of the film felt a stronger sense of perceptual presence (M = 2.49; SD = 1.29) than those who viewed the 3-D version (M = 2.15; SD = 1.09) (t = -1.08, p = .09).

3.1 Additional Analysis

Additional analyses were conducted to investigate the role of the 3-D glasses on viewer perceptions. A series of ANCOVAs (analyses of covariance) were conducted with each of the dependent variables and the independent variable (3-D vs. 2-D). The covariate was the question (to which participants responded on a 7 point Likert scale from very strongly disagree (1) to very strongly agree (7)): “I paid more attention to the glasses than to the movie.”

The results were that for each of the significant dependent variables that had yielded statistical differences favoring the 2-D version of the film (i.e., immersion, spatial presence and perceptual realism), the differences when controlling for attention paid to glasses were no longer significant, thus providing evidence that the glasses provided a moderating effect on presence perceptions.

[ FURTHER ANALYSES USING ADDITIONAL QUESTIONNAIRE ITEMS TO ISOLATE THE REASON FOR THE REVERSE FINDINGS WILL BE CONDUCTED PRIOR TO THE CONFERENCE! ]

4. Discussion

These results present an interesting conundrum for presence researchers. Previous research and much supposition point to viewing 3-D movies or television leading to greater levels of presence. However, the current results do not support this. All the hypotheses were not supported or were significant in the opposite direction with the 2-D version eliciting greater levels of immersion, spatial presence and perceptual realism. This suggests that perhaps more sensatory stimulation is not always necessary (or the best choice) to achieve a sensation of presence in media users, contrary to previous findings [14, 15].

The ANCOVA results suggest that the use of the paper 3-D glasses may have proved more distracting than the improved sense of “being there” the 3-D version of the film segment provided. The glasses (which are the same as the glasses used when the movie was shown in theaters) seemed to distract the audience from the film, which would logically interfere with their presence sensations.

This suggests that the “hardware” necessary for 3-D experiences needs to be able to be used with little effort on the part of the media user and needs to be less intrusive than the current commercial version available. However, the financial success of this and other promotions for 3-D
movies implies that audiences believe the claims of an enhanced experience.

In future studies it would be interesting to manipulate the 3-D device that is used to produce the 3-D effect for viewers. The results of this study suggest that when the apparatus is flimsy, uncomfortable, moves around too easily, etc., the media users may be distracted from the media experience; it is also possible that the perceived lack of quality of the apparatus (glasses) negatively affects their opinion of the experience itself.

These results might be atypical because of the use of college students when the target audience for the film was considerably younger. Prior observations lead the researchers to believe that young children do perceive the 3-D version to be more “real”. However, it’s not clear that young children could answer the questionnaire items in a reliable and valid manner. Future research should address the relative suitability of the content viewed for the audience members/study participants.

Conclusions

This study attempted to replicate previous findings regarding presence and stereoscopic displays using commercial film segments. The results do not support the earlier findings and suggest that the device used to create 3-D effects can impact upon media users’ experiences of presence.

References

Creative presence: Supporting artistic collaborations

Brigid Costello(1), Alastair Weakley(2), Ernest Edmonds(2)

1: University of New South Wales, Sydney, Australia
2: Creativity & Cognition Studios, University of Technology, Sydney, Australia
{bm.costello@unsw.edu.au, alastair@weakley.org.uk, ernest@ernestedmonds.com}

Abstract
Our discussion focuses on the design of creativity support systems. We describe a dedicated support system and the results of its evaluation. We then analyze a creative collaboration that was carried out without such a dedicated support system. We propose a new goal for designers of creativity support systems: facilitating creative presence. In addition, we describe four key aspects that such support systems should address: speed, status, synthesis and sketching.

1. Introduction
This paper is concerned with collaboration between remote participants on creative projects. We present first a creativity support system that has been developed and evaluated. Secondly we describe and analyze a remote creative collaboration that took place during the development of an artwork. In the first instance the tool provided a constrained environment for a test collaboration. In the second instance the paper describes an actual project that took place without a dedicated support system. While there are different insights to be gained from each instance we found that the analysis of both methods highlighted the need for creativity support systems to capture a sense of the presence of remote collaborators as well as the creative process. We go on to discuss the implications of this for designers of creativity support systems.

In the first part of the discussion we describe WISA (the Web-Interactive Scrapbook Application) [1]. This system aims to support collaboration between loosely-formed groups of people engaged in design work. Design problems have been described as 'wicked' and it is commonly the case that a single designer does not have the expertise to solve such problems single-handed [2]. For this reason, designers often need to collaborate with one another as well as with experts in other domains. Design is not a straightforward problem-solving process: at the start there is seldom a well-defined problem to be solved. In fact, the process commonly involves both the definition of the problem to be solved and the generation of a solution to that problem [3].

Supporting the definition and solution of design problems was the main aim of the WISA system.

The second part of the paper describes a remote collaboration that took place during the development of an artwork. In most cases, the development of an artwork occurs outside the commercial sphere. Although there may be some money made from an artwork or some support for production gained via government grants, many artists fund their artistic practice out of their own pocket. A concern to keep costs low often, therefore, reduces the technological options available to or considered by artists. The small scale of many artistic collaborations also influences the technological options that are chosen. With no large teams of people to manage, no clients, no large budgets and no time-based costs to calculate, many features of commercially available project management software are irrelevant. The main aim of the analysis of this collaboration was, therefore, to ascertain what kind of technological support is relevant to remote artistic collaborations.

Both these studies focused on creativity support but from different perspectives and not surprisingly they each highlighted different issues for designers of creativity support systems. Whereas the second study foregrounded creative process, the first brought home the importance of collaborative presence. We describe the combination of these two concepts as creative presence and argue that designing for creative presence is an essential consideration in the design of support systems for artistic collaboration. Furthermore, we suggest that designers must not only consider ways of capturing creative presence but should also ensure that creative presence can be extracted from the system. Finally, based on these ideas, we suggest four important capabilities that future systems might embody.

2. Evaluation of a Creativity Support System:

In this part of the paper we describe WISA, a system that has been developed to support remote collaboration amongst people engaged in creative projects. The rationale behind the development of this system is discussed and we review some findings from an evaluation of it. In this evaluation, the system was found to support the sharing of tacit knowledge between its users. Significantly, the
evaluation also highlighted areas where systems like this could be improved in order to provide a greater and more direct sense of connection among their users.

WISA supports the collection and organization of a body of research material which designers might build up as they explore the problems they are faced with. Additionally, by making these collections available to other users, WISA supports its user group in sharing knowledge about one another's skills and experiences. Jointly-owned collections can be made by collaborating users.

2.1 Practical Description:

The WISA system can be used to make, organize and share collections of useful material, each item in the collection in this case is called a resource. The resources may be pictures, text, sounds, references to people and so on. Like cuttings pasted into a conventional paper scrapbook, the resource objects can be annotated and arranged or grouped in two-dimensional space on the screen. Users may create any number of projects each of which corresponds to a page in a paper scrapbook. The WISA system exceeds the capabilities of conventional scrapbooks in that it allows users to browse and search within one another's collections. Searches can be sensitive to the groupings of the resource objects: when a search is carried out, it is possible for the results to include not only objects that match the search query directly but also those that have been placed near to ones that match.

A further aspect that differentiates the system from a conventional scrapbook and one that is also important in the task domain of design is that it records the history of the users' actions and allows this to be reviewed. There are three main reasons for recording and reviewing the design process in the conventional face to face environment "(a) to support reasoning processes in design, (b) to facilitate communication among the various players in the design process... and (c) to further the cumulation and development of design knowledge across design projects and products" [4]. All of these are valid reasons for including this facility in WISA, but in this case there is an additional justification. The WISA system is largely concerned with interaction that takes place asynchronously and in a sense indirectly: the current version of WISA does not provide facilities for direct communication between its users. Instead, users may learn about one another's skills and interests by studying the collections they have made and how the objects within them have been arranged. Reviewing the development of the collection might give an indication as to the thought processes of the user. The asynchronous nature of WISA does not necessarily mean that it cannot support the feeling of connectedness [5].
Inferences about what they were doing and the reasons behind their actions can come from studying the traces they left behind, such as their writings, art, and so on. With WISA one can study others’ past actions and past work and from this make inferences about what they were doing and the reasons behind their actions.

### 2.2 Implementation of the System:

WISA is a Web-based system implemented using PHP, MySQL and Javascript; it can be accessed using a normal Web browser. Collections are grouped into projects which correspond to pages in a conventional scrapbook. Each project may only be modified by its owner (or owners), but all users can see the contents of any project. Individual resources in a project appear on the screen in the form of tiles. These tiles can be arranged and grouped as desired by dragging and dropping (figure 1). When the history of a project is replayed, the tiles move around the screen reenacting the project until the current arrangement is reached.

### 2.3 Evaluation of WISA:

In order to evaluate WISA, small groups of people were asked to use it to carry out a collaborative planning task [1]. The system does not provide facilities for direct communication between its users and this, not surprisingly, was found to be a problem. Additionally, on returning to a project, it was hard for the user to see what, if anything, was found to be a problem. With WISA one can study others’ past actions and past work and from this make inferences about what they were doing and the reasons behind their actions. Communication between its users and this, not surprisingly, was found to be a problem. Additionally, on returning to a project, it was hard for the user to see what, if anything, had been changed. However, there were instances where the users discovered new common interests: where they came to know one another better through using this system. While one might say that there was an awareness of the other users this was indirect: a knowledge about rather than of the other person- they were not directly present to one another.

In the WISA system, the user makes a collection for their own purpose. Incidentally, this collection may be of use to other users of the system: the owner need not actively be involved in any sharing process. Instead, by actively carrying out searches, other users pull information from the system. A number of the users’ comments lead us to the hypothesis that users would feel a greater sense of connection were the system also to actively push information to them. For example, they should be alerted to changes that have been made since their last visit to a project.

In face to face collaboration, communication is not limited to deliberate speech acts. Instead, many interactions may be stimulated (or inhibited) by something we overhear, by something we notice on a colleague's desk or by observation of a facial expression [7]. Traditional remote collaboration systems such as the telephone, or videoconferencing systems tend not to succeed in transmitting all of this information. With the telephone one must actively make a connection by dialling the remote person, and much work continues to be done in improving videoconferencing systems by incorporating for example devices to allow the user to actively point at objects at the remote location [8]. WISA takes an alternative approach: project owners do not actively push information at one another- instead users casually observe other's projects: the equivalent of looking at the books on their shelf or their facial expression.

While clearly the WISA system is not designed to cater for all the needs of people collaborating remotely, studying the use of such a constrained system has led to insights about the facilities that should be included in a more extensive communications facility. We found, for example, that users returning to a project that had been modified by a collaborator found it hard to understand what had been changed and why. While the history replay facility was interesting, as one evaluator described it, it did not provide sufficient flexibility in the way that the project could be reviewed. Perhaps more importantly, there was not a sufficiently clear indication of the other user's past presence in a project. Like footprints in the sand, the history feature was intended to give a clue not just about where people had been in the project but also about what they did while they were there and why they did it. It is clear that this aspect should be emphasized in future systems. In the next section of the paper we describe a collaboration between two artists as they worked on a project from remote locations. Here again we find that the traces left by the collaborators to be a useful resource for future work.

### 3. Analysis of a Creative Collaboration:

In this section we describe the process of developing and producing a new media artwork via a remote creative collaboration between Brigid Costello and a fellow artist. The problems and benefits of collaborating via technologically mediated communication are discussed. An analysis of this collaborative experience highlights the need for a system to capture creative presence and suggests a development direction for the history feature of the WISA system.

The following analysis is based on a creative collaboration that occurred between December 2003 and June 2004. The collaboration involved two artists who worked together to propose and produce a new media artwork. At the beginning of the process both artists were living in different Australian cities. Just before the main production phase of the project one of the artists relocated to Japan. Due to this geographical separation the collaboration process primarily occurred via technologically-mediated communication: in this case email. This was augmented only by a few telephone calls, two face-to-face meetings and two chat sessions. Our analysis is based on a log of the artists’ email communications and on discussions with each of the artists.

#### 3.1 Managing Information:

One of the first comments made by each artist when reflecting on the collaborative process was that they needed better methods for managing and collecting the data they were exchanging. During previous remote collaborations with each other the artists had relied much more heavily on...
telephone conversations. The increased geographic separation in this instance made telephone calls too expensive and none were made after one of the artists relocated to Japan. Compared to previous collaborations there was, therefore, a much greater volume of email communication. During the collaboration the artists exchanged nearly 200 emails. Sixty percent of these were exchanged during the peak production period between the 20th May and the 2nd of June. It was as this point that one of the artists considered making a bulletin board style log of the emails as it was becoming difficult to keep up with the creative decisions that were being made. The log was never made as it was judged too time-consuming a task at that late stage of the project.  

Another factor that made data management more difficult was that each artist, for various reasons, used several different computers and different software packages to collect their email. Messages and files were downloaded on one computer and then not available when work was being done using another computer. One of the artists had a 20 MB limit on their email account and this sometimes meant that emails were deleted in order to free up space for further communication. There were several emails asking for files to be resent because one or other of the artists could not find them.

3.2 Communication Issues:

The collaboration also suffered from the common communication problems associated with asynchronous non face-to-face communication [9]. For example, there were some cases where ideas were misunderstood due to a lack of detail in email text. Both artists said that it took them longer to write an email expressing an idea than it would to speak the idea over the phone. In the interests of saving time email communications were often abbreviated and thereby more likely to be misinterpreted. Given the visual nature of the work the artists were producing it was not surprising that often these misunderstandings were cleared up with sketches or diagrams. When asked why they did not exchange such sketches more often, time was again a major factor. The process of opening a drawing program, saving the file in the right format and then attaching the file to an email was felt to take too long.

The collaboration did not encounter any major communication problems due to different time zones. Although Japan and Australia are geographically quite distant there is only a one-hour time difference between them. What did cause some issues though was the difference in times of day that each artist did their work [10]. While one artist was working on the project full-time and tended to work between 8am and 6pm on weekdays the other artist was fitting the project in around other paid work and tended to work between 5pm and 1am and on weekends. In the busiest period of the production project there are occurrences where one artist sends 3-5 emails asking questions and does not receive any replies for several hours and in one case for a whole day. These delays in responses again provide opportunities for miscommunication by making it hard for the artists to keep track of the context of their dialogue. Interestingly it was observed that the artists often accommodated for this by making much more use of email quotations after such large time delays. Conversely, when both artists are at their desks at the same time and emails are being exchanged quickly there is very little use of quotation.

It was also evident during the early stages of the project that some communication problems were occurring because at that stage neither artist was working on it fulltime. There were two periods where there were two-week gaps between emails and it was apparent that the artists needed to clarify the creative decisions that they had already made because they had forgotten what had been decided. This lack of clarity about creative decisions was not just occurring because the artists were working on other projects and were not fully focused on the collaborative project. It was also caused by the artists’ reliance at this stage of the project on undocumented telephone calls and face-to-face conversations. The artists both saw the recording of their creative decisions as one of the clear benefits of collaborating via a text-based communication tool like email. They were particularly surprised to see the great number of ideas that were suggested but not used during the brainstorming period of the project and felt that being able to review this data could be useful for future projects.

Where email was judged to perform poorly was during times when very major creative decisions had to be made. The asynchronous nature of email was seen as a hindrance to the debate involved in making creative decisions. The artists felt that misconceptions took too long to identify and clear up and that it was much harder to ‘spark’ off each other. While the artists were both happy to brainstorm the initial idea via email, the major discussions and decisions about the first shape of the project were made via phone calls or face-to-face meetings. In the middle stage of the project when it became clear that it was necessary to make some big changes to the project definition the artists used synchronous chat to debate those changes. After finding the email records so useful there were some regrets that these chat exchanges were not recorded.

4 Capturing Creative Presence:

Even from this brief analysis it is apparent that such a small-scale remote collaboration would benefit from a coordinated approach to the management of communicative data. Such data needs to be collected together in one place that is specific to the project and accessible from multiple computers and locations. It is also evident that a software tool to support remote artistic collaborations would need to include a bulletin board, loggable voice and/or text chat, a loggable sketch tool, an editable project specification document and a file management tool. While these observations are useful they do not particularly cover any ground not already encompassed by existing project management software. What it is most important to realise about the combination of these tools, however, is the shift in focus away from economic concerns of budgets etc and towards an almost complete focus on supporting and capturing the creative decision-making process.

290
This shift in focus from economics to creative process entails an accompanying shift in design goals for the creator of such a support tool. The designer needs to aim not only to enhance but also to successfully capture both the presence of each collaborator and the 'sparks' that are occurring between them. We use the term 'creative presence' to describe this combination of collaborative presence and creative process. While it was remarkable how much creative presence was in the email communications that we analysed, it was also evident that this sense of presence could be enhanced through careful design of a support tool. When looked at from this perspective it is also apparent that many of the issues described in the previous sections arose through deficiencies in capturing or enhancing creative presence. For example, difficulties in digesting and tracking email communications were caused by email software’s division of email into ‘in’ and ‘out’ boxes which in effect destroyed the conversational presence of the exchanges. We would suggest that a creative process support tool with enhanced creative presence would not only improve the process of remote artistic collaborations but given the wealth of ideas discussed and discarded during such a process would also provide an important data resource for future projects.

4.1 Extracting Creative Presence:

In both the WISA system and the creative collaboration described above, the asynchronous nature of the communication led to the generation of material that could be referred to later. While this is a largely inevitable side-effect of the use of many asynchronous communication technologies such as email, in WISA this recording was a feature of the system and one that was found to be interesting by the test participants. Similarly, the collaborating artists remarked that on returning to their notes they rediscovered a number of interesting ideas that might, had their discussions been largely by telephone for example, have been lost. In the WISA system, the playback facility has only limited functionality: it is not possible to play the project in reverse or to skip forwards at high speed for example. Also, while there is an indication of who moved an object when a WISA project is replayed, this does not seem to go far enough. Users of WISA, for example wanted to be alerted to the recent presence and actions of one another. It seems that it is desirable to create an awareness of the historic presence of another. This ability to come to understand the creative intent by the trace of the project as it developed, not just by looking at the end result is something that should be enhanced in future systems. We have described above instances where this has been captured: it is on the recovery of and access to this information that future work should concentrate. Access to the raw data is not such a problem as is access to the creative intention: the reason behind an action. Thinking of WISA, one might study, for example, the development of a particular grouping of objects. Was the grouping created in one go or were there many tentative shufflings and reshufflings as the user explored some concept? It is just this sort of, seemingly inconsequential, information that is often lost when system design focuses on raw data. And it is just this sort of information that can be of great use in creative collaborations. The ability to recover these subtle clues about intentions and creative practice can greatly enhance both the process of a collaboration and its outcome.

4.2 Designing for Creative Presence:

In this section we propose some ways that our concept of creative presence might influence the implementation of a creativity support system. However, before we can discuss designing for creative presence it is important to clarify what exactly is encompassed by this term. We defined creative presence earlier as a combination of two concepts, collaborative presence and creative process. The first concept, collaborative presence, relates to the representation of what each individual collaborator does, says or thinks. This definition of collaborative presence also importantly encompasses the representation of any dialog between these individual presences. Designing for collaborative presence, therefore, relates to designing for vitality, connectedness, social presence and awareness. The second concept, creative process, is concerned more with being able to 'read' than 'write' presence. Creative process relates to representing the intentions behind and meaning of communications. It also relates to enabling through the system design the human abilities that Boden identifies as an essential part of the creative process, the abilities of "Noticing, remembering, seeing, speaking, hearing, understanding language and recognizing analogies" [11]. A final important ability that a system design needs to enhance is the human ability to "redescribe" and "transform" their ideas and processes [11]. Designing for creative process, therefore, relates to designing for searchability, threading, sorting, archiving and historical records. Additionally a designer needs to ensure that their system design provides opportunities for collaborators to discover new ideas rather than merely to explore a predefined conceptual space and to this end it should allow for multiple and fluid representations and transformations. So how might designing for creative presence influence the design of the tools within a creativity support system for artists? A complete discussion of this question would take far more room than this paper allows. However, there are four preliminary areas that we would like to discuss: status, synthesis, sketching and speed.

In creative work, particularly when it is synchronous, immediacy must extend beyond deliberate communication to encompass passive notification of status. In face-to-face working one can easily see if a collaborator is busy or tired for example but in remote working this is often not possible and it can be difficult to behave sensitively to the remote collaborator's situation. It is in precisely these situations that it is helpful if the system can push information rather than requiring the user to actively search for and retrieve it. This would also help the user who has been interrupted and who may not be in a position to make a deliberate notification of this fact. At other times, of course, deliberate notification may be desirable. We have, then, a
Sketching is an important part of the creative process. Because it enhances awareness, the design of a status feature relates primarily to designing for collaborative presence. It does also, however, partly relate to designing for creative process because it improves the ability of collaborators to ‘hear’ each other and helps with the reading of intentions.

The ability to spot new connections, to make new syntheses of concepts and ideas is only partly supported by conventional threading systems. Data, such as the artists’ emails described above in truth contain many simultaneous threads. Moreover, threads may have different significance at different times and to different people. Dynamic highlighting and tracing of common features in data that may be in multiple formats would, therefore, be highly desirable. It would also be valuable to provide multiple methods for synthesizing these common features and multiple ways of representing each synthesis. Here the design focus is obviously more on creative process. It is also possible, however, that this feature may create new opportunities for the connectedness we associate with collaborative presence.

Sketching is an important part of the creative process because it allows commitment to be deferred [12]. Where sketching is not appropriate, such as when an idea is to be communicated in writing, this ability to avoid commitment, to avoid the idea appearing to be formalized and complete, may be lost. A sketch may appear tentative, and may express many possibilities simultaneously but this can be hard to do in writing for example, particularly when it is typed. A facility for adding sketchy, explanatory notes, perhaps by voice recording as well as by visual sketching would help to solve this problem. The vitality that a sketching feature would bring could enhance collaborative presence. This feature would also potentially improve the creative process by facilitating the transformation and re-description processes described above.

Speed is a factor that influences not just what purpose a tool is used for but whether a tool is used at all. It should not be forgotten that what we are concerned with here is the design of a support system. The artists clearly regarded their work on their project as their highest priority and this influenced their choice of tools. Finding the use of a dedicated drawing application too slow for their purposes they resorted to plain text. Having the right tool to hand at the critical moment can allow one to capture and explore a creative insight before creative flow [13] is lost. It is also evident that there are times when different speeds of communication are necessary. A chat application whether voice or text based is clearly useful for high-speed exchanges of ideas. At other times, though, it may be regarded as causing communications to take up too much time or as taking focus away from the task at hand. In the same way that people are able to use text messaging and phone conversations for different purposes, collaborators need to have a suite of communication tools with different levels of immediacy. There is clearly a requirement for some sort of balance between the collaborative presence afforded by high-speed communications and the creative process which may sometimes benefit from a more reflective approach. Of course, this balance may need to be shifted as work progresses.

We propose, then, a number of preliminary features as desirable in a creativity support system; a suite of communication tools with varying levels of immediacy, a tool to push and pull status information, a tool for synthesizing data across multiple formats and dynamically representing this synthesis and a multimodal sketch facility. Above all such a system must be readily available and easily accessible in timely fashion.

5 Conclusion:

We have described both a system that has been developed to support creative collaboration and an artistic collaboration that was carried out without a dedicated support system. In both cases the lack of communications facilities that would normally be available to collaborators working together face-to-face led to insights about the nature of what we have termed creative presence and highlighted the value of designing a system that focuses on both the writing and reading of creative presence. We plan to continue this research by developing these ideas into a creativity support system for small and large scale remote artistic collaborations. There is potential for this to have other wide-ranging applications, for example, as a support tool for commercial designers, for creativity researchers or for art and design students. In each case, the system’s focus on creative presence would be valuable both for current and future collaborations.

References


Is Presence Present in Augmented Reality Systems?

Brian F. Goldiez, Jeffrey W. Dawson
Institute for Simulation and Training, University of Central Florida
{bgoldiez@ist.ucf.edu, jdawson@ist.ucf.edu}

Abstract

Presence is a construct based in virtual environment (VE) systems. There are various definitions of the term and approaches to measure presence that are grouped into objective and subjective measures. Augmented reality (AR) systems are beginning to emerge. AR uses a computer to add information (or augment) the real world. Some have suggested a continuum between VE and AR. This paper suggests a possible stronger link between AR and VE as well as suggests a theoretical basis for applying the presence concept consistently between the two.

1. Introduction

The answer to the question asked in the title of this paper is a qualified “yes”. The qualification to whether presence is present in AR is dependent on several factors including the definition of presence chosen, the degree and validity to which presence is measurable, the definition and state of AR technology, and the task(s) chosen where the technology is being applied.

The concept of presence grew from the hypotheses of Sheridan [1] and has been further refined by many authors such as Durlach, Heeter, Slater, Singer, Witmer, and others [2]. Virtual reality, virtual environment, virtual presence, and artificial reality are all similarly defined by Sheridan as something experienced by a person when sensory information is generated only by and within a computer and delivered through display technology. Sheridan goes on to describe that the sensory experience must compel the individual to feel that they are present in an environment other than the one they are actually in. He conjectures that in an ideal sense and with sufficiently good technology, an individual would not be able to distinguish between actual and virtual presence.

A contemporary definition of presence in part states that “Presence (a shortened version of the term ‘telepresence’) is a psychological state or subjective perception in which even though part or all of an individual’s current experience is generated by and/or filtered through human-made technology, part or all of the individual’s perception fails to accurately acknowledge the role of the technology in the experience. Except in the most extreme cases, the individual can indicate correctly that s/he is using the technology, but at some level and to some degree, her/his perceptions overlook that knowledge and objects, events, entities, and environments are perceived as if the technology was not involved in the experience” [3 from ISPR]. Essentially, the technology is transparent enough that entities or the artifacts the technology introduces into the environment are perceived to be as real or functional as material objects.

Historically presence has been of interest in VE systems. VE’s baseline is in a virtual or artificial environment the computer creates. Augmented reality is an emerging technology in which the real world is overlaid with additional information from a computer-generated sensory display. The real world is the baseline upon which information is added, as contrasted with virtual reality where the desired state is to completely immerse the human’s sensory systems within a computer created environment. The definition of presence above, however, allows one to consider it being present in AR systems as well as VEs.

2. Presence

Presence, a phenomenon of human experience, is triggered by external events that are technologically mediated. Slater et al. [4] describe presence as having two components: behavioral (or physical) and subjective (or psychological). Behavioral presence refers to observable responses to stimuli. Subjective presence refers to what an individual will express in response to questions about their experience in a virtual environment or about “being there”. There are various techniques used to evaluate and assess subjective and objective presence.

It has been suggested by Sheridan [1] that there are components of presence. The components seem to fall into three groups; self, physical, and social. Self involves a representation of the individual in the virtual space that in some way provides continuity between the physical and virtual representation of the person. Physical represents the person’s ability to physically affect and be affected by the virtual environment. Social represents the ability of the person to interact with other life forms represented in the virtual environment. Heeter [5] suggested a similar grouping for the purpose of measuring subjective presence. An apportionment of the presence construct among several orthogonal components might be useful for AR systems where certain features are accommodated by the real world and others are supplied by computer augmentation.
At present there is no single universally accepted method for measuring presence. However, there are a number of factors that have been suggested to influence presence. These factors include ease of interaction, length of exposure, user characteristics, etc. A more complete list can be found in the work of Sadowski [6]. Although no single approach for measuring presence has emerged, approaches generally follow either the subjective measures or objective corroborative measures. Ijsselsteijn, 2000 [7] and Slater, Usoh, and Steed, 1994 [8] respectively provide examples for objective and subjective measurement.

3. Augmented Reality and Virtual Environment Relationships

Augmented reality generally describes a technology in which computer graphics are overlaid on an actual scene. Barfield and Caudell [9] characterize augmented reality systems as overlaying computer graphics onto the real world scene through a helmet mounted display. The authors acknowledge that augmented reality systems allow information to be removed from the real world in addition to being added. They also acknowledge that augmentation may occur in additional sensory modalities, other than the visual. Whereas virtual environments completely immerse the user, augmented reality allows the user to see the real world with computer-generated images superimposed.

AR technology is somewhat immature and therefore has several barriers to overcome before it becomes widely usable in studies involving human users. For example, wearable computers are currently under development affording some freedom of movement to the user of AR systems. Accompanying mobility is the need for tracking approaches with sufficient precision, degrees of freedom, and area of coverage. Some AR barriers can leverage technologies and needs in VEs.

AR and VE sit on opposite ends of a continuum defined by Milgram [10]. AR is based in reality while VE’s seek to isolate or carefully control inputs to the user’s sensory system to define a computer generated reality. However, it is not clear where VEs stop and ARs begin. For example, AR borrows some technologies from VE, such as computer graphics and localized tracking methods. The authors of this paper believe that in the future the lines demarcating VEs and ARs described by Milgram will blur with VE becoming a subset of AR because of AR’s potential to provide all of the functionality of a VE. This is merely a conjecture, but it causes one to consider that technologies and techniques in VEs could be useful to AR. In this regard, presence is a concept worthy of exploring in AR.

4. Presence in Augmented Reality

Presence requires an appropriate forum (e.g., task) and technological platform to be triggered. Additionally, a contemporary definition, such as the one included in the ISPR [3] is useful to limit the scope of presence being considered. The forum implies a task of sufficient impact or of a sufficiently compelling nature that it will cause a suspension of disbelief. The technological platform implies using AR equipment that is sufficiently robust and with an appropriate form factor so that it can be overlooked by the user. The forum and technological platform could be related and an increase in one factor could allow for a decrease in the other. For example, an increase in the forum could result in the user allowing a decrease in robustness of the technological factor so long as the technology is consistent with and supportive of the task.

The work of Meehan et al. [11] illustrates where presence is introduced by a mixture of real and virtual. In this case the situation is based in the virtual realm with selective addition of the real. Vertigo induced stress is achieved by mixing a physical ledge (of a few centimeters) with a virtual image of an opening to the floor below.

It is useful to consider decomposition of the presence construct as suggested by Heeter [5] to better understand the extent to which specific features affect presence. The decomposition allows one to disregard, for instance, the self-component which is an integral part of the real component of an augmented reality system. Decomposition also allows one to consider the potential compelling nature of other components of presence, such as social interactions, that might be a key part of the task or forum in which the augmentation is being applied.

Mantovani and Riva believe that presence is culturally mediated through artifacts and principles [12]. They reject the concept that reality is external to peoples’ minds and consider it to be coconstructed between people and their environments. This concept is useful for considering presence in AR, because in AR people can interact with their environment through the use of computer-generated tools or overlays onto the current environment. Mantovani and Riva note that “The criterion for presence does not consist of simply reproducing the conditions of physical presence but in constructing environments in which actors may function in an ecologically valid way” (p. 547). In AR, computer-generated tools, beings, or objects that people interact with in accomplishing tasks will enhance presence.

Technological maturity concerning the robustness of current AR systems (i.e., the platform used) should be considered in a discussion of presence in AR. Current mobile AR systems have limited graphics capabilities due to the limitations of wearable computers and their form factors. These technical factors might limit the needed transparency required from equipment for the onset of presence and also limit the types of tasks that can be performed using AR.

The limitations in technology for AR are today’s limitations, but should not restrict our views of what might be possible and how presence can be a part of and influence the future of AR. If nothing else, AR systems with high levels of presence should have a positive influence on people using AR systems.

Future AR systems will be more robust and better conform to human ergonomics and kinesiology. One need only look at the early work of Ivan Sutherland in the 1960’s compared to today’s technology in order to envision AR systems that in the next decade will contain improved...
graphics, tracking, audio, network interfaces, and displays in smaller packaging in a pervasive computing environment.

AR systems have been developed for a variety of purposes such as assembly and diagnostics. Research is ongoing by the authors in using AR for decision making and navigation. Technological limitations and maturity limit the degree of utility being explored for human use, but the ongoing reduction in size and increase in power of electronics, coupled with new display technologies, will greatly expand AR applications.

4.1. Measuring Presence in AR

Decomposing presence into the components of subjective personal, social, and environmental components suggested by Heeter [5] provides an approach for considering presence in AR. "Subjective personal presence", according to Heeter, measures how much and why a person feels like s/he is in a virtual world. In AR, since information is overlaid onto the real world, this component of presence can be considered to be already strong (by replacing "virtual" with "real" world in its definition). Social presence measures how much other entities (people or beings) react to a person. In AR, helpful software agents or synthetic beings overlaid in a real terrain can provide a sense of social presence. Environmental presence can exist in AR when the environment or a computer system reacts to a person.

Measurement methods, such as the Immersive Tendencies Questionnaire/Presence Questionnaire (ITQ/PQ) of Witmer and Singer [13] or the Slater, Usoh, and Steed (SUS) [8], seem to be the most widely used methods of measuring presence, but there are several others. Can these existing methods be applied directly to AR? Generally, not in their current form, but they can be adapted or used as a platform for creating measures that are applicable to VE and AR.

A look at the ITQ and PQ illustrates why direct use is not advised. For example, question 4 in the ITQ states, “Do you ever become so involved in a movie that you are not aware of things happening around you?” The authors of this paper believe such a question might be confusing to users because it can have an embedded meaning of “being there” in addition to attentional issues. Likewise, question 32 of the PQ asks, “Was information provided through different senses in the virtual environment (e.g., vision, hearing, touch) consistent?” This question should at least be changed to reflect technologies other than VE. There are also questions that could be added. For example, one might ask, “During your experience was the equipment burdensome, such that it interfered with performing the needed tasks?”

One approach to creating a survey instrument to measure presence in AR would be to take, for example, the PQ questionnaire and delete the VE immersive-related questions, substituting those appropriate for VE and AR, and then use experimental research data to refine and validate the instrument. Another approach would be to create semantic factors conceptually specific to computer mediated presence, make questions to measure these factors in a survey, then with factor analysis refine and validate the survey. If the conceptual semantic constructs to measure AR presence are clearly defined, measuring them is straightforward.

4.2. Other Considerations of Presence in AR

Controlling presence in AR is anticipated to be more difficult than in VEs because AR is not fully immersive and events in the real world can disengage the sense of presence. Additionally, if one assumes that the representation of self is not a factor in AR presence, then the two other components are what remain to cause the onset and sustenance of presence in AR.

Presence control of physical events also needs further research to overcome technical limitations. For example, the dynamic range in brightness in the real world is far greater than can be achieved in current computer graphics and displays. Therefore, it would be difficult to model a situation where a dimmer switch was used to adjust the light level in a real room that also contained computer generated items. Likewise it might be difficult to throw a simulated ball against a solid real door when the door is suddenly opened. Similar problems can be envisioned with social interactions.

5. A Plan for Exploring Presence in AR

The above examples are illustrative and not intended to deter either AR research or approaches to study presence in AR. The examples are intended to be further impetus to the research community to press forward. The authors of this paper are involved in several synergistic activities and suggest others work toward similar goals.

The first activity is to advance current presence measurement approaches to accommodate AR systems and validate the modifications for AR while still accommodating VE systems. A review of the ITQ/PQ has shown some questions being exclusively oriented to VEs. These questions need to be replaced or reworded to accommodate AR systems. Careful consideration of task and existing AR system capabilities needs to be considered and accommodated. Modifications of existing presence measurement methods are needed if for no other reason than to acknowledge the different system characteristics and operational settings of AR versus VE. Consideration should also be given to the types of tasks and operational conditions one might encounter in using AR.

The second is exploring the realm of presence in AR systems as a stimulus to provide a better systems orientation to AR technological development. The idea is to conjecture where the introduction and control of presence might be beneficial in AR systems and conduct technological and human computer interaction related research to achieve the desired results. The design of effective AR systems rests on their ability to be seamlessly integrated into the users’ environments, so that use of AR entities and artifacts seems as natural as using any material object.
References


On the importance of reliable real-time sensorimotor dependencies for establishing telepresence

Wijnand IJsselsteijn¹ and Miriam Reiner²
¹ Eindhoven University of Technology, Eindhoven, The Netherlands
² Technion, Israel Institute of Technology, Haifa, Israel
[w.a.ijsselsteijn@tue.nl; miriamr@technion.ac.il]

Abstract

In the current paper we review the concept of (tele)presence as it relates to the active exploration of the virtual, remote, or real environments. The same sensory and brain systems responsible for a flexible mapping of our position within a spatial layout are also remarkably adaptable to include non-biological elements in the perceptual-motor loop, provided reliable, real-time sensorimotor correlations can be established. Telepresence technologies, especially those used in fluent teleoperation, essentially enable the remapping of far space to near space. The importance of actively exploring the environment through tele-bodily actions will be discussed, in particular as they relate to appropriate haptic feedback patterns for telepresence.

Keywords— telepresence, sensorimotor dependencies, perceptual-motor loop, near space, far space, haptics, teleoperation

1. Introduction

Interactive systems that allow users to control and manipulate real-world objects within a remote real environment are known as teleoperator systems. Remote-controlled manipulators (e.g. robot arms) and vehicles are being employed to enable human work in hazardous or challenging environments such as space exploration, undersea operations, or hazardous waste clean-up. They also allow for transforming the temporal and spatial scale of operation, as is the case with for instance minimally invasive surgery. In teleoperation, the human operator directly and continuously guides and causes each change in the remote manipulator. Sensors at the remote site (e.g. a stereoscopic camera, force sensors) provide continuous feedback about the slave's position in relation to the remote object, thereby closing the continuous perception-action loop that involves the operator, the master system with which she interacts locally, and the remote slave system. In the context of telerobotics, telepresence is closely associated to the sense of distal attribution, the externalisation of the self to include remote tools that phenomenologically become extensions of one's own body, even if they are not physically part of it [1].

Whereas teleoperation systems enable the manipulation of remote real-world environments and objects within it, virtual environments (VEs) allow users to interact with synthetic or computer-generated environments. In its most well-known incarnation, VEs are presented to the user via a head-mounted display (HMD) where visual information is presented to the eyes via small CRTs or LCDs, and auditory information can be presented using headphones. Importantly, the HMD is fitted with a position tracking device which provides the necessary information for the computer to calculate and render the appropriate visual and auditory perspective, congruent with the user's head and body movements. Haptic information, although not yet usually included in present-day VEs, can be added through the use of for instance an exoskeletal glove or arm, acting both as sensor and actuator.

Telepresence (in relation to teleoperation) and virtual presence (in relation to VEs) both address the psychological phenomenon of presence – the sense ‘being there’ in a mediated environment – essentially a displacement of the participant’s perception of self-location or an “illusory shift in point of view” as Dennett [2] phrased it. Perceived transparency of the medium is crucial, i.e. a sense of direct perceptual stimulation and potential for action, without an awareness of the remoteness in time or space of the simulated or reproduced realities.

Systematic research into the causes and effects of presence has started since the early 1990s, and is currently picking up speed. Although the terminology used tends to vary across authors, there appears to be a broad agreement on the major concepts. Fundamental understanding of presence is still quite rudimentary however, leading Alan Newell to comment that “researchers are working with only a seat-of-the-pants notion of the underlying concepts” and there is “an immense need for a theory and a plausible model of telepresence” (cited in [3], p.365). A generally accepted theory of presence has yet to emerge, but various approaches and models have recently been formulated (e.g. [3], [4], [5], [6], [7], [8]). Although each approach has its
own specific background, focus, and terminology, several approaches appear mutually consistent such that an integration or synthesis at a later stage is conceivable.

In the current paper, we will review the concept of (tele)presence as it relates to the active exploration of the virtual, remote, or real environments. We will argue that reliable, real-time sensorimotor correlations are of essential importance in establishing a sense of presence.

2. Determinants of telepresence

A large number of factors that may potentially influence the sense of presence have already been suggested in the literature [3], [9], [10], [11]. Two general categories of variables can determine a user's presence: (i) media characteristics, and (ii) user characteristics. Depending on the levels of appropriate, rich, and consistent sensory stimulation, varying levels of presence can be produced. It is interesting to note that both non-interactive, photorealistic displays (e.g. IMAX 3D), as well as interactive, non-realistic displays (VEs) are able to engender substantial levels of presence, although interactivity clearly appears to be the more important factor of the two [12], [13].

Sheridan [14] proposed three categories of determinants of presence:

1. The extent of sensory information presented to the participant, i.e. the amount of salient sensory information presented in a consistent manner to the appropriate senses of the user.
2. The level of control the participant has over the various sensor and interface mechanisms (tracked HMD, dataglove, etc.). This refers to the various sensorimotor contingencies, i.e. the mapping or correlation between the user's actions and the perceptible spatio-temporal effects of those actions.
3. The participant's ability to modify the environment, i.e. the ability to interact with the virtual or remote environment and to affect a change within that environment.

These three factors all refer mainly to the media form, that is, to the physical, objective properties of the media technology. Importantly, presence research is about relating those media form variables to the human response. The experience of presence is the psychological-neurological counterpart of immersive technology. It is of clear theoretical and practical value to establish what the optimal mix of cues might be for different application contexts, or, if the optimum is unattainable, which elements are most critical to the experience of presence.

The first and third of Sheridan’s determinants are a reflection of bodily presence in a physical environment. Indeed looking at presence in the most immersive environment known to us provides one possible framework of what is needed to experience presence. However, perception is not a process of simple template matching. It is more than mere ‘sensing’ of the environment through our various sense organs and subsequently comparing these sensations against passive representations of stored information. Rather, it is a highly activity-dependent and embodied process, integrating sensory data from multiple sensory modalities, which are continuously being transformed by ongoing actions (including head and eye movements), and shaped further by top-down cognitive and emotional processes. How is sensory information used to build affordances of action in a given environment [15]? It is probably rooted in evolutionary processes of adaptation to interact efficiently on the physical environment. Thus, usefulness of information becomes a central notion, relating to the extent that sensory patterns are perceived by the user, recognized as informative, and used in order to finely tune responsive acts, which in themselves, change the environment [16], [17], [18]. This is a cognitive process. Several questions are involved in identifying the components of the cognitive process intertwined with usefulness of sensory patterns: what is the threshold of sensory patterns; what are the possible arrays of multisensory patterns so that a useful mental response is triggered?

Importantly, the participant's ability to modify the environment, the third determinant that Sheridan suggests, is not an isolated component. Looking again at a fully immersive physical environment shows that sensory patterns perceived are also used for, and changed by, bodily response [16]. A surgeon for instance knows how much pressure to exert while inserting a trocar, by the feedback force that she feels. Perception and action are tightly coupled, continuously interlocked phenomena: perception guides our actions, and our actions are continuously affecting the percept of the environment.

The way the world responds to our actions can be conceived of as a set of reality tests [6]. If the world transforms in a way that is consistent with our representations of the invariants of the physical environment, we are more likely to accept the world at face value – and will feel present within it, even if this world is mediated by technology. Examples of such ‘reality tests’ include appropriate movement parallax1 as we move our heads, or appropriate haptic feedback as we move our limbs and skin with respect to surfaces and objects. An interesting extension of remote bodily presence is reported by Armel and Ramachandran [19]. They found that subjects project their real-hand sensations on a remote rubber hand, as if sensations originate in the remote hand, if synchronized tapping or striking on both hands are felt and seen. The

1 Parallax refers to the fact that when two objects are at different distances from a moving observer, their viewing angles shift at different rates. Movement parallax occurs when the motion is self-induced or active (e.g. head movements), whereas motion parallax refers to the case when motion is imposed, or passive (e.g. motion pictures).
illusions is diminished if the real hand is visible. Armel and Ramachandran suggest that the illusion and skin conductance response were due to perceptual assimilation of the remote rubber hand into one’s body image. This further suggests an illusion of perceiving sensory input from remote objects, as if being part of the user’s body.

Sheridan’s second determinant is different from the first and third – in that it does not correspond to a factor of presence in the physical reality - humans do not control interfaces in order to manipulate the environment. Hence this second determinant, the control over the interface, may indicate the level of ‘transparency’ of the interface [20]. Framed differently however, any tool that humans employ to perceive or manipulate their environment (a pair of glasses, a blind man’s cane, a pencil) can be regarded as a piece of telepresence technology, an extension of the human body that becomes seamlessly incorporated in the continuous sensorimotor loop. While the user’s acts of controlling interfaces may become automatic with extended experience, complicated control of an interface acts as a negative determinant in presence.

Determinants of presence in the physical environment may provide one possible theoretical framework for understanding presence, but not the only one. A sense of presence may also be achieved in environments that have no real-world equivalent, and the rules of interaction do not necessarily follow the rules of interaction in physical presence. See for instance the fast learning in video games that require artificial manipulation of objects. Adaptation to new consistent sensory patterns and new rules of action may allow emergence of presence in imaginary worlds.

The secret in generating the sense of presence is not in devising the ultimate technology that will imitate sensory patterns from a physical reality, but rather the minimal combination of sensory cues that are sufficient to generate the sense of ‘being and acting there’. As Heeter [21] noted, “the alchemy of presence in VR is in part a science of tradeoffs”. It is not clear at present how much each feature or perceptual cue contributes to eliciting a sense of presence for the participant (i.e. the relative weighting), or how these cues interact with each other. This is one of the issues presence research is currently aiming to resolve.

3. Remapping space through telepresence technologies

In general, the space that surrounds an individual can be meaningfully segmented into a number of ranges, usually three or four, based on principles of human perception and action. Several models have been proposed (e.g. [22], [23]), all of which distinguish between a peripersonal space (the immediate behavioural space surrounding the person) and a far or extrapersonal space. Referring to haptic space, the peripersonal space corresponds to what Lederman, Klatzky, Collins & Wardell [24] refer to as the manipulable space, i.e. within hand’s reach, whereas the extrapersonal realm would be regarded as ambulatory space, requiring exploration by movements of the body. Animal and human brain studies have confirmed this distinction between peripersonal and extrapersonal space, showing that space is not homogeneously represented in the brain [25], [26].

Neuropsychological evidence supports the notion that coding of space as peripersonal and extrapersonal is not only determined by the hand-reaching distance, but it is also dependent on how the brain represents the extension of the body space through tool use (the ‘distal attribution’ mentioned earlier). Berti and Frassinetti [27] describe a clinical case, patient P.P. who, after a right hemisphere stroke, showed a dissociation between near and far spaces in the manifestation of neglect. Using a line bisection task, the neglect was apparent in near space, but not in far space when bisecion in the far space was performed with a projection lighten. However, neglect appeared when in the far space bisecion was performed with a stick (used by the patient to reach the line) and it was as severe as neglect in the near space. Thus, this study provides evidence that an artificial extension of a person’s body (the stick) causes a remapping of far space as near space – essentially telepresence.

Like the stick in the Berti and Frassinetti study, telepresence technologies may overcome the boundaries of spatial segmentation. Telepresence can thus be viewed as an attempt to extend the personal space beyond the boundaries of the physical two meters. The boundaries of our spatial reach are defined by the focus of action, and are determined by the parameters of the interface. Within a virtual world, where sensory patterns are artificially mediated, one may touch a virtual object. Similarly, in a teleoperation environments, the user is capable of manipulating an object located at any distance, hence the diameter of personal space is extended to a particular diameter that matches the remote arm, and action space is all the remote visible space. Practically, with telepresence, the personal space and action space can be of an indefinite diameter. Vista space can then be regarded as the space that is not included in action space – the visual and auditory background. This ambient extrapersonal space plays an important role in spatial orientation, postural control, and locomotion, and is of particular relevance to presence in relation to non-interactive immersive environments, where layout and motion stimulate the peripheral visual system, controlling our more visceral responding.

4. Sensorimotor dependencies and the active exploration of environments

Perception serves the individual’s need to control relevant moment-to-moment behaviour or action within a changing environment. The development of visual perception of object shape and environmental layout is strongly dependent on consistent correlations between vision and input from other sensory systems (mainly touch and kinesthetics) through active exploratory behaviour of
the environment, establishing a stable yet flexible multisensory representation of space.

At a brain level, modality-specific feature maps can project onto one another through re-entrant connections, which allows disjunctive feature characteristics (e.g. visual and haptic properties of a stimulus) to be connected in the responses of higher-order networks. Sensorimotor correlations will initially be driven by the temporally ongoing parallel signalling between primary cortical areas receiving the sense data associated with stimulus objects at a given time and place. Next, stable feature correlations establish reciprocal connections between previously disjunct feature maps, thus allowing for higher order perception and categorisation of objects and environments [28], [29].

In a classic study, Held [30] convincingly demonstrated the relation between locomotor experience and the understanding of spatial relations. In an experiment involving dark-reared kittens, the study showed that kittens that were allowed to actively control their perception of an environment during a 42 days training phase, responded appropriately to a visual cliff – a test that requires depth perception and the understanding of spatial layout. In contrast, kittens that were only allowed passive perception during training did not show the appropriate reflexes, thereby providing support for the view that the development of depth perception is action-dependent. Observations in humans have led to similar conclusions. For instance, Verkuyl (in [31], p. 14) indicated that so-called Sachtelen children, who do not possess upper extremities due to the use of a sleeping drug in the early stages of the mother’s pregnancy, had severe problems in 3D-perception.

Similarly, telemanipulation experiments using the Delft Virtual Window System [32], [33] demonstrated a significant perceptual advantage of active observers, whose head movements controlled the movements of a remote camera (generating movement parallax), over passive observers, who received identical visual input (i.e. motion parallax), yet without the ability to actively change the viewpoint. This is in line with results found in relation to virtual environments, where Welch et al. [12] showed that participants who had active control over a simulated environment indicated higher levels of presence than participants who were passively exposed to the same environment.

Studies of adaptation to prismatic displacements provide further support for the importance of establishing reliable sensorimotor correlation maps through actively negotiating the physical environment. Held & Hein [34] studied prismatic displacements in humans under three conditions: active arm movement, passive arm movement, and no arm movement. In the active arm movement condition the subject swung her arm back and forth in the frontal plane, in the passive condition it was transported in the same manner by means of a moving cradle to which it was strapped. Results, as measured in terms of visual-motor negative aftereffects, showed that adaptation was only produced in the active movement condition and not in the passive or no-movement conditions. Another classic experiment on visual displacement [35] used lenses that turn the world upside-down. The study showed that full adaptation to this situation (i.e. seeing the world right side up again) occurred after a few days only when subjects were allowed to actively explore a complex world. When a subject was simply pushed around in a wheelchair, he did not show this perceptual adaptation to the lenses [35]. Active exploration not only influences the perception of spatial layout, but also facilitates subsequent recognition [36]. The neural basis of object and space perception provides further evidence that sensory and motor representations are closely tied together. Our peripersonal visual space appears to be represented to a large extent in terms of movement-based space, i.e. space in which objects are reachable or graspable (e.g. [25]).

5. Haptics and telepresence

Telepresence can be interpreted as the sense of ‘being there’, which often is mainly based on visual appearances. The view taken here is that telepresence applications largely require tele-bodily-actions, which need to be finely tuned in real time, to feedback from the environment. Hence presence includes here the capability to manipulate objects, at least as efficient as in the physical environment. What are the ‘appropriate’ haptic feedback patterns that lead to emergence of presence? Several factors have been suggested.

1. Consistency of haptic patterns in time - haptic patterns are consistent with haptic memories of past similar experience. Through manipulation of objects in the environment since first days of life, the user has developed a reservoir of image-schemata that relate patterns of distribution of forces on the hand to particular information/interpretation [15], [37]. If the patterns of force felt are not familiar, there is no information conveyed. New patterns may be learnt, given that the haptic feedback is consistent. Thus, this first factor refers to the cognitive-bodily capability to identify a pattern of force exerted on the hand as meaningful [17].

2. Consistency across users - requires that different users that experience a similar haptic pattern in a similar context, will extract a similar interpretation. This is especially crucial for haptic collaboration. For instance, two surgeons working remotely through a tele-surgery system, and ‘touch’ a bone, will both interpret the haptic pattern as ‘this is a bone’ [38].

3. Consistency across sensory modalities - this requires that the information conveyed by haptic patterns is consistent with the information conveyed by other sensory modalities in a given context. This may be considered as coherency of the sensory array.
The first two conditions suggest a ‘haptic language’-like reference system. By reference system we mean that a haptic pattern, a spatial field exerted and sensed by the body, is attributed a meaning. The cognitive-bodily system includes a loosely defined system of haptic patterns and associated meanings. Not only one individual uses haptic patterns to extract meanings - these meanings are shared by other users that share the same experience. Indeed a haptic language was identified in remote, fine manipulation in a telesurgery system [38], and more recently in mobile telecommunication using so-called ‘squeezy’ phones.

Haptic patterns are then used to construct interpretations about the environment. The user integrates the various interpretations into a hypothesis about presence in a particular world. This process is similar to using visual cues to construct a hypothesis about visual presence in a particular world [4], [39], [40], [41].

Fulfilment of the above three types of consistencies suggests that the feel of presence during action on the environment emerges out of resonance between haptic patterns experienced and past memories. Resonance is then, the level of mutual synergy between sensory patterns and purposeful acts of the hand on the environment. The ‘hand’ develops expectations as to the kinds of haptic sensations involved in a particular action. If the haptic feedback is different from the expected, and inconsistent, resonance will not occur. Resonance occurs when sensory memories are activated by sensory cues perceived from the environment, and simply describes whether one can rely on sensory input to support successful action [17].

Resonance can be described as a cycle of exploration of the environment. It includes: recognition of sensory patterns and their associated interpretations, motivations, intentions and goals that determine action that in turn lead to new sensory patterns, etc. [42]. To enable recognition, new haptic experience is interpreted in view of past memories of contexts and sensory-haptic experience [43]. If new sensory input follows selected cues of sensory memories, acts are finely tuned with sensory input, i.e. acts can be viewed as supported by the environment. This can be described as a closed cycle of interaction between an individual and the environment, tuned to explore the features of the environment, in order to formulate a hypothesis of presence. Figure 1 is a description of the exploration cycle from the point of view of the individual – the integration of cognitive and bodily acts to generate presence (for an extended description, please see [17]).

6. Conclusion

In line with the determinants of presence suggested by Sheridan [14], the perception of ourselves as part of a space not only depends on a passive perception of spatial layout but also on the ability to actively explore an environment, allowing the perceptual systems to construct a spatial map based on sensorimotor dependencies (see also: [44]). As mediated environments support real-time action at a distance or in virtual space, the participant is able to control certain aspects of the environment, and, as a consequence, his or her perception of the environment. By incorporating telepresence technology that supports our bodily perceptual and control movements as part of the ongoing perceptual-motor loop, the correlations between motor actions and multisensory inputs remain intact, thereby confirming one of our most important reality tests. This conception of presence places a clear emphasis on the possibilities of interaction in the peripersonal or manipulatory space, the space in which sensory and motor systems act in unison to grasp and manipulate objects, and less on the extrapersonal or ambulatory space, which seems to be particularly important for the presence-generating abilities of non-interactive media applications (e.g. widescreen cinema). Being there becomes the perceived ability to do there, as perception critically depends on successfully supported action - in line with the ecological approach to perception [45], and its subsequent application to VEs [46], [47].

![Figure 1. The integration of cognitive and bodily components in the exploration cycle of the environment](image-url)

In order for sensorimotor correlations to be mapped by the brain, they need to be reliable, robust, and without significant delays. Haptic sensory patterns need to follow the constituents of a haptic reference system: the sensorimotor patterns need to be consistent in time, so that sensory-haptic memories are developed, consistent across human users, and consistent across sensory modalities [38]. It should be noted though that telepresence does not necessarily occur immediately as a consequence of a stable correlation between local actions of the operator affecting remote effectors, and remote sensors feeding information back in real-time to the senses of the operator. Because of the intermediate interfacing technology, considerable practice may be required before teleoperators can fully experience a sense of telepresence, indicating that the lawful relation by which efference governs afference needs to be modelled in the brain [1], [48]. Nevertheless, the same sensory and brain systems responsible for flexibly mapping spatial layout and our position in real world environments, are also remarkably adaptable to include non-biological elements in the perceptual-motor loop, be they a blind man’s cane or an advanced teleoperator arm. 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 [8].

Acknowledgement

The authors would like to acknowledge the support of the IST-OMNIPRES and IST-PRESENCIA projects, both of which are funded under the FET Presence Research Initiative.

References


A Method For Designing Virtual Places

Rod McCall, Shaleph O’Neill, David Benyon, Michael Smyth and Fiona Carroll
School of Computing, Napier University, Edinburgh, E10 5DT, United Kingdom
{initial.surname@napier.ac.uk}

Abstract

This paper discusses our experiences of using a range of methods and techniques to measure the sense of place in real and virtual environments. The paper presents a discussion on presence and how this is linked directly with our sense of place. From here we discuss the development of ‘The Place Probe’ a bundle of measurement techniques that allows for the direct comparison of real and virtual scenes. We show how the data gathered from the probe can be used to inform the design of virtual places. The whole approach is predicated on the premise that by understanding place we can improve our sense of presence in virtual environments.

1 Introduction

The BENOGO project is a fifth framework project in the ‘Presence’ initiative of the Future and Emerging Technologies programme. The project seeks new ways to give people a sense of ‘being there’ without having to go there. Using photo-realistic image-based rendering of real scenes, the BENOGO technology provides a real time experience of the scene. Stereo images are generated from 2D photographs and rendered in an area such as a head mounted display or six-sided CAVE that allows for real-time rendering of the images as tracked by head mounted tracking devices. The overriding aim is to inform the design of these virtual places. In doing this there is a need to investigate what it is that gives a place its distinctive sense of place (as opposed to it being ‘placeless’ [1]) and how to produce an environment in which people feel a sense of presence; they have a feeling of ‘being there’.

Our interest, then, is in (a) our sense of presence in a particular place and (b) how to compare real and virtual places effectively. These two problems led us to develop the ‘place probe’, a set of techniques that locates the idea of place at the centre of presence research.

In essence a physical environment is a space, or a setting, whereas a place is somewhere to which people have attached additional meanings, feelings and interpretations. For example a room has little or no significance for us when we enter it for no purpose. However, if we enter the same room when undertaking an exam it may take on additional feelings such as fear or success, which are purely personal but often related to other physical cues such as other people, silence and the sound of clocks. Different spaces may become different places for different people. For example, the design of city centre parks may provide convenient lunchtime seating for office workers but as night falls may also provide unforeseen challenges for the city’s skateboarders. While the ‘found space’ [2] remains the same for each group it is contended that the sense of place is quite different. In order to provide people with a real sense of ‘being there’ (i.e. a sense of presence) there is a need for a method that allows designers of virtual environments to capture key points about both the setting (physical attributes) and the experience (meanings, feelings and other abstract properties).

The paper is organized as follows. Section 2 provides a brief discussion of place and presence. This is followed by a review of existing methods for measuring a sense of presence and of our experiences in using these methods. Section 4 introduces the place probe and explains the rationale for including the various components. These are illustrated with examples from the use of the probe in empirical work. Section 8 illustrates how the data from the probe is captured as a number of interacting patterns of place that can be used by designers and evaluators of virtual places. A brief conclusion provides insight as to how this work might develop.

2 The Link Between Presence and Place

There are many definitions of presence. For example, Lombard and Ditton [3] describe it as the ‘illusion of non-mediation’. Others prefer the notion that it is of ‘being there’. For the purposes of the work contained here a definition of ‘the feeling of being somewhere’ (whether that be in a real or virtual environment) is used. As there is a need distinguish a sense of place as part of presence as opposed to a sense of placelessness. We want people not just to tick a box saying they had a sense of ‘being there’ we want them to tick a box saying they had a sense of ‘being somewhere’ (specific).

Figure 1 The form and content of a medium used in virtual environments has an impact
In order for people to ignore the media they are using (e.g. the head mount display or cave) and feel a specific sense of place there is a need to explore which aspects of presence are most relevant. At the basic level Sheridan [4] indicated that presence was derived from (i) characteristics of the medium and (ii) characteristics of the user. Of which characteristics of the medium can be broken down into media form and media content, the work here focuses primarily on the media content aspect. That is to say which aspects of the place we can utilize in order for people to feel they are in the specific place and thus experience a high degree of presence.

In terms of what constitutes place, a number of commentators have offered definitions. For example Relph’s view of place [1] describes three properties; activities, physical properties and meanings (see table 1), a view shared by Norberg-Schultz [5] on his view of existential spaces (see figure 2). In many ways the Relph and Norberg-Schultz view of place share many aspects of the separation between media form and content. With content consisting of the aspects found in the Relph and Norberg-Schultz models, for example meaning and expressive space, when combined with the physical attributes such as layout.

<table>
<thead>
<tr>
<th>Property</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical Properties</td>
<td>Buildings, furniture, other people</td>
</tr>
<tr>
<td>Activities</td>
<td>Walking, reading, sitting</td>
</tr>
<tr>
<td>Meanings</td>
<td>Fun, boring, sad</td>
</tr>
</tbody>
</table>

Table 1 The Relph model of place

The idea of place has been extended by Gustafson [6] who places the person, and their relationship to others and the environment at the center of their experience of a place. The Gustafson model covers many of the areas found in other models, such as the importance of physical properties (such as environment) and a person’s interpretation of it (through their feeling of self).

- **Self** - experience, memories, emotions and activities.
- **Environment** - natural/built /symbolic or historical/institutional environment.
- **Others** - characteristics, behaviours, traits of the inhabitants of these places.
- **Self/others** - places become meaningful as a result of people living there;
- **Others/environment** - atmosphere, climate and street life.
- **Environment/self** - respondents knowledge of the place, familiarity;
- **Self/Others/Environment** - themes that involve all three main parts i.e. traditions, festivals and anniversaries.

The work here makes use of both the Relph and Gustafson’s models of place. However, the emphasis on experiences over a longer term that characterizes much of Gustafson’s model is not relevant. It is also the case that the social nature of Gustafson framework while relevant for single user environments, probably lends itself more to work on collaborative virtual environments where there are issues of social and co-presence are more relevant. At present our focus is capturing the essential characteristics of an individual’s experience of a place.

3 Experiences of Existing Methods

During 2003-2004 a number of studies of real and virtual environments were conducted [7], using a variety of investigative methods. The objective of these studies was to measure the sense of presence in virtual environments, and where possible to compare that to the sense of presence in real environments. Another objective was to explore the features that made people feel present in a particular location. From a methods perspective we were interested in exploring what measuring instruments were available and the types of data each method was capable of uncovering.

3.1 Questionnaires

Measures of presence have often focused on the use of questionnaires such as the one developed by Goldsmiths College (University of London) for the UK’s Independent Television Commission [8]. It is a cross media presence questionnaire exploring four different measures; sense of spatial presence, the level of engagement, the sense of naturalness, and the negative effects experienced in an environment. There is no overall score for presence in this questionnaire as presence is regarded as a culmination of all factors.
While the ITC-SOPI questionnaire provides useful data on the levels of presence felt within an environment it is difficult to use the results for providing input into the design of a system or to highlight areas of weakness.

3.2 Video Analysis

The studies made use of video analysis with talk-aloud protocols as a means of understanding what people were doing while they experienced a real or virtual environment. Video analysis of the kind used provides vast quantities of data that were analysed from a semiotic perspective.

Table 2 Transcript from a Video Session

<table>
<thead>
<tr>
<th>Time</th>
<th>Visual</th>
<th>Audio</th>
</tr>
</thead>
<tbody>
<tr>
<td>02:58:00</td>
<td>Table (and camera legs)</td>
<td>Doors opening and closing</td>
</tr>
</tbody>
</table>

Transcription
- Right there’s something not right here with the table
- Uhuh it was kind of moving towards me, under me
- Right ok

Table 3 Questions used during one of the interviews is contained in table 3.

1. Initial open question ‘describe the main features that you experienced while in the room’
2. Describe the sense of scale that you experienced while in the room?
3. What were you aware of while you were in the room?
4. Describe your personal feelings about the room
5. Did you experience enjoyment while exploring the room?
6. Did you feel part of, or engaged by, the experience?
7. While exploring did you experience a sense of movement, either of the objects in the room or yourself? E.g. Differences in sounds, texture gradients etc.
8. What in your opinion were the three most striking features of the room?

Table 4 Questions used in the structured interviews

From here an essentially enumerative strategy based on categorizing the data into various themes was used. For example how real/natural the environment/experience looked and felt to the participant, what they were aware of? How involved they were in the scenario? And what technical issues arose? To make analysis simpler each of these themes was represented with a code: realism, aesthetic, technical issues, involvement and physical objects. Each of these codes, permits the identification relationships and associations.

3.3 Structured Interviews

During early studies structured interviews were used to find out about the experience people were having of the real and virtual places. Such interviews allow questioning to be concentrated on specific areas of interest and to explore particular responses made by the participants. A list of sample questions used during one of the interviews is contained in table 3.

3.4 Repertory Grid Analysis

A repertory grid [9] is a method of analysing the meanings that a person has attached to something, for example an object, activity, location or the virtual world used in this study. A participant is first asked to provide a series of elements (or descriptions) of their experience, they are then asked to assign these elements along a series of bi-polar constructs (see figure 4). The bi-polar constructs are usually derived from the elements selected by the participant, or (as in the case of this study) they may be supplied by the evaluator. The constructs used were based on Relph’s model of place (see table 1).
Table 4 A list of constructs used in the repertory grids

<table>
<thead>
<tr>
<th>Constructs</th>
<th>Permanent</th>
<th>Temporary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colourful</td>
<td>Colourless</td>
<td></td>
</tr>
<tr>
<td>Noisy</td>
<td>Quiet</td>
<td></td>
</tr>
<tr>
<td>Still</td>
<td>Full of Movement</td>
<td></td>
</tr>
<tr>
<td>I could explore</td>
<td>I could not explore</td>
<td></td>
</tr>
<tr>
<td>I could meet my friends here</td>
<td>I could not meet my friends here</td>
<td></td>
</tr>
<tr>
<td>I would like to work here</td>
<td>I would not like to work here</td>
<td></td>
</tr>
<tr>
<td>I could touch this</td>
<td>I could not touch this</td>
<td></td>
</tr>
<tr>
<td>Surprising</td>
<td>Unsurprising</td>
<td></td>
</tr>
<tr>
<td>Attractive</td>
<td>Ugly</td>
<td></td>
</tr>
<tr>
<td>Interesting</td>
<td>Dull</td>
<td></td>
</tr>
<tr>
<td>Designed for Work</td>
<td>Designed for pleasure</td>
<td></td>
</tr>
</tbody>
</table>

Table 5. Words used to describe one environment

<table>
<thead>
<tr>
<th>Construct</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stairs</td>
<td></td>
<td></td>
<td>Stressful</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pleasant</td>
<td></td>
<td>Grey</td>
<td>Dark</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Desk</td>
<td></td>
<td>People</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Signs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In the example below a person is asked the question. They are then asked to place words they have chosen within the grid (table 5), based on whether they would like to work (a score of 1) or not like to work (a score of 5) within the environment.

3.4.1 Example

Question asked “Can you give me eight words which best describe the environment you’ve just been in?”

Elements (words) chosen: dark, pleasant, stressful, busy, grey, stairs, desk, people, signs

Construct:
1 = I would like to work here
3 = No overall opinion
5 = I would NOT like to work here

4 Comments on Use of Methods

In using a combination of methods the studies were able to uncover themes and issues that appeared across multiple sources. However methods such as repertory grids and interviews both require the evaluator to be present and are time consuming to conduct and in the case of interviews also take some time to analyze.

5 The Place Probe

A key factor in developing a new technology is to provide useful feedback in the design process. While all the methods mentioned above proved to be useful in exploring participants sense of place, they all took a considerable time to administer and provided little useful information for the technologists working on the project in terms of how to enhance the sense of place in the VE. For the BENOGO project it became paramount to devise a method that was both quick and easy to administer in various locations, both real and virtual, and that provided essential benchmarking information that could be used by the technologists to improve the VE. Based on the experiences of the methods discussed earlier a new way of capturing core aspects of real and virtual places was devised. This new method called 'The place probe' [10] was designed to be more flexible in its approach to capturing data, to produce results relevant for the virtual reality technology being used and to be easy to administer while basing itself around the key ideas of the existing methods.

The concept of the probe drew on earlier ideas by Gaver [11] and Baillie [12]. The probe is designed to be used by visitors to the real or virtual scenes, where they can complete it either in the presence of the evaluators or on their own, as it is easy to read and understand. This allows it to be used on locations where the evaluators may not be present. It has also been designed to allow for relatively quick data analysis, for example by asking for keywords or scores, through to more complex data such as those found in sketch maps and written descriptions.

The probe is designed to capture key information about real scenes so that they can be incorporated in to the virtual equivalent. It is also intended to allow for a comparison (a process referred to as benchmarking) between a real and virtual version so that evaluators can uncover areas of difference. The data ranges from specific information on a scene, such as that uncovered during written descriptions and sketch maps to high level data about the overall experience (in the semantic differentials).

Each data source (with the exception of the semantic differentials and photographs) were analysed by a member of the team using a grounded theory approach. This allows for themes to emerge that exist within the data rather than adopting a series of pre-set categories. In order to verify the data samples are then chosen at random from each data source and given to the other evaluators who then note down any themes. The data found and emerging themes were then compared.

Prior to selecting a real world location a checklist was used to decide if it is was suitable. The checklist was developed to accommodate a number of technical considerations which effect the choice of locations used for the virtual environments for example; lighting, the amount of movement offered to the participant and the geometry of the space. As the project evolves it is
expected that the criteria will change, moreover others using the probe will no doubt need to alter the probe to suite their own requirements.

5.1 Profile

Basic profile information is obtained to permit easy comparison between people, for example males, females or different age ranges. An example from a navigational perspective is that of children who were asked to find their way around a university campus [15]. The children all remembered features at their eye-level or below for example dogs and bins, however they become lost when these features were removed. While navigation is not the only task which people undertake in virtual and real environments, the example illustrates the variation in the types of features remembered by different groups of people.

5.2 Descriptions

This part is similar to a visitors book and asks people to write down a short description of their experience. This approach allows for a totally free form response and is not restricted by asking for information on specific topics.

Atlas.ti was used to analyse the descriptive component of the probe. Atlas.ti was useful as it allowed certain quotes within the content to be highlighted and then to be easily organised and grouped into common codes/themes for example in the viewpoint real study, a quote like 'we can pick out landmarks we know – the castle, Tyn Church and St. Nicolas' and 'a good point of view. St Vitus is a marvellous church' can be categorised into a theme like physical location. The themes/codes were chosen as a means of best representing/categorising a group of quotes. These individual themes/codes were then discussed and compared with the themes of the other researchers and an agreement was reached on whose theme/code was most appropriate for the situation.

5.3 Sketch Maps

Sketch maps (see figure 3) can be used to assess which aspects of an environment a person recalls, and areas where they stood and walked around. The objective being to uncover the most salient aspects of the environment and if required concentrate any computing resources on them. The sketch maps may also prove useful in identifying which areas of the environment people are most likely to visit and hence support for allowing avatar movement in these areas of the virtual environment can be provided. A method of analysis from Billinghurst and Weghorst [16] was chosen, however the procedure was simplified by not looking at aspects such as overall sketch map accuracy and orientation, instead the emphasis was on how many times a feature was drawn in by all the participants.

The sketch maps provide useful information on aspects of the scene which people consider most relevant, for example they draw the Cathedral but not another large less prominent building next to it. They are also open to some issues such as drawing ability, subject exposure to the environment and the evaluators ability to recognise objects within the drawing.

![Figure 3 A sketch map drawn by one child of the Technical Museum in Prague](image)

A number of evaluators would examine a series of maps to assess the level of agreement of categorising objects. In the example below there is some disagreement on the name of the marked object, however on checking travel guides for Prague it is possible to identify the correct name of the building.

The sketch maps from the real and virtual scenes are compared to see if people are drawing the environments in the same level of detail. For example the study of the hilltop location in Prague found that distant objects in the virtual environment were not accompanied by supporting information such as their name and tended to appear as collection of buildings rather then being drawn separately.

5.4 Semantic Differentials

The objective of semantic differentials is to uncover any connotative associations that people have with certain words and the environment they have just visited. To do this people are asked to rate their experience of the environment on a series of bi-polar scales, for example in the table below the person has illustrated that they found the environment 'very attractive' whereas they could equally have indicated that it was 'very ugly'. Semantic differentials produce a set of scores related to how people feel towards the environment and are easy to analyse.

The differentials were based on our work with Rep Grids mentioned earlier and are essentially a quick method of gauging participants responses to an environment in relation to Relph’s three categories; physical properties, activities afforded and meaning. Within these categories themes that were uncovered from the repertory grid studies are used, for example within the physical features section we find attractive, big, colourful and noisy. Within the activities section, temporary, available, versatile, and interactive. Within the meanings section pleasant, interesting and stressful.
### Table 2 The semantic differentials used

<table>
<thead>
<tr>
<th></th>
<th>Very</th>
<th>Quite</th>
<th>Nearly</th>
<th>Ugly</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attractive</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Big</td>
<td></td>
<td></td>
<td></td>
<td>Ugly</td>
</tr>
<tr>
<td>Colourful</td>
<td></td>
<td></td>
<td>Colourless</td>
<td></td>
</tr>
<tr>
<td>Noisy</td>
<td></td>
<td></td>
<td>Quiet</td>
<td></td>
</tr>
<tr>
<td>Temporary</td>
<td></td>
<td></td>
<td>Permanent</td>
<td></td>
</tr>
<tr>
<td>Available</td>
<td></td>
<td></td>
<td>Unavailable</td>
<td></td>
</tr>
<tr>
<td>Versatile</td>
<td></td>
<td></td>
<td>Limited</td>
<td></td>
</tr>
<tr>
<td>Interactive</td>
<td></td>
<td></td>
<td>Passive</td>
<td></td>
</tr>
<tr>
<td>Pleasant</td>
<td></td>
<td></td>
<td>Unpleasant</td>
<td></td>
</tr>
<tr>
<td>Interesting</td>
<td></td>
<td></td>
<td>Boring</td>
<td></td>
</tr>
<tr>
<td>Stressful</td>
<td></td>
<td></td>
<td>Relaxing</td>
<td></td>
</tr>
</tbody>
</table>

Results from the differentials are useful in providing information on the high level experience people have of a particular place for example whether it is ugly or attractive. It is also possible to quickly compare the data obtained for real and virtual scenes and uncover areas of difference. For example, results from one study found people in a virtual environment find less attractive than its real counterpart. Data from the semantic differentials is useful for corroborating results from other sources, and it can be used as a starting point from which to explore data found in other sources.

### 5.5 Photograph

A set of six photographs was taken at each real world location visited by the evaluators. These were then given to the participants in the study who were asked to select the one which best represented their experience of the location they had or were visiting. This approach is useful for finding which view people felt best represented the environment they had visited. In turn this provided the designers with information on the most common standing locations of people within the place. For example this could be used to indicate a starting location in the environment, a standing point (if the technology allows limited movement) or a location where images can be captured from (if the technology uses real world images e.g. photo-realistic VR).

### 5.6 Six Words

Participants are asked to write down the six words which best describe their experience.

The six words section involves counting the number of incidences of common groups or terms. In common with the descriptions the use of a dictionary and thesaurus was useful in pinning down specific meanings in cases where the evaluators were unsure.

---

### 6 Study

The first study which made use of the probe for design and evaluation was conducted during 2003. For this a number of locations in Prague were visited with a view to selecting those which were most appropriate for the technology. The locations were chosen based on a list of properties known as the ‘The Place Suitability Checklist’.

### 7 Themes Workshop

After analysing each source of data a workshop is held. This is a forum where all evaluators gather to discuss the themes which emerged within the data with a view to agreeing on a common set of templates. In order to do this each evaluator is asked to discuss the main themes which emerged within their data (e.g. the descriptions), in some cases noting them on a whiteboard (See figures 4 and 5). The themes are then checked against the samples chosen from each data type to ensure that similar items are being identified and to discuss any discrepancies.

### Figure 4 A group of themes on the left being situated within the Gustafson model

During the next stage the evaluators begin to explore which themes appear across the range of data sources (e.g. within the descriptions and six words), with a view to retaining those which appear in more than one data source and possibly eliminating those which appear in only one source. They also look for links between themes in different data sources so that any which refer to the same feature or phenomena can be grouped together.

The ultimate objective of developing the themes is to provide designers and evaluators with a series of templates which can be used to describe a given location. This is similar in many respects to the idea of architectural design patterns by Alexander [17] which are sets of high-level common description of architectural features such as paths & goals and café’s. However as documented later there is need to develop in a way which is relevant to the technology which will be used to operate the virtual environment.
The final stage of the data analysis involves grouping the range of themes within a chosen model of place. To date the studies have used the Gustafson model (see figure 2 with descriptions), which is drawn primarily from a sociological perspective and Relph (see table 1 with descriptions). The Gustafson model of place seemed unduly complex for the task which was primarily communicating results to technologists who were building the virtual environments, where as the Relph model provided three distinct and readily understandable categories.

8 Place Model

The place model is the result of the final stage of the analysis (see figure 6). In the example model derived from a technical museum in Prague there are three main levels, which are based on Relph's model of place and the technical aspects of the VR system. The top area (dark) represents the meanings people have attached to their experience, the central area physical properties of the environment they have noticed and the lower (dark) area represents the technical issues which affect the environment. The central circle represents an activity people undertook while in the museum that should be encouraged within the virtual environment.

In addition to the place model designers and evaluators can view on-line supporting documentation. At present this take the form of a webpage which contains a definition, links to related templates and relevant data from the studies.

The place model represents a tool for communicating information to the developers of virtual reality systems. The aim being to allow for such information on the real environment to play an important role in the early design stages of a project. For example in the above place diagram the technical requirement that there should be sufficient resolution is linked with the need to be able to clearly identify exhibits within the environment.
museum. This could be interpreted not only as providing a sufficiently high resolution for the entire environment, but also augmenting certain objects at higher resolutions so that people can get close to them and still be able to recognize them. The issue of closeness is highlighted as a desirable property on its own but is also linked to a range of other features such as the occlusion of objects within the space. While occlusion of objects is a natural property of the real environment, the ability to provide it within photo-realistic virtual environments is problematic and there are limitations on how and where it can be provided. As a result it is important to indicate this is an important issue of the environment which if possible should be addressed within any implementation of the space.

From an evaluation perspective the issues and links between the topics in the place chart provide a method of finding out whether people are having similar experiences within the real and virtual environments, for example are they able to look at objects which are close-by without the resolution causing problems. Also the data from semantic differentials can be used to find if people are having the same views on the experience.

9 Conclusions

The concept of presence is open to much discussion and debate, not only on whether it exists but through to definitions and how systems can be better designed to support it. However it is clear when comparing place and presence that there is a degree of overlap and that the idea of place is rarely fully explored by the presence community. This paper explores the link between presence and place and how that can be exploited to provide relevant design and evaluation advice for developers of virtual environments. In contrast with certain other methods this results in a situation where specific properties can be discussed and ultimately included in a virtual environment, on the premise that if these features are provided then a higher degree of presence will occur. Rather than abstract notions of presence which produce scores but provide little or no indication of what is right or wrong with a particular environment.

The ‘Place Probe’ discussed in this paper is in development but provides a quick and reliable method of capturing information about real and virtual scenes. While the method builds on aspects of place from literature the results are not tied to any specific models of place and adopt a grounded theory approach, with specific frameworks (e.g. Relph or Gustafson) only being used in the final stages of analysis. This permits the data to be framed in specific contexts, for example designers and evaluators of collaborative systems may chose the Gustafson framework as it places emphasis on the relationships between people, their environment and others. Where as if this is not relevant then the Relph model may prove useful.

In conclusion this paper presents a review of our experiences of using existing methods of capturing place and presence and how based these a new method ‘The Place Probe’ was devised. The ‘place probe’ has been used in a variety of studies of real and virtual scenes and provides relevant information and can be applied to a range of environments and contexts.

10 Acknowledgements

The work in this paper is funded by European Union BENOGO project (IST-2001-39184). We acknowledge the work of the other project partners and the members of staff at the Technical Museum in Prague and other locations visited during our studies.

11 References

A Constructionist Cognitive Model of Presence

David Nunez
Collaborative Visual Computing Laboratory, University of Cape Town
dave@cs.uct.ac.za

Abstract
This paper proposes a cognitive model of presence which emphasizes the interaction of top-down and bottom-up processes, and the subsequent creation of temporary representations of an environment in working memory. These temporary representations (which are termed constructions) are formations of meaning about the environment, which allow a subject to make inferences and interact in that environment. This paper argues that this constructionist position follows from the existence of hypothesis selection models of presence such as proposed by Slater & Steed. Finally, the paper contrasts this model to that of Wirth et al and discusses some of the implications of this model for presence in unfamiliar or unrealistic environments.

1. Cognitive models of presence
With the recent move towards a comprehensive theory of presence, cognitive science has surfaced as an important basis for understanding presence. For example [1], in discussing breaks in presence, propose a filtering, information-processing model in which one of two information streams (real environment/virtual environment) is selected for processing. Many other examples of the application of cognitive theory in presence exist - for instance [2] and [3]. Although it is generally accepted that presence is a complex, multi-factor construct, cognition seems to be significant enough a factor that an understanding of the cognition of virtual environments (and indeed real environments) can lead to invaluable insights into fundamental aspects of presence. Therefore, although this paper presents an exclusively cognitive model of presence, it should be understood as an attempt to understand the role of cognition in presence, rather than an argument that a complete understanding of presence can come from examining cognition alone.

2. Constructionism in cognition
Constructionism involves the interaction of top-down and bottom-up processes which create a mental construct reflecting some external or imagined situation in such a way that inferences and decisions can be made in relation to that situation [4] [5]. Furthermore, constructionism is generally associated with a dynamic state of cognition in which bottom-up processing activates particular top-down processes which in turn lead to a bias in bottom-up processing [6]. The ‘rabbit/duck’ switching illusion discussed in [1] is an example of two constructions arising from the same bottom-up data, but mediated by a different top-down bias. Constructions are capable of adapting to changes in the external world, but only to a degree; if the products of bottom-up processes lead to a significant mismatch with the construction, a reconstruction will occur so as to better represent the stimulus situation [7]. That constructionism is useful in presence research has already been argued by [8], and indeed, constructionism already exists in the literature in various forms. [9] for instance, argues for the interaction of internal and external factors as a determinant of presence. More recently, [10] describes a concept (perception and action in 3d space) in which a continual re-consideration of the environment in terms of the user’s cognition, perception and emotions occurs. Also, [11] proposes a model in which presence is an active process in which sensory and conceptual data are used to construct mental models of the environment (see section 7 below for a discussion of this model).

3. The need for constructionism: The virtual stimuli problem
In our view, constructionism is a useful tool in modeling presence because it solves an important problem in current thinking: namely, the problem of virtual stimuli. Often in the literature a distinction is made between stimuli which originate in the virtual environment, and stimuli which originate from other sources (for example, [1] and [12]). This distinction is useful because it allows one to ask questions in terms of relevant information and distracters. However, this distinction is also somewhat problematic. From a physiological point of view, there is only one stream of input for a user of a virtual environment – all stimuli are real. All sound, regardless of its origin, is received by the same receptors in the ear. Similarly, all visual stimulation, regardless of its point of origin, is received
by the same receptors in the retina. Once these receptors are stimulated, the original stimulus ends its role in perception; the receptors generate neural signals from which cognition proceeds. Therefore, once the stimulated receptors have fired their impulses, all information about the origin of the percept must be inferred via a series of cognitive processes - its position in space, its semantic relationship to other percepts, what object that stimulus represents, and where that stimulus originated.

4. Cognitive constructions

The ‘two streams’ of information between which a user selects [1] therefore exist not as external streams of information one of which is selected, but rather as two alternative interpretations or constructions of a complex set of perceptions. In one of these constructions (the ‘virtual construction’) the user interprets their mental state (the perceptions and the inferences drawn from these) as a coherent virtual environment. In the other construction (the ‘real construction’) the user will interpret their mental state as being in the experimental venue, viewing a display. In each case, information which matches the user’s basic idea (“this is the virtual environment”) will be added to the construction, while information which does not match the construction will be attended out, or act as a distraction [11]. The construction thus becomes not only a basis of inference for perceptions, but also works to filter out irrelevant stimuli and bias perceptions to create a coherent model of the environment. In the constructionist view, there need not only be two possible constructions – depending on the mental state of the subject, any number of possible constructions are possible, many of which could be considered as evidence of presence. For example, placing the subject inside a cave system displaying the inside of a hospital could lead to, depending on the emotional or semantic aspects of the active construction, the experience of a frightening hospital, a comforting hospital, a friendly VR laboratory or a stressful VR laboratory. It should be noted that the choice of which stimuli are used to form the construction depends not on the actual origin of the stimulus (the real or virtual environment), but rather on the current state of the construction, and the biases on attention and interpretation which it exerts. Thus, it is possible for stimuli arising outside the virtual environment (such as a phone ringing in the office next door) to be incorporated into the construction of the virtual environment.

5. A constructionist cognitive model of presence

The constructionist cognitive model, proposed here, makes use of a basic information processing architecture (such as used by [13]). In this architecture (depicted in Figure 1), stimuli are selected for further processing by an attention filter. The selected inputs are then processed and transformed in working memory. These transformations are performed using rules and data contained in long-term memory. Thus, the contents of working memory are constructed by means of bottom-up input (stimuli from the attention filter), as well as from top-down input (contained in knowledge structures and transformation rules) from long-term memory. The products of this processing in working memory represent temporary structures such as mental models, themes, cognitive maps and other constructs which allow the subject to make inferences about the world and thus interact with and navigate through it.

![Figure 1: Information flow in the constructionist cognitive model](image-url)

The model represents a continual upward and downward flow of information and, by consequence, the continual construction and reconstruction of temporary structures in working memory. Information taken in by the senses is first partly processed by the attention filter, which itself contains a bias towards permitting or blocking certain types of information on the basis of relevance [14]. This information is then organized in working memory into temporary constructions such as semantic chunks or mental models [13]. The rules by which this organization occurs are encoded as activated knowledge structures (e.g. schemata or scripts) in long term memory [15]. These structures not only provide rules by which to organize information, but also fill in missing data with that learned in previous experiences [15]. As this occurs, the temporary structures in working memory are adapted to create a state which accommodates the perceptual information provided by
the attention filter as well as the active structures in long-term memory. At the same time, the active structures in long-term memory are affecting the attention filter by creating a bias towards filtering out unrelated information.

5.1 Coherence, construction and re-construction

Constructions are formed partly on the basis of previous experience and learning, and as such represent attempts to fit a current stimulus set into a meaningful whole (encoded by existing knowledge structures in long-term memory). In essence, forming a construction is a process of inferring meaning. Therefore, it follows that not every stimulus will be compatible with the construction that is active (for instance, a helicopter would not ordinarily be compatible with a construction of prehistoric Earth). Stimuli which are not able to fit into the construction are selected out by attention, which is biased (by the construction) to only allow in relevant stimuli - in this way, constructions are kept coherent. However, it is also possible for a stimulus, related or not, to demand attention to itself and thus force its way into consciousness [7]. This usually occurs in the case of stimuli which represent large, sudden differences between the outside world and the perceiver’s expectations. Thus a sudden noise or movement, an expected tug of a cable or similar stimulus will force its way into consciousness where, if a large disparity between the perception and active construction exists, will force a re-evaluation of the situation. When this occurs, a reconstruction takes place which incorporates the new stimuli. However, when this reconstruction occurs, the re-assignment of cognitive resources is felt – first as an odd sense of mismatch or confusion (as the construction fails to match the situation), and then as a subsiding of that sensation and with a sense of realization of the new situation as the reconstruction occurs [7].

5.2 Meaning activation & feedback

In this model, the meaning which the subject constructs about the environment is crucial for their behaviour, and thus affects all levels of the model. Although meaning originates in the top-most level (knowledge structures in long term memory), it can flow downwards to working memory, as well as to the attention filter in the form of a bias for selection. The subject’s current construction also affects the temporary objects being formed in working memory. Thus, knowledge structures contribute to the construction of temporary structures in a ‘broad brush’ approach (due to the activation of categories, concepts, etc. [15]), although the specific details of the construction are determined by the stimuli. In this way, the meaning inferred by the subject will allow for interpretation and prediction of the state of the environment, while the perceptual stimuli allowed in by the attention filter will ensure successful behaviour which matches the true state of the environment.

5.4 Presence in the model

In this model, presence exists as the degree to which the construction is about the content of the virtual environment rather than about the display system or the experimental situation. This definition is in line with the concept of cognitive presence proposed by [3] as well as with the concept of ‘pretence’ proposed by Slater, in which presence is taken as when the subject is acting and thinking ‘as if’ in the virtual environment [17]. In order for this to occur, the construction will need to have a high degree of match to the stimuli (as would occur in a high-fidelity, highly immersive display); if not, the subject will experience a re-construction, which will disturb cognition about the virtual environment and thus reduce presence. Although the subject’s own biases and previous experience will contribute to their experience (which one might call the ‘willing suspension of disbelief’) large disturbances in the stimuli will demand attention and are more likely to cause reconstructions (as occurs in a break in presence).

6. Construction flexibility and unrealistic environments

Although constructions tend to limit the information in consciousness to create a sense of coherence in the world, they do allow a great deal of flexibility in interpreting stimuli. A construction operates simply by means of creating expectations, and attempting to interpret meaning in terms of those expectations. In short, constructions operate on the level of concepts rather than specific objects. Thus, if we have constructed a restaurant from the available perceptions, then we will have an expectation for an interaction with a waiter at some point. We will also have general expectations about that waiter (he will take a food order, will display a particular degree of politeness and so on), but not specific expectations such as whether the waiter will be a man or woman, young or old, dressed in a particular way or not [15]. What exactly those expectations will be will depend on cultural conditioning and personal experience [4], but they will always be general enough to allow flexibility enough to allow successful interaction in the world.

Because constructions rely so heavily on the expectations and data provided by top-down processes, an interesting question arises about constructions of unrealistic or highly unfamiliar environments. In such a situation, our model predicts that the closest fitting available knowledge will be activated, but due to the poor fit with bottom-up data, constant reconstructions will occur until enough information has been gathered to begin forming new top-level structures about the environment. Anti-presence [16] could be taken as an example of this process. Describing anti-presence [16] notes that people witnessing traumatic events often feel as if they were experiencing a film or other mediated environment rather than reality. From the constructionist model, one would argue that the closest fitting top-down
knowledge people tend to have about traumatic events are all from films and other media, and thus it is these structures which become active during the constructions. Gradually however, with increased experience, top-level structures to deal with the new type of situation will form, which will allow more useful constructions to be formed.

7. Cognitive constructions vs. spatial situation models

Wirth et al [11] have proposed a model which emphasizes the subject’s construction of a mental model of the spatial situation. In this model, a spatial situation model (SSM), which encodes information about a physical space and the observer’s relation to it, is constructed from information presented by the encoding medium as well as from internal factors which include previous knowledge and context. In this model, the SSMs are formed and then tested against further input as the subject interacts with the world. The cognitive construction model we propose, although similar in its basic constructionist philosophy, is different from Wirth et al’s model in several important respects. Firstly, the constructionist model differs in that it sees constructions operating not only at the level of spatial or mental models of the environment, but rather at the higher level of meaning about the environment. Furthermore, the construction of meaning affects lower level cognitive processes (such as attention), which in turn affects the activation of higher level processes, thus making cognitive construction a highly interactive, continual process. Finally, the constructionist model explicitly models effects on attention, in line with findings from priming studies in attention (for example, [14]). Thus while this model is completely compatible with Wirth et al’s (in that spatial situational models can be seen as intermediate level constructed structures), construction is a far more fundamental process which involves all levels of cognition in an interactive way.

8. Conclusion

The constructionist cognitive model of presence is a useful tool not only for investigating the role of cognition in presence, but also for asking such questions as the importance of previous experience and thematic coherence as factors in presence. Although the model is only in its conceptual stages, it is proving to be a useful framework for the theoretical development of the role of cognition in presence.

Acknowledgements

I would like to acknowledge the support of the CAVES (Collaborative African Virtual Environments System) project in providing the means for conducting this research.

References

Discrimination Among Three Types of Anomalies in a Visual-Haptic Virtual Environment

Miriam Reiner, Gad Halevy, David Hecht, Moram Furman, Daniel Vainsencher
Department of Education in Science and Technology
{miriamr@tx.technion.ac.il, gadh@tx.technion.ac.il, davidhecht@beethoven.com, moranfurman@yahoo.com, danielv@tx.technion.ac.il}

Abstract
We performed a pilot experiment aimed to study behavioral and cognitive responses to different types of anomalies in a virtual world of haptic objects and 3D visual stimuli. Subjects’ reports and behavior during the experiment indicate that the virtual world that had been constructed creates in most subjects a high degree of presence in the VR. For some of the anomalies that were included in the pilot experiment, the behavioral responses gave clear indications of their effectiveness. For others, the pilot experiment suggested modifications and variations for the following phases of the experiment.

Keywords--- Presence, Haptics, Anomalies, BIPs, EEG

1. Introduction
Unlike in the field of vision, there still lacks a theoretical framework of how haptic experiences are integrated into an internal cognitive model of what is perceived [1-2]. The present experiment is part of a study aimed to explore the role of haptics sensory feedback in the context of presence, especially in relation to visual sensations, and bodily action.

It has been shown in [3] that interruptions of the VR illusion (termed ‘Breaks In Presence’ - BIP) lowered the reported measures of presence. It is also suggested in [4] that BIPs also affect some physiological measurements of subjects. The authors of the latter paper noted that the results reported are not conclusive, and one possible reason is the use of relatively insensitive biological measurements (GSR and heart rate).

The present paper describes a pilot experiment that is part of a larger experiment aimed to extend the above work in three directions:

- Study of physiological and cognitive responses to different types of anomalies in virtual-environments.
- VE that include haptic feedback in addition to the visual cues.
- Measurements of a more sensitive physiological indicator, specifically EEG, as proposed in [5].

We defined some anomalies that fall into three rough categories:
1. A Break In Presence - BIP,
2. An inconsistency between the senses (what is seen and what is felt)
3. A surprise in the virtual world (an unexpected event which can happen and would be a surprise in the real world as well).

Because we assume that part of what makes a multimodal VE immersive is the combination of the different sensory modalities, we expected an inconsistency to be experienced as a BIP.

The pilot experiment was aimed at a number of objectives which are prerequisite for the main experiment:

(1) To test the functioning of the virtual-world in real time, and its interaction with experimental subjects, including: haptic channel (PHANToM), visual channel (3D display), and recording of behavioral data.

(2) To study qualitatively, using behavioral responses and interviews, the ability of our novel virtual world to create a state of presence in subjects.

(3) To test if the suggested examples of anomalies are effective in general, and to study subjects’ spontaneous behavioral responses to them.

We present some tentative conclusions about the effects of the different anomalies, and some conjectures on why some anomalies lack effect.

2. Methods
2.1. Experimental world

The ReachIn environment was used to design and develop a new ‘world’ that allows 3D realistic, visual-haptic-synchronized-interaction. The virtual space visible to the viewer was simple: a box having a transparent front, roof, but no bottom. In it, are floating six small cubes.
The setting of the experiment is a game, which is affected by the following features of the world:

- A small gravitation-like force slowly pulls the cubes down and out the bottom of the screen.
- A small wall (vertical square) was placed at the bottom of the visual space, and served as the target.
- The subjects used the PHANToM to push the cubes. The cubes resisted pushes, so they seem quite massive. The goal of the game is to hit the target once with each cube.
- Subjects’ achievement in hitting the target was graded. The score goes up with each cube that hits, and down as time passes and no hits happen. The game ends when the subject presses a small panel in the back of the box. They do this after all cubes had either touched the target or left the screen.

The task as described above is a baseline, and in addition to it, the subjects also experienced several modified versions. These modified versions of the game included one of six following different anomalies.

1. Inconsistencies between visual and haptic cues:
   - All cubes become invisible (but can still be touched and give haptic feedback) for a short period - 10 seconds.
   - One of the cubes passes through the target instead of colliding with it.

2. Surprises - an unexpected haptic experience:
   - One of the boxes is heavier than the others.
   - One of the boxes, when touched, starts to roll around its own axis in addition to moving (otherwise the cubes in the game are always aligned – see figure 1).

3. Breaks-in-presence (BIPs)
   - A personal message appears on screen, and all cubes move quickly down. The game stops suddenly for 2 seconds, and a white screen is displayed.

Figure 1 shows a snapshot of the VR visual field during one of the experiments.

2.2. Experimental methodology

Eight subjects (5 males, 3 females, ages 23-28) took part in the experiments. Six subjects were naive about the purpose of the experiment. Before the experiment, subjects were asked to fill a preliminary questionnaire aimed mainly to assess their level of experience with computers and VRs. Subjects were introduced with the equipment and instructed to interact with the VE. During the experiment, subjects wore 3D goggles and used the PHANToM in their dominant hand. At each experimental session, subject tried to win scores by hitting the cubes at the target within the VR. Each of the anomalies happened at some point during the games.

During the experiment, the system documented various events along the time axis: the subject touching the boxes, times of occurrence of anomalies, etc. In addition, during the experiments, special physical and verbal responses of the subjects were documented. Following the experiment, an interview was held. The interview included: report of special events; subjective feeling of the degree of presence in the VR; BIPs; and feedback on the functioning of the system. We note that the use of questionnaires to assess presence has been critiqued in [6]. We consider the questionnaire to complement the recorded physical and verbal responses during the experiment, in producing a coarse indication of presence. This is sufficient for the purposes of the current study.

3. Results

3.1. Reported degree of presence, and spontaneous responses during the experiment

As shown in Figure 2, most subjects reported a high degree of presence in the VR.
Figure 2: Reported degrees of presence during the experiment.

Another evidence for the high degree of presence is the wealth of spontaneous responses (verbal and physical) during the experiment. The responses included: Expressions of joy and pleasure (“oh! that’s nice”, “cool”, laughing, etc.); Expressions of frustration (“oh, my score is running down”); Expressions of surprise (“what’s going on here?”); Physical responses (use of the free hand); Talking to the cubes.

3.2. Behavioral responses during the experiment

For each subject, we recorded the times of touching the boxes with the stylus. Figure 3 shows the times of touching boxes for each of the subjects, in an experimental situation where all boxes seemed to disappear, but could be felt haptically between t=10 and t=20 sec. Each rectangle in the figure represents time of touching a box by the corresponding subject. As can be seen, most subjects stopped interacting with the boxes when they were not seen, assuming they were not active any more.

Figure 3: Box touching times during an experiment where boxes seemed to disappear for 10 seconds.

3.3. Responses to the different types of anomalies

The types of responses to each type of anomaly in the VR are summarized as follows:

(1) Unexpected haptic experience (an unexpected event, yet consistent with previous sensory experiences in the real physical world): In accordance with our previous expectations, subjects did not report experiencing a BIP, only surprise.

(2) Conflict between visual and haptic cues (a situation that stands in contrast to previous sensory experiences in the physical reality): In contrast to our expectations, subjects did not report experiencing a BIP, only surprise.

(3) Clashes between virtual world and the real physical world (an intrusion of the external physical reality into the virtual world): In accordance with our previous expectations, in many occasions subjects experienced a BIP (“I became aware of the physical reality when that beep sounded. I thought – what is this noise?”)

4. Discussion

Our pilot experiment indicates that the virtual world described here creates in most subjects a high degree of presence. For example, one of the subjects said: “I couldn’t connect the boundaries of the virtual world to the real world. Sometimes I thought that my hand is pretty close to the reflecting mirror, and then I realized I was actually far from it. In a sense, the virtual space was more real for me than the physical space.”

For some of the six anomalies that were included in the pilot experiment, the behavioral responses gave clear indications of their effectiveness. For others, the pilot experiment suggested modifications and variations for the following phases of the experiment. The main observations from the pilot experiment are the following:

(1) During unexpected haptic experience subjects experienced surprise, but not a BIP.

(2) During conflicts between visual and haptic cues subjects did not report experiencing a BIP, only surprise.

For the anomaly of cubes that are temporarily invisible, this may be explained if we consider that most participants stopped touching the cubes, and therefore didn't experience their haptic effect while they were invisible. Here the inconsistency depends on the users’ reaction – it is only perceived if the user continues to explore haptically. Also, we have noticed that the anomaly, in which one cube passes through the target instead of bouncing off it, is sometimes not reported by the subjects. For this anomaly, it is possible that the subjects are not noticing the difference at all, perhaps because the altered behavior is not at the focus of the interaction.

In order to study, in the main experiment, whether there is a difference in physiological responses between this and the other classes of anomaly, we will design more reliable anomalies of this class.

(3) During clashes between virtual world and physical reality, in many occasions subjects experienced a BIP.

References


Augmented Reality for Museum Exhibits.

Rocio Ruiz Rodarte
Instituto Tecnológico y de Estudios Superiores de Monterrey ITESM-CEM
{caruiz@itesm.mx}

Abstract
This paper’s proposal is the use of Augmented Reality (AR) as a museographic tool for superimposing digital reproductions of archaeological elements, over physical remains as they are actually exhibited. The AR installation is part of a museum project under current development, intended to portray an ancient Maya city which ruled the area for 1200 years before it was abandoned and left to be concealed for centuries in the rain forest until its recent discovery. The museum installation is meant to provide a sense of being in the presence of the past by digitally superimposing no longer existing archaeological relics. The aim of the project is to show museum visitors two perspectives of a Mayan funerary chamber: the original aspect as Maya inhabitants buried their governor in the 8th century and its present aspect as the archaeologists found it in 1995. This paper also describes different issues addressed during the implementation of the project: the three dimensional digital modeling of the Maya burial, the enhancement of its rendering looks, the AR software implementation, the lightning requirements for the site and some issues related to immersive devices.

Keywords--- archaeology, virtual museography, augmented reality.

1. Introduction
Augmented Reality has successfully joined all the advantages that virtual environments bring to those that the real world already has, thus generating one of the most immersive applications. The opportunity of combining both real and virtual objects in the users’ sight allows researchers to play with one specific aspect of the displayed scene: its temporality. Museographers and designers can easily implement this feature to enhance regular museums’ exhibits.

Museum visitors usually struggle with the inquisitiveness of how things were before destruction or even during looting. By superimposing virtual objects to physical mock-ups, the visitors’ experience can be completed by adding virtual reproductions of all the no longer existing elements. These elements keep their position and orientation while the visitors stare at the rest of the installation, increasing the realism of the scene and its feeling of presence.

This project is part of the virtual museography currently developed for a site museum at an archaeological Maya Ancient City called Calakmul, listed by UNESCO as a Cultural Humanity Heritage Site.

Figure 1. Aerial view of Calakmul
This archaeological site is located deep inside a protected biosphere in Mexico in the Yucatan Peninsula only 18 miles away from the Guatemalan border.

Figure 2. Yucatan’s Peninsula, Mexico.
Calakmul’s isolation is part of its magnetism and yet, due to this solitude, not one of its treasures can be left on site for exhibition. To display the invaluable jade masks and funerary relics, a wide selection of digital and Virtual Reality products is proposed to be used as means of local exhibition. With these three dimensional reproductions, world visitors will be able to appreciate its rich features regardless of their language or previous knowledge.

Amongst the 17 tombs discovered so far in Calakmul, the tomb of king "Jaguar's Claw” is one of its
main findings due to the many remains that guided the archaeologists to assume how the burial was originally settled. This particular tomb has been enhanced with an AR installation to provide the visitors with two different views of the same situation: the actual condition of the funerary chamber and a virtual recreation of its hypothetical layout back in the 8th century. The visitors will be able to see through a flat display or through I-glasses, how the burial mummy was supposedly laid out together with its many offerings.

Different issues addressed by the research team during the development of this installation are described:

- Three dimensional modeling of the burial
- Addition of “shadowed” surfaces to increase realistic depiction.
- Design of patterns for the recognition algorithm
- Other uses for Past/Present AR overlay

2. Data Acquisition and Modeling

The archaeologist who found and excavated the burial has guided the digital reconstruction of king "Jaguar's Claw" tomb. From his written reports and drafts, several alternative reproductions were modeled from the burial, its ceramic offerings and the jade jewelry, leading to the achievement of his research hypothesis.

![Figure 3. Archaeologist notes from burial](image)

To model and texturize the ceramics that laid in the burial, a complete set of 360º photographs of each ceramic vase, placed on a rotating table, was taken. The shots obtained by a digital camera set on a tripod, were cleaned and blended together until a whole roll-out of the design of each vase was obtained. This procedure can be partially automated by QTVR-object software but human assistance is much more accurate. The final flattened images were saved in jpg format, which is needed for the final AR scene.

Modeling and texturizing processes were both done with 3Dmax software and exported to VRML format. Saving scenes in VRML format allows the use of the same models for both Augmented Reality applications and interactive VRML kiosks. In fact, every individual vase of the AR funerary reproduction is an artifact exhibited amongst many others in virtual showcases (PC stands) for the museum and are part of the museum 3D digital catalog.

![Figure 4. Burial at museum & reproduction](image)

2.1. Shadow Rendering

VRML format cannot render real-time shadows. To avoid this problem without the need of adding shadow volume algorithms or any other sophisticated procedure, fake shadowed surfaces were added to the three dimensional model. This could be considered as cheating, however, it is done to increase the realism without increasing the difficulty that these applications could mean for museographers and exhibit designers.

The first thing to do was to place enough omni lights with a reduced multiplier value all over the 3D scene. This action makes all textures and objects visible. Lights on one side of the scene should have a higher intensity to increase the volume and to indicate a source of lightning.

![Figure 5. “Shadow” surfaces.](image)
After defining this source, distorted surfaces simulating shadowed areas are placed beneath the objects that could produce shadows.

These surfaces are mapped with textures made with the same materials used for the rest of the scene but retouched on Photoshop software with a reduced brightness and an increased contrast.

3. Implementation

The installation uses ARToolKit software [1] developed by Phd. Hirokazu Kato from the University of Hiroshima City and Ph.D. Mark Billinghurst at the Human Interface Technology Laboratory, University of Washington, Seattle, Washington, USA.

ARToolKit software, developed in C language, calculates the position and orientation of a webcam relative to a pattern inside a square of specific proportions. The input information from the environment is provided to a computer by a simple webcam. The video data is then converted to a high contrast image. Kato’s algorithm seeks on the high contrast video input, for squares of certain proportions and, in doing so, it identifies patterns inside such squares. It also identifies the orientation that this pattern has in relation to the original orientation saved for it. Once the pattern, its relative position and its orientation are identified, a computer graphic is placed on top of it with the same position and orientation.

The software uses a very quick pattern recognition algorithm that lets the user perceive the virtual graphic stuck to the square even though it is moved or spun. As long as the pattern is framed by a square of established proportions and it stays complete on the users view field, the object will keep visible and movable.

The first prototype of the burial created for the exhibition, superimposes a whole virtual scene to an empty physical replica of the chamber’s walls. Visitors can stare to a flat monitor as looking through a window or use I-glasses to have a deeper presence feeling.

This first AR installation, recreates the entire burial with all its offerings as a sole 3D object, to be rendered in an empty physical reproduction of the original funerary chamber. Since museum visitors will be placed in different positions around the installation, three different patterns call the same object (whole burial) at different orientations.

For example, the pattern at the left side of the physical installation calls a burial whose origin is placed at the head of the mummy; while a pattern placed at the middle of the installation, calls the mummy’s model whose origin is placed at its body’s waist.

Multiple figure 5 shows the same 3D mummy with different origin and orientation depending on the location of the pattern that calls it. The 3D scene is the same one but it is saved as three different VRML objects assigned to three different patterns.

Having more virtual objects than real ones, it dealt with an initial logistic problem for the museum. This issue serves to explain here a concept related to how Augmented Reality shifts between Virtual Reality on one side and Reality on the other.

This first prototype is closer to a Virtual Reality scene and could be called Augmented Virtuality [2].

The ARToolKit software could render as many 3D objects as patterns recognized in the same scenario. As soon as physical replicas are done by the museum and a more accurate mock-up is settled, the installation will have several patterns all over it to place individual virtual archaeological artifacts. The combination of less virtual objects with more real replicas will turn the installation from Augmented Virtuality into Augmented Reality.
3.1. Pattern Designs

Since the AR software, turns the whole image provided by the webcam, into high contrast images, it provides the opportunity to play with colors that are equivalent to black and white when converted to high contrast tones, for elaborating designs that could work for the pattern recognition algorithm without affecting the museographic looks.

The burial itself has no drawings or hieroglyphs where patterns need to be placed, even though, something that could work as a pattern must be placed in the same visitor’s view field. To reach that requirement, the museum’s logo is used as a pattern. Its colors were slightly adjusted so they could blend into a black & white square whenever the software converts them to high contrast images.

![Figure 8. Tones equivalence into high contrast](image)

Since the AR pattern recognition software struggles with luminosity changes, the capture of the same patterns with different illumination is a solution to avoid pattern recognition problems at several daylight conditions, lamp luminescence temperature, etc.

3.2. Further uses for Past/Present AR overlay.

There are many opportunities to overlay the present and the past in museum’s exhibitions. AR is a technology that allows manipulation in both directions, superimposing past to present or vice versa.

There is a second example for this idea currently under development. A physical mock-up of the ancient city of Tenochtitlan was made superimposing virtual reproductions of present buildings on the mock-up. In this case, the physical part of the installations deals with the past while the overlapping virtual objects represent the present. The installation is a didactic tool that allows students to understand how the actual city of Mexico was built over the remains of the Aztec capital.

In this case, very little modeling is used. Instead of detailed geometries, retouched photographs of actual buildings are mapped to 3D boxes. Several photos from the same facade are taken from an orthogonal position and with no perspective warp. The 3D boxes representing Mexico City’s actual buildings are called by de ARToolKit over each of the patterns placed on a Tenochtitlan mock-up. Each pattern appears as a number related to a list of names from downtown streets.

![Figure 10. Building recreations superimposed to mock-up.](image)

Conclusions

This paper describes the use of Augmented Reality technology as a means to present the concept of temporality in museums. This technology enables a realistic experience that complements what time has destroyed.

ARToolKit software allows museographers to implement these exhibits in an easy way and low costs that can be easily adjusted to regular museums’ budgets.

We have successfully combined virtual reproductions from the museums’ digital 3D catalog, for AR installations and VRML PC Kiosks. This activity has let us optimized the digitizing effort from archaeological findings and turned them into didactic devices for museum visitors.

These didactic devices, complemented with I-glasses, will provide museum’s visitors with a unique presence feeling of being in the face of the Past.

Acknowledgements

The project is partially supported by the National Institute of Anthropology and History (INAH in Spanish by its abbreviation initials), Mexico: National Coordination of Museums, INAH, the INAH-Campeche Center and the Archaeological Project of Calakmul.

References

Presence in Educational and Entertaining Virtual Environments

Sacau, Ana, Coutinho Susana
University Fernando Pessoa
{pssacau@ufp.pt, susanacoutinho@tvtel.pt}

Abstract
This paper analyse the sensation of Presence in virtual environments developed for educational and entertainment goals. Based in a user-centred point of view we built a questionnaire supported in the two core constructs of absorption and significance. Results showed a strong correlation between absorption and significance and significant differences between both kinds of environments. Significant higher means were found for the educational environment. No significant differences were found between woman and man as a whole sample but we found gender differences for the entertainment environment where the avatar was female. We didn’t find any relationship between Presence scores and age.

Keywords— Presence, absorption, significance, virtual environment.

1. Introduction

The concept of Presence had been studied from several points of view including engineering, psychology and communication.

Even many authors comes from social sciences field, a literature review shows that the research about Presence still presents a strong technological approach. “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” (Sas and O’Hare, 2003, p.527) [1]. The basic design still consists of the manipulation of some technological variables and the examination of their effects on Presence. Studies about field of view (see for example, Arthur, 2000 [2]; Dillon et al, 2002 [3]; Duh et al, 2002 [4]; Hendrix, 1994 [5]; Ijsselsteijn et al, 1998 [6]; Kim and Biocca, 1997 [7]; Lin et al, 2002 [8]; Prothero and Hoffman, 1995 [9]) or viewing conditions (mostly stereoscopically versus monoscopically viewing) (see for example Arthur et al, 1997 [10]; Freeman et al, 1999 [11]; Ijsselsteijn et al 2001 [12]) are quite extensive but only one study examines, for example, the cognitive style of participants. From our point of view, this technological approach presents an important limitation, that is, to understand the human being as a passive user of technology. “Whether reading a book or using a very sophisticated virtual reality immersive system, the human being is an active performer” (Sacau, 2004) [13]. Nevertheless, only few researches take into account the assessment of the psychological variables related with Presence as a main objective of their works.

One of these studies was conducted by Sas and O’Hare (2003)[1]. These authors evaluated the relationship between some cognitive factors with presence assessed by a questionnaire. Their experimental design included the exploration of a virtual environment and a searching task. The linear regression performed has shown that presence was a linear function of the creative imagination, absorption and a subscale of empathy named fantasy (Beta= .463; .270 and .232, respectively).

This study was developed from an exclusive psychological point of view using a quantitative methodology. The authors didn’t change or manipulate any technological factor.

Other studies used a more qualitative methodology (Freeman and Avons, 2000 [14]; Rétaux, 2002 [15]). Freeman and Avons (2000) employed a focus group methodology to explore presence related with immersive TV. They concluded, “some comments suggested that participants felt more present in scenes that they would have been interested in, regardless the presentation mode” (Freeman and Avons, 2000, online). Applying a direct method of questioning participants, Rétaux also said that participant’s answers revealed “it is the ability of the game to match their goals that allows them to feel like they are in the game” (Rétaux, 2002, p.163).

Some other individual factors analysed are spatial ability (Witmer and Singer, 1998) [16], prior training (Freeman et al, 1999 [11]; Rétaux, 2002 [15]) and emotion (Dillon et al, 2002 [3]; Chapman et al, 2003 [17]).

This paper tries to defend a user-centred approach for the study of the sensation of Presence. We focus our study of Presence in two psychological constructs: Absorption and Significance. According to Wild, et al. (1995) [18] we understand absorption as a motivation and a skill. From our point of view, Presence will occurs if the user has the capability (skills) to flow into a different reality from the real one and to reach a different state of consciousness. But
it is also necessary that the user permits (volition) the suspension of disbelief.

Significance reports to the emotional, intellectual and biographical meaning of any experience. From our point of view, the users ask themselves if the experience of a virtual presentation makes sense in terms of content or narration but also in terms of functionality, personal intentions and goals (Sacau, 2004) [13].

2. Method

The main goal of this study is to analyse the sensation of being there into educational and entertainment virtual environments. To achieve this goal we used two experimental groups and we developed a questionnaire of Presence based on a user-centred approach. No technological variables were included in this study. The content of the environment was the only factor we manipulated. Our main point of interest was to explore the psychological factors of Absorption and Significance when subjects are exposed to different contents.

2.1. Participants

60 undergraduate students participated in our study (30 subjects per group). The age mean was 22 years old. The minimum was 17 and the maximum was 31 years old. The sample was counterbalance by gender; 50% of the subjects were females and the other 50% were males. This distribution was also balance between groups.

Students from computer engineering and regular players of videogames were excluded. Before running the experimental sessions with the entertainment environment group, we asked the participants about their knowledge on Tomb Raider games. We only selected subjects who never played any Tomb Raider version.

2.2. Instrument

We developed a questionnaire of Presence based in a user-centred approach. Twenty items, 9 for the construct “Absorption”, 9 for the construct “Significance” and two direct items of Presence compose the questionnaire.

Items from “Absorption” were built up based on the theoretical proposals developed by authors as Lombard and Ditton (1997) [19], Roche and McConkey (1990) [20] and Wild et al. (1995)[18]. Only one of these absorption items was generated using an adaptation of one item from the Tellegen Absorption Scale (TAS) published by Tellegen and Atkinson (1974) [21]. The reason is related with the non-authorisation to translate the TAS to Portuguese. This deprivation of authority to use the TAS in any other language but English compelled us to build our own absorption scale.

The significance items were supported on the works by Dillon et al. (2002) [3], Freeman and Avons (2000) [14], Ijsselsteijn et al. (2000) [22], Osgood et al. (1978) [23], Slater (1999) [24], and Witmer and Singer (1998) [16].

Two items measured Presence in a direct way. These items were articulated based on Slater (1999) [24] who defined Presence as the sensation of being there and as a function of the level in which participants remember the experience like visiting a place and not just like a group of images in a computer.

Global Presence scores were computed adding up the Absorption, the Significance and the two direct measures of Presence scores. This Global Presence measure is designated here as Presence to distinguish it from the two direct measures of Presence.

The questionnaire was originally built in Portuguese. Following we show two examples of items, one from absorption and one from significance and the two direct items of Presence. All of them are exposed in the original language (Portuguese) and their translation in English.

Absorption:
“Senti-me completamente dentro do ambiente virtual”
“I felt completely inside the virtual environment”

Significance:
“O mundo representado no ambiente virtual fez sentido para mim”
“The world represented by the virtual environment made sense for me. “

Directs measures of Presence (from Slater, 1999):
“When estava no ambiente virtual tive a sensação de estar lá”
“When I was in the virtual environment I had the sensation of being there”

“Senti que apenas visualizei um conjunto de imagens”
“I felt I just visualised an assemblage of images”

We used a Likert five-points scaling for answering. We represented the five points by letters to avoid tendency answers like “more is better”. Only the ending points (A and E) of the answer scale were labelled. “A” letter was labelled as “I completely agree” and “E” letter was labelled as “I completely disagree”.

The questionnaire included a set of final questions about gender, age and two control questions “I don’t like games or any other virtual environments” and “I didn’t like this virtual environment”. These two last questions had a dichotomised answer “I agree” and “I disagree”.

We run a pre-test with 8 subjects (four women and four men) to verify the contextual and linguistic suitability of the questionnaire. The pre-test reproduced the main experimental setting in terms of time, location and instructions. The only difference was that we asked the subjects to mention any difficulty or doubt with the questionnaire. After the analysis of the pre-test results we simplified and re-edited two items.
2.3. Apparatus and material

We used a laptop Compaq Presario2500, Intel Pentium IV, 2.8 GHz, 512 MB RAM, 60Gb, a mouse Logitech cordless (for the educational environment), a standard keyboard (for the entertainment environment) and headphones Sony MDR-P180.

The training level permits the player to explore the environment without any danger. Anyone interact with Lara Croft, just her majordomo appears but without any interference. The Lara Croft’s house is a non-interactive environment with two floors with many rooms as a dinning room, a kitchen or a library.

![Figure 1. Mozart’s museum first floor.](image1)

![Figure 2. A room in the 3rd floor.](image2)

The stimuli

The educational virtual environment was a Mozart’s Museum developed by MEC project (IST-2001-37661) and designed as a House of Learning. This virtual environment is an interactive museum dedicated to the famous music composer Wolfgang Amadeus Mozart.

It is a building with four floors where the visitor can learn about the Mozart’s life and compositions. Figures 1 and 2 show two rooms of the museum. Each floor has thematic rooms related with the Mozart’s childhood, his family, his early travels around Europe, and so on. The last floor is dedicated to his masterpieces and his dead. The visitor can hear some of the most well-know pieces including an excerpt of the famous Requiem. The virtual museum is full of interactive objects as a violin or a harpsichord, which plays music, and pictures about Mozart’s life.

The entertainment virtual environment was selected from the game “Tomb Raider III” produced by Core Design Limited and edited by Eidos Interactive Limited.

We selected the training level named Lara Croft’s home. Figures 3 and 4 show two places inside the house. We didn’t use a playing level to prevent the risk of a participant “dies” during the game. Nevertheless, two subjects were removed from the experiment because they died during the exploration: one of them jumps into the fire of the chimney and the other one jumps inside the swimming pool and he didn’t know how to swim and went out.

Before each experimental session it was necessary to deactivate the off voice presented in the training level.

We select two environments built with different purposes (educational and entertainment) but with similar configuration and usability. Anyway, we show in the table 1 the main differences between both environments.

2.3. Procedure

The participants were contacted individually at the University Fernando Pessoa. The purpose and the procedure of the study were explained and the participants were invited to join the experiment. If they accepted they were randomly assigned to one of the experimental groups. Their participation was not paid. After the instructions, the participants explored the virtual environment during seven minutes. Then, they answered the questionnaire. The questionnaire was self-administrated.

The experimental sessions run in different classrooms of the university with the same conditions of light, dimensions and noise.

3. Results

To analyse the data we used SPSS 10.0 for Windows. First we calculate the descriptive statistics for each subscale. Table 2 shows these results.
A first reading of this results point up that the whole sample reveals high levels of Presence, Absorption and Significance.

A Pearson’s correlation coefficient shows (see Table 3) a strong, positive and significant relation between Absorption, Significance and direct measures of Presence. Those subjects who think that the virtual environment was meaningful also felt higher levels of Absorption as well as higher levels of Presence measured by two items.

To analyse the differences on the sense of Presence and their components depending on the content of the environment we apply a t-test for independent samples. The results reveal a significant difference between both environments (see Table 4). Surprisingly, subjects who explore the educational environment reported higher levels of Presence, and Absorption and experienced the environment as more meaningful.

There are two possible reasons for these results. The first explanation is related with novelty. Even we excluded subjects who play the game in any of its versions, Tomb Raider is a very well-known game. Most of young people know the game although they haven’t ever played it or even they don’t like any kind of videogames. Opposing to that, it is possible that none of the participants had seen before a virtual interactive museum.

The second explanation could be related with the amount of information and objects presented in the educational environment.

We didn’t find any study that empirically analyse novelty or amount of information as important factors to enhance Presence but some authors consider that a more enjoyable exploration of a virtual environment could influence the sensation of being there (Lombard and Ditton, 1997) [19].

Maybe the novelty and the amount of information and interactive objects made the exploration of the Mozart’s Museum more enjoyable than an emblematic videogame as Tomb Raider.

A third possibility must be pointed out. Subjects who explore the Lara Croft’s house knew that they would explore a videogame just before the experiment session start.

---

Table 1: Main differences between both virtual environments.

<table>
<thead>
<tr>
<th>Mozart’s Museum</th>
<th>Lara Croft’s House</th>
</tr>
</thead>
<tbody>
<tr>
<td>Without avatar (first person view)</td>
<td>With avatar (third person view)</td>
</tr>
<tr>
<td>Music in all floors. No more sounds than Mozart’s music.</td>
<td>Without music. The only sounds present are the Lara Croft’s steps and the sound of fire in the chimney.</td>
</tr>
<tr>
<td>Interactive objects (instruments play music and doors open when the visitor comes close to them)</td>
<td>No interaction (all doors are opened).</td>
</tr>
<tr>
<td>Exploration using a cordless mouse</td>
<td>Exploration using the keyboard</td>
</tr>
</tbody>
</table>

---

Figure 3. Lara Croft’s sleeping room.  
Figure 4. Lara Croft’s swimming pool.
Table 2. Descriptive statistics of Presence, Absorption, Significance and direct measures of Presence.

<table>
<thead>
<tr>
<th></th>
<th>Presence</th>
<th>Absorption</th>
<th>Significance</th>
<th>Direct measures of Presence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>23.0</td>
<td>10.0</td>
<td>9.0</td>
<td>2.0</td>
</tr>
<tr>
<td>Maximum</td>
<td>72.0</td>
<td>35.0</td>
<td>34.0</td>
<td>8.0</td>
</tr>
<tr>
<td>Mean</td>
<td>44.8</td>
<td>20.1</td>
<td>20.0</td>
<td>3.9</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>11.8</td>
<td>5.1</td>
<td>6.3</td>
<td>1.4</td>
</tr>
</tbody>
</table>

n = 60. Presence amplitude (20-100); Absorption amplitude (9-45); Significance amplitude (9-45); Direct items of Presence (2-10). Lower values indicate higher levels of Presence, Absorption and Significance.

Promptly, they could create expectations about what is supposed to do in a videogame, just game. Then the instructions gave by the researcher asked for them to explore, just explore, the virtual house. The failed expectations could provoke the results we find.

When playing is avoid, Tomb Raider is empty of meaning and the capability to be absorbed by the game is broken.

No significant differences were found between both environments when we analyse direct measures of Presence. Even no statistically significant, the mean values show that participants who explore the educational environment report higher rates of Presence. Presence levels, composed by Absorption, and Significance items plus the two direct measures of Presence, also showed statistical differences. This result was expected once Absorption and Significance items have a great weigh on the final score.

Table 3. Correlation coefficients (r).

<table>
<thead>
<tr>
<th></th>
<th>Absorption</th>
<th>Significance</th>
<th>Direct measures of Presence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absorption</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pearson’s coefficient</td>
<td>1.00</td>
<td>.78</td>
<td>.66</td>
</tr>
<tr>
<td>P</td>
<td>.00</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>60</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>Significance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pearson’s coefficient</td>
<td>1.00</td>
<td>.62</td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>60</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4. Group comparison with t-test.

<table>
<thead>
<tr>
<th></th>
<th>M¹</th>
<th>Sd</th>
<th>t</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absorption</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Educational</td>
<td>18.16</td>
<td>4.35</td>
<td>-3.18</td>
<td>58</td>
<td>.002</td>
</tr>
<tr>
<td>Entertainment</td>
<td>22.06</td>
<td>5.11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Significance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Educational</td>
<td>18.53</td>
<td>6.21</td>
<td>-2.82</td>
<td>58</td>
<td>.007</td>
</tr>
<tr>
<td>Entertainment</td>
<td>22.86</td>
<td>5.71</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Direct measures of Presence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Educational</td>
<td>3.70</td>
<td>1.48</td>
<td>-1.50</td>
<td>58</td>
<td>.139</td>
</tr>
<tr>
<td>Entertainment</td>
<td>4.26</td>
<td>1.43</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Presence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Educational</td>
<td>40.40</td>
<td>11.31</td>
<td>-3.08</td>
<td>58</td>
<td>.003</td>
</tr>
<tr>
<td>Entertainment</td>
<td>49.20</td>
<td>10.80</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

¹ Lower values indicate higher levels.
To control the influence of general interest on videogames on Presence we apply a t-test for independent groups. We differentiated those individuals who reported to be interested in games and other virtual environments from those who didn’t. Most of the participants (61.66%) affirmed to be interested in game and virtual environments. Only 38.33% of the sample answered they were not interested at all. The t-test didn’t showed significant differences between both groups (t=1.07; df=58; p>.05). This results reveals that Presence is not related with the general interest in games and other virtual environments.

A clear view about this result arise when we analyse the second control question “I didn’t like this virtual environment”. 100% of participants affirmed to like the virtual environment explored. This result shows the attraction power of virtual applications even some people are not concern about them.

A main comparison between genders didn’t show any differences on Presence (t=-1.20; df=58; p>.05). Only when we made separated analysis for educational and entertainment groups we found significant differences between male and female who explored the entertainment environments. Women reported higher levels of Presence (t=-2.40; df=28; p=.02). This result is in conformity with Biocca (1997) [25] who refers a possible identification of the user with the avatar depending of its characteristics in terms of gender, clothing, etc. This result seems to confirm that women felt more identified with the female avatar of Tomb Raider. It didn’t happen with the participants who explored the educational environment where users have a first person point of view.

We didn’t find any relation between Presence and age (r=-.141; p>.05). In this point it is important to remember the small range of ages of our sample (min.=17, max. =31).

The statistical quality of the questionnaire is being assessed. The analysis is not finish but preliminary results showed that the questionnaire presents accepted levels of reliability. Internal consistency calculated for the whole questionnaire using Cronbach’s Alpha method was upper that .90. Factorial analysis is still in course.

Conclusions

Research on Presence should pay more attention to the individuals. A bigger effort must be applied to comprehend the psychological factors related with Presence. Individuals are not passive users of technology. Virtual (as real) worlds are full of meaning and users own different psychological characteristics and capabilities. With this paper we try to encourage a user-centred approach for the future research on Presence. Presence could be gathered as a function of human factors acting in response to the technological variables.

This paper describes that Absorption and Significance are dependent on the content of the virtual environment. We tried to measure Presence based on these two psychological factors, the capability of the individual to feel absorption and how they award meaning to their experiences.

Acknowledgements

This research was developed partially under the MEC project supported by the European Union’s FET Presence Research Initiative (IST-2001-37661). The Mozart’s museum was developed by MEC team. We are very grateful for all MEC member’s contribution.

References


Abstract
This paper presents an investigation into how sound can be used to create a sense of place (that is, the sense of being in that particular place). In the study, people were asked to speak aloud and tell us what they could hear. The analysis of these concurrent verbalizations is a particular focus of this work. We also demonstrate the usefulness of this data and offer a novel interpretation based on the work of the phenomenologist Martin Heidegger.

1. Introduction

Qualitative methods, though increasing in popularity, are still not widely used in Presence research. The norm has been to adopt quantitative techniques, the most frequently used of which is the questionnaire. The questionnaire, as an instrument for investigating presence, has a great number of advantages, not least of which they are easy to administer, score and code while being (potentially) valid and accurate. However the use of questionnaires have been the subject to a number of criticisms [1]. Of these, there is one irrefutable objection to their use is that they are necessarily administered after the event. As the event in question is being present (i.e. in a virtual environment) a questionnaire can only elicit impressions and memories of ‘how it was’, rather than ‘how it is’. While this objection does not in any way invalidate a role of questionnaires, there is clearly a place for a more immediate method for determining an individual’s sense of presence while in situ. In response to these difficulties a series of objective measures have and are being developed and these include such things as physiological measurements [2,3].

The approach to a more immediate and situated appreciation of presence presented in this paper is based upon the elicitation and subsequent analysis of verbal protocols, which have been concurrently vocalized during a VR / presence episode. The use of talk-aloud or concurrent vocalization is widely used in other human-computer studies, which are briefly reviewed in section 2. However before this review we wish to highlight our intention to extend their use in two ways. Firstly, we do not wish to confine ourselves to a purely cognitive interpretation by assuming that these verbal protocols are merely the contents of the individual’s working memory [4,5]. Instead we believe (and will demonstrate) that they contain both phenomenological and reflective / cognitive elements. These phenomenological elements will give us insight into what people can see, hear and feel at that time in the virtual world – in short how they are coping with it and what they are telling us about their experiences. While the reflective or cognitive statements might witness the individual making sense of the experience with reference to their prior experience and memories.

We do, of course, recognize that the very process of concurrent vocalization may interact with whatever processes are involved in being present in a virtual environment. Indeed it is reasonable to expect that concurrent vocalization may degrade task performance, particularly if the task involved working memory or linguistic reasoning. However Ericsson and Simon (ibid) have argued that this is not the case and they have reviewed a number of studies, which appeared to support their position. Despite this, there is evidence to the contrary. Russo, Johnson and Stephens [6] have demonstrated that concurrent vocalization does reduce the accuracy of mental arithmetic, which is a typical working memory-intensive task (incidentally it also seems to improve the accuracy of a gambling task). Given the demands of a VR / presence episode we might speculate that the production of VR / presence may actually increase an individual’s sense of presence by increasing the sense of being involved.

The second extension to standard use of verbal protocols concerns their analysis. Typically, as will be seen in the next section, verbal protocols are taken to be indicators of (cognitive) information processing such as: decision making or spatial reasoning. While not denying the value of the human information processing paradigm, we wish to propose and explore the usefulness of a complementary phenomenological interpretation. In many ways this is quite an ambitious undertaking: using, in this instance, Martin Heidegger’s existential-phenomenological philosophy to understand how people cope with and experience a technologically mediated experience is no small matter.

2. Cognition, Coping, and Verbal Protocols

The technique of protocol analysis of verbal data has been employed within cognitive psychology and human computer interaction (HCI) research for almost 30 years. As we have seen, concurrent verbalizations are taken to
be representative of that individual’s cognition, specifically, the contents of the subject's working memory. Perhaps the keyword here is 'cognition' – if an individual is engaged in cognition - then this argument holds, but if they are not, what then? Alternatives to cognition might include – daydreaming, being distracted by an itch, feeling hungry or engaging in a routine activity. Adopting this complementary phenomenological perspective may give us additional insights into the immediate contents of an individual’s consciousness while engaged with virtual reality.

### 2.1 Cognition and Reflection

Conventionally, the elicitation of verbal protocols involves asking a person to articulate (speak aloud) what they are thinking as they perform a task. Thus, collecting verbal protocols reveals both what a person is thinking, and how people are coping with the situation or task. Detienne and Soloway [7] describe the insight as a ‘sort of window onto subjects’ processing strategies’. Part of the appeal of verbal protocols lies with their usefulness in revealing an individual’s problem solving, or coping behaviour [8]. Using the technique, it is possible to infer a participant’s cognitive processes. The requirements of the problem-solving task will direct what information is processed by the subject. Concurrent verbalization articulates that information processing (Ericsson and Simon ibid). As an example, verbalizations made during software debugging could be expected to reveal the various factors considered by the person as potentially contributing to a bug. Similarly, verbalizations made during software maintenance could highlight the hypotheses considered and rejected in the process of formulating an understanding of the software.

### 2.2 A Phenomenological Treatment

Dennett [9] distinguishes between phenomenology, which is the study of phenomena and Phenomenology, which refers to a family of the philosophical schools of thought. The former is concerned with describing experiences as they appear in consciousness, without recourse to explanation, theory, or other assumptions. Phenomenological psychology, for example, is concerned with the study of personal experience, and subjective perception of phenomena rather than ‘objective truths’.

Dennett suggests that these phenomena can be grouped into (1) experiences of the ‘external’ world such as sights and sounds; (2) experiences of the ‘internal’ world such as daydreams, talking to oneself and (3) affect – pains, hunger, and emotional responses such as surprise or desire. While these very different experiences could be divided and partitioned in a dozen different ways we can (probably) agree that they are direct (i.e. unmediated), subjective, personal and qualitative in nature.

Phenomenology may also be defined as the interpretive study of human experience, the aim of which is to examine and clarify human situations, events, meanings, and experiences “as they spontaneously occur in the course of daily life” (as noted on page 3 [10]). The goal of phenomenology is “a rigorous description of human life as it is lived [ … ] in all of its first-person concreteness, urgency, and ambiguity” (page 5, [11]).

Phenomenology is also the name of a number of different but overlapping schools of philosophical thought. These schools of phenomenological thought can be broadly divided into two traditions, namely those which draw upon the writings and thoughts of Edmund Husserl (transcendental phenomenology) and of Martin Heidegger (existential phenomenology) respectively. Since our interest here is existential in nature, Heidegger’s work is the more relevant of the two. We take two things from Heidegger and his later commentators and phenomenological researchers: (i) everyday, practical coping (ii) Heidegger’s treatment of language from the perspective of telling.

### 2.3 Everyday, Practical Coping

By practical coping, Dreyfus (a philosopher and influential commentator on Heidegger) means the mostly smooth and unobtrusive responsiveness to circumstances that enables human beings to get around in the world. Everyday, practical coping is not necessarily cognitive and may only invoke cognition when some form of breakdown occurs. The scope of practical coping extends from the mundane such as using a knife and fork, sitting working at a desk, to the highly skilled such grandmaster chess or writing an academic paper. Tools and devices are central to coping and these are often used as a backdrop to other activities: talking on a mobile phone while irritating the other passengers on a train; holding a conversation while tying a shoe-lace; typing a letter while drinking a cup of coffee and a thousand other combinations – including wearing an HMD while concurrently vocalizing.

Dreyfus clarifies Heidegger’s basic theses as:

1. People have skills for coping with equipment, other people, and themselves;
2. Our shared everyday coping practices conform to norms;
3. The interrelated totality of equipment, norms and social roles form a whole which Heidegger calls “significance.”
4. Significance is the basis of average intelligibility, and
5. This average intelligibility can be further articulated in language. As Heidegger puts it “We have the same thing in view, because it is in the same averageness that we have a common understanding of what is said” (Being & Time 212).

---

1 We are confident that many philosophers will disagree with this rather simple treatment of a complex topic.
While the language of Heidegger and Dreyfus may be a little unfamiliar, what is apparent is that our everyday coping with the world, tools and technology is the basis by which we make sense of those things we are using. We share a common understanding with respect to tools and equipment and this intelligibility can be articulated in language.

2.4 Heidegger, Language and Telling

Heidegger [12] has noted “It is language that tells us about the nature of a thing …”. In the same essay, Heidegger, demonstrates at great length, how language, and the use of language affords insights into the nature of being-in-the-world. However for the purposes of this discussion we now turn to his treatment of telling. The reason for this is captured in the following trio of cryptic-sounding observations which Heidegger offers in his Being and Time (abbreviated to BT), “ ‘The … foundation of language is discourse or talk’ and ‘Discourse is existentially equiprimodial with state-of-mind and understanding’ and finally, ‘The way in which discourse gets expressed is language’ (BT 161).

Drawing heavily on Dreyfus’ commentary on Being and Time to understand these assertions – we note that Heidegger argues that while talking is the foundation of language, by talking he actually means telling. He uses the German word Rede, which Dreyfus prefers to translate as ‘telling’. ‘Telling’ should be understood as in the expression ‘to tell the time’, or what a bank teller does or being able to tell that an image is upside down or a surgeon telling apart different kinds of tissue. All of which indicate that ‘telling’ is not necessarily linguistic, instead telling is about picking out and pointing out significations in the world which is then manifest as language. Or as Heidegger puts it himself, “Discoursing or talking is the way in which we articulate significantly the intelligibility of being-in-the-world” – linking us back to everyday practical coping.

The next point is further revealing, Heidegger argues that discourse (telling) does not precede or follow state-of-mind or understanding it occurs concurrently (it is equiprimodial). To illustrate this, the following fairly long quotation from Heidegger should help:

“What we “first” hear it is never noise or complexes of sounds but the creaking wagon, the motor-cycle. We hear the column on the march, the north wind, the woodpecker tapping, the fire cracking. It requires a very artificial and complicated frame of mind to ‘hear’ a ‘pure noise’. The fact that motor-cycles and wagons are what we as being-in-the-world already dwells alongside what is ready-to-hand within-the-world; it certainly does not dwell proximally alongside ‘sensations’; nor would it first have to be given shape…” - BT 164.

Heidegger writes in much the same vein on the issue of vision and visual perception.

In summary, Heidegger argues that much of our being-in-the-world can be characterized by everyday practical coping (and cognitively mediated behavior when required). The processes of coping reveal or disclose the nature, structure and significances of the world. These significances can also be articulated by means of language, which is built upon telling which occurs equiprimodial with understanding.

So if people concurrently vocalize while enjoying a technologically created world what can they tell us about this world?

3. Listening to people

For the purposes of this study we have sought to investigate the extent to which we can create a sense of place using a soundfield alone. This study continues the work we have reported elsewhere [13] into the role of sound in recreating real places.

3.1 Capturing and recreating a soundfield

A custom eight-channel digital audio recording/playback system, was utilized in order to reproduce the central computing lab at Napier University (Jack Kilby Computer Centre) during a typical afternoon.

Figure 1 - Capturing the sound of the Jack Kilby Computer Centre

The recording involved eight identical omni-directional tie-clip microphones, with subsequent speaker positioning matching the microphones in both floor position & height (figures 1 & 2). These were positioned into an ellipse at approximately average ear-height when seated, in order to emulate the majority of the inhabitants’ positions. Omni-directional microphones were chosen in order to maximize any natural reflections as well as to ensure that nothing was “off-axis” such as is the case of directional microphones.

The recording was made in a single thirty-minute pass onto eight separate channels, a separate eight channel microphone pre-amp was used to minimize distortion and ensure consistency in both dynamics and frequency. Each channel was recorded at 96kHz and 24 bits, which gave us an theoretical dynamic range of 144 dB ensuring that the full audible range was covered. The high sampling rate meant that not only could ultrasonic frequencies be recorded, ensuring that associated phase cancellation could be reproduced, but also that the short
time delays, with an accuracy of circa fifteen microseconds, that we rely on in order to accurately locate sounds could be reproduced, something which is not possible at the standard CD sampling rate of 44.1 kHz.

Calibration between the physical soundscape and its subsequent reproduction was achieved utilizing a sound pressure level (spl) meter. The meter was set to the C scale and recorded an average of c.48dBC, the A scale would have rolled off too much bass, whereas the C scale more accurately represents the acoustic energy present during the recording. For reproduction eight compact monitors were supplemented by four sub bass units, whilst bass transmission can normally be considered omni-directional, the low spl levels made accurate positioning of low frequency sounds, such as people walking on hollow resonant floors, difficult. The use of four sub bass units solved this problem, achieving a more accurate representation, than that normally associated with a 5.1 or 7.1 system, where the sub bass unit is normally located in front of the listener. This also compensated for the reduced frequency transmission range associated with compact monitors. When participants were describing the virtual soundfield they were recorded using a standard stereo tie-clip microphone onto a DAT set to 48kHz 16 bit, this allowed an accurate stereo image in order to emulate the participant’s listening experience with reference to their own voice.

3.2 Participants

Forty participants were invited to participate in the study and were randomly assigned to one of the four conditions. The study was conducted over a period of two consecutive weeks. The participants varied with respect to their age, sex and background. All participants took part in the study on a voluntary basis, and all were required to have a high command of spoken English.

3.3 Experimental Conditions

The study had four conditions with 10 participants randomly assigned to each condition:

<table>
<thead>
<tr>
<th>Condition</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Participants were physically present in the Jack Kilby Computer Centre (JKCC) while being asked to speak aloud what they could hear for 15 minutes.</td>
</tr>
<tr>
<td>2</td>
<td>As condition 1 but with the participants having been blindfolded.</td>
</tr>
<tr>
<td>3</td>
<td>Participants were exposed to a recreated soundfield of the JKCC for 15 minutes. They were asked to describe speak aloud what they were hearing.</td>
</tr>
<tr>
<td>4</td>
<td>As condition 3 but with the participants having been blindfolded.</td>
</tr>
</tbody>
</table>

In order to prevent the responses from the physical environment being merely a reflection of what participants could see and therefore interpret, half of the respondents were blindfolded. In a similar manner half of the participants who experienced the recording were also blindfolded to prevent the knowledge of the loudspeaker positions affecting their responses. This allowed us an insight into whether knowledge of the reproduction system would affect the responses, which it subsequently did not.

3.4 Procedure

For all conditions a stereo microphone was attached to the collar of each participant in order to record what they said aloud. A Sony DAT Walkman™ was used to
record their words. The participants were told that the task would last approximately fifteen minutes and that they could ask any questions afterwards.

For conditions 3 and 4 the participants were guided into a room and seated at a table where they were asked to listen to the recording and describe what they could hear for 15 minutes. They were also told that they could end the experiment at any point. For condition 4 they were blindfolded and unaware that they would be seated in the midst of eight speakers and four sub bass units (see figure 2). After fifteen minutes, they were guided back out of the room and their blindfold was removed. At no stage during the experiment, could the participant see the room and its contents.

4. Data Analysis

The recorded verbalizations were transcribed and read. The transcribed text files were then analyzed using ATLAS/ti². Perhaps the most striking finding was the degree of individual differences within each condition, for example, here are two extracts from two participants:

00:00:06 to 00:00:17
Voices ... paper crunching ... more voices...

00:00:24 to 00:00:26
Maybe the wind...

00:00:49 to 00:00:51
No changes still voices

00:01:14 to 00:01:16
Someone is coughing in the background

00:01:26 to 00:01:27
More people coughing ...

00:01:41 to 00:01:46
Its like sitting in some kind of... railway station or something

This protocol is fairly laconic, mainly phenomenological (‘paper crunching’) intermixed with more interpretative statements ‘kind of... railway station’. In contrast, the second extract is much more cognitive or reflective.

00:00:00 to 00:00:29
It sounds like a station it sounds like Waverley station¹ ... although it could be a corridor ... Amm ... fell a bit lost sitting still when everyone else is moving around and getting on with their stuff ...

00:00:49 to 00:01:00

Now I'm trying to listen to what the people behind me are saying ... amm ... and it feels strange 'coz they're talking in another language and it feels like I should be listening ... some sort of buzzer...

00:01:08 to 00:01:14
Its like a tape sound. ticking that's probably the tape ...

00:01:21 to 00:01:47
that sounds like somebody pressing keys, typing ... coughing ... that sounds like I'm in an office... it just feels pretty normal ... like you ... like you were sitting in an office... always worrying when you hear someone laughing and you don't know what's its about

00:01:53 to 00:02:08
sounds like somebody opening curtains or blinds... or even taking photos... which is a bit disconcerting

These two participants both commented on what they were hearing (voices, typing, laughter) and occasionally produced an interpretation of those sounds (like a railway station or a corridor, ticking that’s probably the tape) very much in line with what we expected. These verbal protocols are not just the contents of peoples’ working memories but reflect what they are hearing and how they feel which is what they are telling us about.

To make sense of these data we have chosen to identify themes. This standard qualitative approach has been used successfully both by the authors and others ([14, 15]). To this end the protocols were re-read independently by the authors until we were reasonably confident that we had identified a number of recurrent themes. At this point we compared notes and eschewing the use of a statistical test⁴ to quantify the level of agreement agreed a set of themes. In all forty-five themes were identified, we present six of them by way of illustration.

4.1 Theme 1: the environment

A consistent and persistent theme in many of the verbal protocols is that the real and recreated JKCC is a noisy environment. These phenomena correspond to Dennett’s first category, namely, experiences of the ‘external’ world such as sights and sounds. However we note that individuals’ vocalizations are a mixture of the phenomenological and the more reflective.

²The software supports qualitative content analysis of text, images and audio material, in particular the selection, coding annotating and comparison of segments of raw data. A semantic network editor allows the building and modification of theoretical models.

³The central railway station in Edinburgh.

⁴The use of Cohen’s κ to quantify the level of agreement is occasionally used in these circumstances but the test is not without its critics. Given that our intention is to illustrate the usefulness of the technique of concurrent vocalization and its analysis rather than to make claims about the experiment itself we decided against applying the test.

PRESENCE 2004
You can tell it's a large room because a lot of the sounds are very far away.

You can tell that I am sitting at the end of the hall because the noise is coming from the left hand side and at the right hand side there is a wall.

I have a feeling it's quite a large space... as well... am I don't know if that's to with the any kind of echoic or not but it feels quite eerly and perhaps its because I know where I am... so I've got visual pictures to go with the sounds...

4.2 Theme 2: Reflecting on the experience

There was also abundant evidence of people stopping and reflecting on the experience of listening to and concurrently verbalizing the experience too.

Am I'm trying to think whether the sounds are distracting or not... ammhh its not too bad at the moment some times it is really nippy if you have got a couple of people... am sitting next to you and they are just not intended to do any work at all... they are just sort of chatting away and that's very distracting... but at the moment most of the sound talking... sounds quite purposeful... am it doesn't sound to much like people trying to avoid work... so it's a... I wouldn't find it particularly distracting...

I would be fairly content to work in this environment... it's not a... not distracting it has got a sort of buzz which doesn't pull me away from what I'm thinking about.

things are getting noisier well it seems that way anyway must be getting used to it or something...

an interesting experience...very much like what a blind person must feel when they get moved into an environment that they are unfamiliar with and they are trying to grasp at every cue that they can to make sense of it.

Not only have I been spared any shrieking or anything like that... I have not heard any mobile phones yet.

If I were in a room and this was happening around me, I would be looking at something else to do... I'm getting bored... whatever that is going on doesn't seem to involve me... I would be looking around to either get out or look for something to read... or somebody to talk to...

4.3 Theme 3: My body

While a number of commentators – particularly Maurice Merleau-Ponty [16] – have stressed the importance of the body in experiencing the world, we found relatively few examples.

Constant tapping of my feet as well

I hear myself talking out loud... I hear myself sniffing...

Quite often, I can hear my own clicking mouse.

Maybe it's somebody stapling things

Scrunching noise behind me... is moving from left to right... somebody... crisp packet...

4.4 Theme 4: Clicking, tapping and ringing

Classically phenomenological in character, our participants remarked on these minor intrusions frequently.

Mobile phone ringing... or something like that

Somebody closing a bag or something like that... making sound.

Maybe it's somebody stapling things

Scrunching noise behind me... is moving from left to right... somebody... crisp packet...
4.5 Theme 5: Sex

Heidegger spoke of us being thrown into the world, that is, our inability not to be involved. We are constantly sense-making and telling apart male from female.

P 2: Grace - 2:54 (34:34)
a female voice is coming from the right ... behind but quite close ...

P 2: Grace - 2:55 (34:34)
that was the Gus right in front of me ... still speaking constantly ...

P 5: Matt - 5:11 (16:16)
now I hear somebody talking over there I can't really ... to the right of me ... I can't really hear what she's saying ... a girl talking ...

P 7: Nigel - 7:68 (80:80)
Cough same guy

P10: Stephan - 10:28 (18:18)
a phone going off... she's just answered it ... girl speaking a foreign language... to the left of me ... she keeps talking... she still talking ...

P11: Alice - 11:44 (98:98)
more female laughter there...

P18: Michael - 18:18 (42:42)
I can hear two girls talking behind and to my... right and I can all most make out the words... and so I am trying... am what I am saying is... am ... it's because I can almost pick out the words that I'm unconsciously trying to listen to them

P29: Orlando - 29:38 (46:46)
a girlie laughter... someone reasonably attracted to someone else... hope they can't hear me ... and there is the returned laughter... classic flirtation

4.6 Theme 6: “What was that?”

This theme captures the experience of telling that something had happened but not being quite sure what.

P 7: Nigel - 7:13 (20:20)
Somebody setting something down on a desk...

P11: Alice - 11:40 (86:86)
I am still baffled by the background sounds

P12: Eileen - 12:12 (25:25)
That sounds like something dropped on the floor

P16: Spike - 16:57 (105:105)

Somebody has just dropped something on the left and again

P22: Aspel - 22:16 (43:43)
Ugh ... don't know what that noise was.

P31: Ben - 31:35 (48:48)
Don't know what that noise is

P32: Douglas - 32:29 (69:69)
And there was something falling down

P34: Usha - 34:40 (64:64)
something clanking shut...

5. Discussion

This paper set itself a number of different tasks. The first was to introduce the concept of the concurrent verbalization (CV) an established technique within the computing community, and to extend its use beyond the reflective / cognitive to include the phenomenological. Concentrating on Heidegger’s treatment of telling allows participants to disclose the nature, structure and significances of their auditory world in real time. When subsequently time-coded the speak-alouds allow cross-referencing with the original sound events illustrating how participants cope with and make sense of the virtual world, throughout the experience.

Examples of qualitative thematic analysis were illustrated resulting in forty-five separate themes from forty participants. What can be shown, is that there is engagement on lots of different levels, considerably more than just cognition. These recurrent themes allow participants to tell us what they hear and how they feel about it, providing us with key factors which are required in order to recreate effectively a sense of place.

6. Further work

The next step is to compare the results with work that the second author is currently conducting with professional audio practitioners. Designers of virtual audio are predominately concerned with the technical aspects of dynamics, spectrum and so on, whereas listeners are trying to make sense of their soundscape. This method could be refined by either giving the listeners tasks such as “please identify how many people are in the room” or asked specific questions about where a specific sound source is within the environment.

Additional work also needs to be conducted about the nature of where the participants experience the sound sources as emanating from. In this experiment participants listening to the recording would find it emanating mostly from behind them, while others would experience it mostly coming from the front or left and so on. This was despite the system being calibrated and the seating position being identical. This effect has been realized for some time within the film industry but further work is required in order to evaluate the effects within VR [17].
Further study should also be made into the effect of the participant’s voice when speaking having different acoustic effects than the recorded soundfield. This can be rectified by passing the participant’s voice through an appropriate reverberation unit in order to recreate the effect of speaking in the environment under study. This should partially eliminate this mismatch in a manner similar to hand replication when using data gloves within VR.

7. References

Insideness And Outsideness: Characterizing The Experiences Of Real And Virtual Places

Phil Turner and Susan Turner
School of Computing, Napier University
{p.turner, s.turner}@napier.ac.uk

Abstract
An alternate approach to characterizing and describing experiences of both real and virtually-recreated places is described both in theoretical and empirical terms.

1. Introduction
This paper is an attempt at tackling the evaluation challenge posed by using virtual reality to create (or more strictly, recreate) real places. While the very many existing evaluation instruments are able to elicit sense of presence none fully explore the complementary question ‘is the virtual place you are experiencing like the real place it represents?’ The nub of the difficulty lies with the word like in this question. Is looking like the real place sufficient – probably not, and must it sound like it (probably) or smell like it (possibly). As humans our experience of a place is complex and multifaceted making the questions asked of those who are exposed to our virtual simulacrum extraordinarily difficult. To illuminate this problem we begin by introducing the context of this work before reviewing the adequacy (or otherwise) of the literature on place and sense-of-place before abandoning most of it in favour of a fresh formulation of the problem based on existential-phenomenological thought. We ground this philosophical perspective in two empirical studies.

2. Recreating And Evaluating Real Places
BENOGO is part of the Future and Emerging Technology (FET) programme funded by the European Union (2002-2005), intended to identify key VR research areas. The BENOGO project is developing novel technology while in parallel defining an application for that technology. The technology is concerned with real-time image based rendering while the application area is ‘being there without going’ – to quote the project’s by-line. Rather than visiting a place in person, individuals will be able to experience a specific place, say, the interior of Notre Dame in Paris by donning a head mounted display (HMD) through which photorealistic images will be experienced. Early experiments with the BENOGO technology have hinted at what might be achieved but also highlight the very real challenges of trying to recreate real places convincingly.

In the first of these studies a photo-realistic virtual representation of a glasshouse in the Prague botanical gardens was recreated. Participants were able to experience a 360° panorama of the interior of the tropical glasshouse via a head-mounted display. A second study used a photo-realistic representation of a spacious stairway and landing at the Technical University of Prague, again presented by a head-mounted display. This was augmented by a computer-generated table on the landing behind which participants sat on at real table and chair. The audio environment was silent save for the augmented sound of breaking glass. Participants were asked to imagine themselves as a security guard on the stairs and to report any untoward events to security control room staff (played by a member of the project team).

Experiments have continued in this vein for the first two years of the project, the most recent of which is the recreation of a viewpoint on the city of Prague itself – more of which later.

While there is a range of instruments which might be brought to bear on the evaluation of these virtual environments, none is able to address what we felt the central question of – “do you feel as though you are in the botanics / staircase etc?”. Perhaps this is unsurprising as much of the focus of VR / presence evaluation has been used on the elusive question of whether one feels present in a virtual environment (rather than its surroundings) not whether one feels present in a specified place.

3. The Specificity Of Places
It is a truism to note that real places are specific. My office looks like a million others but is also specific and necessarily different from all others. Its contents are different – it has my books and papers in it; a GPS (global positioning system) measurement would reveal that it is located at a unique longitude and latitude; I am, in some way or other, attached to it in a way which is different from all the other offices in the corridor. It is specific. A presence questionnaire (or any other evaluation instrument) administered to me in my office might tell us to what degree I might feel present in it. If I
were to experience a virtual reality re-creation of my office and them complete a presence questionnaire, we might reflect on the similarities and differences between the two but neither would say anything about whether it feels like my office. The logical conclusion is that a unique instrument would need to be created for each specific place. These problems of capturing specific experience by generically applicable tools and techniques, of course, are not unknown to geographers, sociologist and real world place researchers.

3.1 Empirical studies of place

There is a large collection of place studies to be found in sociology, tourism studies and environmental psychology, among other disciplines. We have space here for only a few representative examples.

In a study typical of the genre, Jaakson [1] interviewed 300 Canadian cottage owners, with the intention of understanding what their ‘recreation homes’ (holiday cottages) meant to them. After analyzing his data, Jaakson was able to identify ten ‘broad themes of meaning’ or narratives. These were:

- routine versus novelty
- continuity and sense of place
- back-to-nature
- work
- identity
- elitism among (fellow) cottagers
- surety
- aspirations
- inversion of everyday life
- time/distance away from ordinary life

All very interesting but very specific to Canadian holiday home owners. In a later study Kaltenborn [2] investigated the meanings of recreation homes to a group of Norwegian cottage owners. In this instance he undertook a questionnaire survey and attempted to operationalises vague terms ‘place attachment’ and ‘place attributes’. A factor analysis of these data revealed that location turned out to be most important factor in accounting for place attachment. Again even using a less qualitative approach, the specificity of places is a major problem. Knowing that location is an important factor in holiday home really is of little (let’s be honest, no value) to VR / presence researchers.

Environmental psychologists too have been much occupied by sense of place. Canter’s work, culminating in his ‘facet theory’ [3] for example, takes in activities (the facet of functional differentiation), physical characteristics (aspects of design), individual, social and cultural experience (place objectives) and the scale of the place (scale of interaction), while Sixsmith [4] identifies personal, social and physical dimensions in a study of the meanings of ‘home’. Jorgensen and Stedman [5] afford a more recent illustration of the empirical approach in this domain. The authors propose that the interpretation of sense of place could benefit from treatment as an attitude, thus drawing on classical psychological theory. Just as any other attitude, sense of place then has cognitive, affective and conative (or behavioural) components. Supporting data was obtained from a questionnaire survey of over 200 owners of rural second homes. It is suggested that the attitudinal components of sense of place are: (a) beliefs about the relationship between self and place (the cognitive component, labeled by the authors as place identity); (b) feelings towards the place (the affective component, place attachment) and (c) behavioral exclusivity of the place compared with alternatives (the conative component, place dependence). Drawing on the data, Jorgensen and Stedman developed a 12-point sense of place questionnaire scale embodying these dimensions.

While place researchers have done their best to both quantify and qualify the nature of place and our experiences of it, there remain too many problems with it for our purposes.

First of all, in many ways the discussion of place is often at a very generic level. To take two more examples, Relph [6] famously categorizes place identity in terms of three categories – physical attributes, affect and activity. Useful for a humanist, phenomenologically-inspired geographer perhaps but not for us. The other extreme is exemplified by Gustavson [7] who has characterized place with three axes and thirteen attributes.

Faced with this ‘Goldilocks’ problem we must find a level of description which is neither too detailed and specific, nor too high level and vague.

4. Space And Spatiality

We experience real places first as children, learning about Cartesian (and other) spaces later. Piaget’s ‘genetic epistemology’ which was based on ideas drawn from biology, continental philosophy an by the occasional observation of children, places the first stage of our intellectual development in our own hands – so to speak. The sensori-motor stage lasting from birth to our second year is characterized by our ability to differentiate ourselves from other objects. Our experience of space – spatiality, he argues underpins the whole of our intellectual development.

As the child grows, passing through the pre-operational (2 – 7 years), the concrete operational (7 – 11 years), the maturing child reaches the formal operational at which she can think logically about abstract propositions such as space, and their relationship with it – including one supposes, presence. We being by experiencing the world first hand – internalizing those experiences – and then move iteratively from the concrete to the abstract, from the material to the cerebral.

We are not the only animals which experience space. The idea of animal territories was first introduced by Eliot Howard in 1920 but was further popularized by Robert Ardrey with, The Territorial Imperative. He writes, "A territory is an area of space, whether of water or earth or air, which an animal or group of animals defends as an exclusive preserve. The word is also used
to describe the inward compulsion in animate beings to possess and defend such a space.”

Spatiality is a reality for all living beings but for the philosopher Martin Heidegger became a major focus of his work.

4.1 Heidegger on “in”

Heidegger is probably most famous for his assertion of our being-in-the-world. We are in the world and not separate from it; we dwell in the world but our understanding of this hinges on the pivotal word – in. In does not refer to containment in, it might better be understood as involved in or absorbed in. Writing this I am in work, engaged in writing, involved in trying to explain Heidegger. Coyne develops this idea in his Technoromanticism [8] when he writes, “This is an experience with which everyone can identify: we are often most absorbed while working in the garden, driving, jogging, watching television, typing at a computer, or just engaged in the day-to-day routine of work. (In this respect, “immersion” is not a special feature of VR. A text editor, CAD system, computer game, or spreadsheet can be just as much an immersive environment.”) To be spatial is to be engaged, involved. Heidegger invokes a variety of etymological arguments (or tricks) to demonstrate what he describes as the primordiality of in as meaning involvement. He writes:

“in” is derived from “innan” – “to reside”, “habitate”, “to dwell”. “An” signifies “I am accustomed”, “I am familiar with”, ‘I look after something’ … The expression ‘bin’ is connected with ‘bei’, and so ‘ich bin’ [‘I am’] means in its turn ‘I reside’ or ‘dwell alongside’ the world which is familiar to me in such and such a way. “Being” [Sein], as the infinitive of ‘ich bin’ (that is to say, when it is understood as an existentiale), signifies ‘to reside alongside …’ ‘to be familiar with … ’ Being-in is thus the formal existential expression for the Being of Dasein’, which has Being-in-the-world as it essential state.

Dasein’s way of being-in consists in dwelling or residing, that is, being ‘alongside’ the world as if it were at home there.

4.2 Insideness and outsideness

One promising starting point for applying Heidegger’s insights is Relph’s dimension of insideness-outsideness [6], a concept which in turn draws upon Heidegger’s writings on in and ‘dwelling’, for example, For an individual to feel inside a place is to feel at home and safe. To experience outsideness is to experience the reverse. Relph notes insideness or outsideness is experienced at differing levels of intensity. He uses Berger’s characterization [9] of the anthropologist’s immersion in an alien culture as a means of labeling these:

- **Behavioral insideness**: taking part in activities local to the culture, but as a detached observer
- **Empathetic insideness**: emotional as well as behavioural involvement, but remaining aware that one is not a full member of the culture
- **Existential**: ‘going native’. (Relph’s term - Berger’s original is cognitive).

Relph extends these distinctions to seven discrete states of insideness or outsideness, which we considered might affords a systematic means of describing and understanding people’s response to place and its recreation in virtual space. They are briefly described below, drawing on the commentary on Relph’s work by Seamon [10]. Each is illustrated by an example from a possible future application domain for BENOGO and similar technologies, the recreation of fragile historical or archaeological sites. Here we take the specific example of visiting a cave such as Lascaux with its ancient artwork.

**Existential insideness**
The feeling of attachment and being at home. The place is experienced without conscious reflection but is full of significances.

The sense of deep attachment and familiarity experienced by an archaeologist who has spent many years studying the cave and its paintings, and immersing herself in the culture which gave rise to them. Unlikely to be experienced by the casual visitor.

**Existential outsideness**
Feeling ‘out of place’. The place may feel alienating, unreal, or unpleasant. As a visitor, feeling so remote from the culture that originally inhabited the cave that the experience has little or no meaning. In extreme cases, perhaps feeling menaced by echoes of ancient ritual or by claustrophobia.

**Objective outsideness**
A deliberate, dispassionate separation from place. Place is a thing to be studied. Exemplified by researchers studying, for example, how the physical contours of the rock have been incorporated in the paintings to give depth to the scenes depicted, or by tourism planners monitoring visitor movements in the cave complex.

**Incidental outsideness**
Place is experienced as the mere background for activities. Difficult to produce a credible example in this specific instance, but the rather unlikely scenario of attending a conference dinner in the cave gives a flavour of incidental insideness. (The stereotypical case here is that of attending a business meeting in a chain hotel in a foreign city.)

**Behavioural insideness**
This type of insideness involves deliberate attention to the way a place ‘works’, and taking part in the activities specific to it. The experience of an archeologist learning the layout of the cave, the whereabouts of the most compelling artworks, how they are thought to have been used in ritual, or participation in the re-enactment of such rituals.

---

1 Heidegger’s word for human beings.
Empathetic insideness

Here a person is trying to be open to a place and to understand it better. The necessary pre-conditions are a genuine interest or degree of care for the place. Visiting the cave, attempting to shut out the 21st century and experience the cave as its original inhabitants may have done, coupled with a deep appreciation of its unique features and concern for its preservation.

Vicarious insideness

Deeply-felt second-hand involvement with place, transported through imagination – through paintings, novels, music, film, VR and other creative media. Experiencing a successful VR recreation of the cave. Approximates to a strong sense of presence.

It should be noted that the states are neither completely orthogonal – people may experience more than one mode of insideness or outsideness – nor do they represent a scale of insideness with an ordered progression from existential outsideness to existential insideness. They are simply different facets of the human relationship to place.

The success of place-related VR applications is predicated upon the evocation of different aspects of insideness, depending on underlying purpose. We can see this by considering just a couple of instances from differing applications. In one of our previous projects, maritime employers were enthusiastic about an application which would allow personnel to learn the geography of a new ship before setting foot on board – here a sense of behavioural insideness is important, but empathetic insideness rather less so. However, for applications which foster a holistic experience of place – such as the virtual Lascaux suggested above – complete success rests on the fostering of empathetic insideness, probably supported by vicarious insideness.

5. Two Exploratory Studies

As we have observed, Relph’s states of insideness afford a convenient structure for the discussion of the experience of place. We now turn to two studies where we used this to analyse the experience of two different real places and the virtual counterpart of one of them. In each case, the seven states were used to code free-form accounts obtained from visitors to the place concerned.

5.1 Study I: The Kettlewell visitors’ book

The first study served to test the practicality of the using insideness as a coding framework as well as exploring the nature of place experience. While on holiday in a cottage in Kettlewell, Yorkshire, we took the opportunity to photocopy the visitors’ book2 we found there. The visitors’ book is a standard A5 notebook with the instructions: Visitors Please add any suitable comment. The entries run from 23rd April 1994 until 13th October 2001. There are over 250 handwritten entries ranging in size from 2 lines to more than 3 pages. The entries are signed usually in the form Mr. and Mrs. Smith, Edinburgh, Scotland. Figures 1 gives a flavour of the contents of the book. Quite simply, people have used their entries to describe their experience of a week or two’s stay.

We have analyzed the first 65 entries, having first satisfied ourselves that their tone and content were consistent with the remainder of the book. They were coded using Relph’s and Seamon’s descriptions of the different states of insideness as a reference. Each entry was coded with as many codes as the evidence of the contents and tone of writing supported: thus many entries were labeled with multiple codes.

Figure 1 – A pair of pages filled with 6 of the shorter entries.

Coding using the literature as reference proved to be a simple and relatively rapid task. The results are shown graphically below at figure 2. The great majority of the entries (91%) displayed strong evidence of behavioural insideness, commenting (often at length) on such activities as walking, eating and drinking in local pubs, playing in the stream and feeding the ducks. Below is part of a typical entry:

...we managed a six-mile walk everyday. The best was voted as being from the river at Kettlewell along to Buckdon, then on the ridge of the Dales with views of Littondale to the left and Wharfedale to the right. The walk from the cottage along the Dales Way to Conistone is quite easy... BBQ nearly every night. Sampled the food at the [pub name omitted] but found it was over-rated.

The vast majority of accounts (88% of the total) were imbued with feelings of interest, attachment, and fondness for the cottage and its surroundings and thus were coded for empathetic insideness. For example:

2 It has not been practical to contact the authors of the entries in the visitors’ book so we have been careful not to reproduce any details which might lead to their embarrassment, identification, or cause them to pursue us through the courts.
We particularly enjoyed the walking and being in close touch with this beautiful and what is for us very different countryside.

Just a few visitors (9%) expressed a sense of deep familiarity or being at home, or in Relphian terms, existential insideness.

We found no evidence for the remaining four states of insideness, viz. Existential outsideness, objective outsideness, incidental insideness and vicarious insideness. It may be that those people who felt less engaged by their holiday destination were not moved to record their experience for the benefit of others, or that choosing to holiday in a place and the activity of holidaying in itself predispose to engagement or involvement. Incidental insideness would perhaps have possible, if visitors had been using the cottage for some other purpose, e.g. writing a book, but none of our sample appeared to be undertaking such activities. And since our visitors were in the place in reality, there was little opportunity for vicarious experience.

5.2 Study II: Real and virtual Prague

This study uses raw data from the study of a viewpoint in Prague and its virtual counterpart conducted by other members of the BENOGO team, as reported in McCall et al. [11]. The Prague viewpoint study employed a multi-dimensional probe to collect experiential data from people experiencing a viewpoint over the city of Prague and the recreation of the same viewpoint in virtual reality mediated by a head-mounted display (HMD). One element of the probe required participants to contribute a short, free-form written account of their experience: this is the data re-analysed here. There were 21 accounts of the real viewpoint and 29 of its virtual counterpart. As in the Kettlewell study, each account was analyzed and coded by reference to the seven states of insideness, and again many accounts attracted multiple codings.

The results contrast strongly with the Kettlewell study. Only 10% of the accounts of the real viewpoint appeared to describe an experience behavioural insideness, and just 43% were characterized by empathetic insideness. Some 5% displayed evidence of existential insideness. Turning to the remaining codes, just over half (52%) were characterized by objective outsideness, there were indications of existential outsideness in 10% of accounts, and incidental outsideness in 14%, but again no evidence of vicarious insideness. These results are shown graphically in Figure 3 (lighter shaded areas). By way of illustration:

You get a good look of the whole of Prague. You can see St. Vitus Cathedral (objective outsideness)

It felt like I had the power of a king over Prague. I was mighty. (existential insideness)

I come here nearly every day with my dog. Today someone spoke to me... (Incidental outsideness)

Results for the virtual viewpoint show an extremely large proportion of accounts displaying objective

Figure 2: Insideness and Outsideness in Kettlewell

![Graph showing the distribution of insideness and outsideness aspects in Kettlewell study](image-url)
outsideness (93%). Otherwise there are instances of existential outsideness in some 41% (largely relating to the effects of the HMD and problems with resolution and similar flaws in the virtual environment), 31% of exhibit vicarious insideness, 28% empathetic insideness, and 7% behavioural insideness. There were no instances of existential insideness. The results are shown graphically in figure 3 (darker shading). As we have just observed, the intrusion of the VR technology seems to promote a sense of outsideness, and almost all accounts have a tone of dispassionate observation, but nonetheless quite a large proportion of people also experience various forms of insideness. The virtual experience seems to be a somewhat attenuated version of its real counterpart. Below is a sample of existential outsideness coupled with vicarious insideness.

You get the feeling that you really have entered a new world. A bit annoying with the low resolution. I got a bit dizzy in the end.

In comparing the two sets of accounts (real and virtual viewpoints) its appears that the experience of place is of a broadly similar overall nature but rather more likely to be weighted towards feelings of outsideness in the virtual case. The large proportion of instances of objective outsideness in data from the real experience as well as the virtual one can perhaps be explained by the nature of the place – from a viewpoint one is necessarily ‘outside’ the subject of the view, and the task required of the participants – to report their experience.

Figure 3: Insideness and Outsideness in Prague

5.3 Kettlewell and Prague

Classifying the data for both studies by states of insideness has effectively illustrated the very different experiences of place for holidaymakers staying a week or so and the transient visitor to a viewpoint. One would expect that staying in a place for a week or more and interaction with people, animals, landscape and pubs to promote a much stronger sense of insideness in its various guises, and this is reflected in the results. Conversely the experience of a viewpoint, whether real or virtual, is of much shorter duration and affords a great deal less in the way of opportunities for interaction – it is, almost by definition, an overview. Again this can be seen reflected in the much stronger ‘outside’ orientation of these sets of results.

6. Discussion And Further Work

At the outset of this paper we set out to answer two questions, namely, (i) when we experience a virtual recreation of a place, do we feel as though we are actually there in that place and (ii) how can we as designers of virtual experiences deal with the specificity of real places.

We can indeed conclude that the use of the insideness-outsideness dimension does give us an insight into whether or not people experience, in this instance, a virtual viewpoint over Prague in much the same way as a
real viewpoint of Prague (the attenuation of these results not withstanding). Again, the results suggest that this dimension gives us an understanding of the question of being present in this particular place complementary to more traditional questionnaire instruments.

As to the second issue of the specificity of places, we may have in practice actually multiplied the problems in recreating places. From the data, we can conclude that looking down on a place is very different from staying in a specific place for a week or two. Places are not neutral. Elsewhere [10] we have discussed the multimedia and multimodal demands in recreating places; now we must revise this to recognise that our relationship with that place is another critical factor. The tourist, who in the real world might visit twelve cities in twelve days, may actually prefer some form of outsideness (the medium of an air conditioned coach and tinted windows comes to mind) rather than insideness and intimacy of trudging up a hot and dusty path to a ruined temple. We cannot divorce the nature of the virtual recreation of a place from our relationship to it and relationship means much more than the range of interaction with it.

Finally, we need to apply the insideness coding structure to other places of differing types before we can be fully satisfied of its relevance and usefulness. That being done, our next task is to operationalise the definitions of the insideness codes, so that they can be used by others working with similar problems at the intersection of place, presence and virtual reality.

7. References

Mining Emotion in Virtual Environments

Josh Williams, Jacquelyn F. Morie
Institute for Creative Technologies (ICT)
{williamsj@ict.usc.edu, morie@ict.usc.edu}

Abstract

Emotion’s effect on the measurement of presence in virtual environments is not well understood. The use of subjective measures alone in identifying emotional responses to virtual stimuli in a free-will environment is not robust enough to fully evaluate this effect. This paper outlines a design for the objective measure of such emotional responses in terms of arousal states as obtained through the use of physiological, behavioral, and retention instruments. The authors believe this process will augment current methods available to researchers for evaluating presence in affective virtual environments.

Keywords--- Immersive Environments, Virtual Environments, Emotion, Virtual Reality, Evaluation Design, Affective Evaluation

1. Introduction

Presence and emotion are undeniably related, but to what degree? How is presence affected by an emotional valence? Could the evocation of emotion compensate for less immersion in other facets of mediation, or result in breaks-in-presence (BIPs) upon saturation? Perhaps a better question is, “Why are we attempting to get answers on all of these fronts simultaneously?”

The Sensory Environments Evaluation (SEE) Program is investigating a design methodology that emphasizes a feels-real, versus sensory-real, metric of virtual environments (VEs). Towards this end, SEE’s prototype environment, DarkCon, is constructed of multi-sensory stimuli designed to induce emotional responses within its participants. In the interest of preserving the validity of each response, the participant is provided agency within a perceivably uncontained space; free-will mitigated only by the interactive limitations of the technology employed. The resultant lack of strictly commensurate experiences across subject groups in evaluation of this design demands a novel collaboration of instruments in order to compare results. In mining the data obtained by these instruments, the emotional queries in relation to presence stated above may begin to be answered, free of experimental isolation.

2. Designing for Objective Measures of Emotional Response

Participants in the current SEE’s DarkCon evaluation environment are instructed to make their way through an unfamiliar hostile environment at night without being discovered. While there, the subject is also tasked with making a number of observations whose accuracy is purported to affect the metric by which their success will be measured.

Figure 1 Subject in the DarkCon environment

The design principals employed in the creation of the DarkCon environment are discussed in the PRESENCE 2003 paper, “The Gestalt of Virtual Environments” [1]. Those that most directly govern the evaluation of emotional affect in such environments are reviewed in greater detail here.

2.1. Arousal States

The ability to classify emotions (fear, anger, joy, calmness, etc.) experienced by participants in VEs through objective measures is near impossible at present. Sophisticated instruments (such as an fMRI) are required to indicate activity in parts of the brain and produce patterns consistent with known emotional states, but these are costly and not easily integrated within interactive VEs [2]. This limitation has lead many researchers of emotion in VEs to rely on subjective measures to determine emotions experienced (e.g. questionnaires, self-reports). The
shortcoming of these methods is that regardless of reporting consistency, the declared emotions cannot be discreetly attached to the space and time during which they were experienced within the environment. Without this information, iterative design based on the results remains as subjective as the measures employed.

It is, however, relatively easy to determine if a participant is experiencing an arousal state using physiological measures such as skin conductance response (SCR), heart rate (HR), and derived heart rate variability (HRV) [3] [4] [5] [6]. As such arousal states are implicated in emotional response, we have elected to use non-invasive physiological monitoring equipment to ascertain if we are eliciting arousal by way of these two measures. We are most specifically concerned with the temporal and spatial locations within the DarkCon experience designated as emotionally salient.

2.2. Agency

Ochsner and Schacter have shown that goal states can exhibit strong influence over the way emotional experiences are encoded, stored, and retrieved [7]. Thereby, in order to shape an emotionally evocative virtual experience, controlling the goal state of participants is necessary to interpret any emotionally-correlated data. This not only emphasizes the role of the instructional set provided to each subject group, but also requires that participants be afforded the ability to freely act upon such goals. By contrast, if participants are under the impression that their experience is independent of their actions taken during it, their goal sets are rendered obsolete. In such a state, their emotional responses are reactive to a series of mediated stimuli, rather than resultant of interacting with stimuli in a mediated reality.

This degree of agency admittedly prohibits the entirety of the DarkCon experience from being strictly correlated between participants, but the design of environment ensures that each will encounter at least 80% of the same emotionally salient locations. In corroborating this design with the instructional set employed, the participant’s perception of free-will is maintained throughout the experience, ensuring the integrity of data correlated across subject groups at these locations. While the remaining 20% lacks such correlates, it serves to reinforce the degree of agency perceived by each participant given the number of possible actions available at these locations.

3. Instruments

In order to discreetly attach the arousal states obtained (instruments described above) to the temporal and spatial locations at which they are experienced, behavioral data is required for each participant’s experience. Given this alignment, the emotional affect of the designated salient locations can be made referential to all other locations throughout the experience, and further strengthened through the inclusion of recall and retention measures. Lastly, the correlation to presence measures provides a guiding feature for further design.

3.1. Behavioral Recordings

The DarkCon environment is equipped to perform a 6-degrees-of-freedom recording of all the participant’s movements inclusive of the following:
- X and Y coordinates – location
- Z coordinate – altitude, and relative posture (stooping or standing) given location
- Pitch and yaw - head orientation (to ascertain what is viewable to the participant)
- Timestamps – attached to values for all of the above at the frame rate (30fps) of mediation

This data from these recordings can be reinserted into the simulation to drive a full multi-modal replay of the experience. The data can also be scrubbed in text-format by using a custom GUI called “Phloem,” (helpful in expediting the identification of anomalies). Phloem’s most important feature, though, is the ability to query specific behavioral aspects (such as how many times an object was in a subject’s field of view) and export of such queries into an integrated spreadsheet for analysis.

![Figure 2 Phloem GUI](image-url)

3.2 Recall and Retention Measures

It has been demonstrated that the way emotionally charged events are encoded into memory differs from the encoding of emotionally neutral events. The amygdala is strongly implicated in this difference. Both appetitive and aversive stimuli are known activators of amygdala activity. Animal and human studies using pharmacological...
interventions post-training, lesion studies, biochemical studies, and electrophysiological studies all support the assertion that emotional events take a different route to permanent encoding [8].

In the case of the latter, the amygdala regulates the release of stress hormones (e.g. the flight or fight syndrome) to produce immediate responses to stress inducements, thereby aiding in effectively storing such critical memories to be called upon in similar situations in the future.

By this rationale, a participant’s emotional response to specific stimuli in the DarkCon environment should improve retention of these stimuli. To determine the affect of emotional stimuli on retention a prompted self-report is administered immediately following each participant’s experience in the environment, and a referential retention test is administered one week after the experience. The emotional arousal across subjects is then correlated to the retention test results.

3.3. Presence Measures

Three standardized presence questionnaires bookend each participant’s experience of the DarkCon environment: Witmer and Singer’s Immersive Tendencies Questionnaire (ITQ), the same authors’ Presence Questionnaire (PQ) [11], and an extended version of Slater, Usoh, and Steed’s Virtual Environment Questionnaire (VEQ) [12] [6]. Each has been separately employed in numerous studies, and jointly employed in at least Usoh’s study. We have extended questions in the ITQ, PQ, and VEQ (as have most studies) to customize them towards the SEE Program’s goals, while maintaining referential integrity.

By establishing these measures of presence for each participant’s experience, we are able to determine the relation of recorded arousal states and behavior to the induction of presence in the DarkCon environment.

4. Example of Methodology

The following is a simplified diagram of the methodology showing sample data as might be derived in an experiment on emotion containing a single independent variable (the data sets are separated by the emotional stimuli at center to which binary responses are provided for each):

<table>
<thead>
<tr>
<th>Variable</th>
<th>Measures</th>
<th>Key: + Increased/Improved</th>
<th>- Normal</th>
<th>Presence Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>I-SET A</td>
<td>Arousal</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Behavior</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Retention</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Emotinal Stimuli</td>
<td>1: Neutral</td>
<td>2: Salient</td>
<td>3: Salient</td>
<td>4: Neutral</td>
</tr>
<tr>
<td></td>
<td>Arousal</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Behavior</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Retention</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

Results Indicate:
- Notable relation between arousal states, retention, and presence, independent of instructional set
- Notable relation between instructional sets and behavior at emotional stimuli #2 and #3
- Need for further examination of data obtained at emotional stimuli #3 and #4 resulting in inverse arousal and behavior

Figure 4 Example Instructional Set (I-Set) Experiment
Conclusions

We have described a methodology for evaluation that employs both objective and subjective measures in a full-agency virtual environment. The use of these instruments should allow researchers to more objectively address emotion’s effect on the measurement of presence.

It is our hope that this process will facilitate a more thorough exploration of the relationship between emotions and presence in immersive worlds.

Acknowledgements

This paper was developed with funds of the Department of the Army under contract number DAAD 19-99-D-0046. Any opinions, findings and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the Department of the Army.

References

Measuring the Subjective Experience of Presence with Think-Aloud Method: Theory, Instruments, Implications

Werner Wirth, Susanne Wolf, Ursina Mögerle, & Saskia Böcking
University of Zurich, University of Munich, University of Zurich, University of Zurich
{w.wirth@ipmz.unizh.ch, wolf@ifkw.lmu.de, u.moegerle@ipmz.unizh.ch, s.boecking@ipmz.unizh.ch}

Abstract

Presence is a multi-facetted construct and thus can be measured best by applying a multi-method design. The construction of presence is considered to be a cognitive process. Regarding those processes, the method of thinking aloud is fruitful because it provides process-oriented data. Throughout this paper the think-aloud method and its use for presence research is explained. Starting with an explanation of the method and its use in social sciences, we will report the development of a code scheme for analysing individuals’ verbal comments. This code scheme is based on the MEC model of Spatial Presence which serves as theoretical foundation. An exploratory study was conducted to get first impressions of how individuals would address constructs of the MEC model when thinking aloud. Results were then used for the development of a code scheme which allows a quantitative analysis of the think-aloud protocols in presence research. The code scheme developed enables us exactly to analyse e.g. in which way which media characteristics activate the construction of presence.

Keywords--- Spatial Presence, Measurement, Think-aloud method

1. Introduction

Most presence researchers agree on the assumption that presence is a subjective sensation [1]. Consequently, the majority of presence studies use subjective measures such as questionnaires and rating scales to assess presence experiences [2]. Compared to these techniques, the think-aloud method offers not only a more open and explorative, but as well deepened and comprehensive assessment of presence. This method requires participants to verbalize their thoughts, perceptions or feelings during an experience. All comments are recorded and thus can be analysed afterwards. The method is well-established in psychological research on problem solving and is also often used in usability testing, but is relatively unknown in the field of presence research. This is not warrantable, since this technique has considerable power in theory building and provides profound insights in cognitive processes and mental states [3]. The aim of this report is to demonstrate a theoretical well-founded think-aloud approach to the measurement of presence. First, a short introduction will be given about the method of thinking aloud and its use in social sciences. Then, the MEC-Model of Spatial Presence will be presented, which constitutes a theoretical framework for the development of a category system for analysing the think-aloud protocols. Based on an exploratory study1, we explain the concrete development of a code scheme which is intended to be used in further investigations to test the ability of the think-aloud method to assess Spatial Presence.

2. The think-aloud method

The think-aloud method was developed at the beginning of the last century in cognitive psychology by Bühler (1907). It was used until the 1930’s to analyse cognitive processes during problem solving [3]. However, because the collected data were hard to verify, the method was discredited. Only in the 1970’s was it again used for research in educational and media psychology, and in communications, and was further developed.

Ericsson and Simon [2] have published the most detailed and up to date theory on data acquisition with think-aloud. Their theoretical framework is based on the multi-store model of memory: According to Ericsson and Simon, as a person does a task, he or she can report information about subjects’ thinking, a person gets the simple instruction just to talk aloud or think aloud. This instruction can be further refined by telling the subject for example to ‘simply speak aloud anything that comes into mind, never mind if you think it is unimportant or a bad thought’. Non-directive reminding instructions, for example ‘just keep talking’, can be used if subjects fall silent for 15 to 60 seconds.

Basically, two kinds of think-aloud techniques can be distinguished: retrospective and concurrent think-aloud. For concurrent think-aloud, subjects think aloud simultaneously while they are handling a so-called primary task, e.g. problem solving, surfing the WWW or navigating a VR environment. For retrospective think-aloud, a specific situation is e.g. recorded on video and verbalized later.

1 We thank Marianne Promberger for her collaboration in this study.
think-aloud method has to be chosen according to the primary task. The choice can be determined by factors such as the speed of the primary task, which can render simultaneous verbalization impossible, or the presence of other persons. Several studies provide empirical support for Ericsson and Simon’s claims that concurrent and retrospective reports (as defined in their framework; note that different forms of retrospective reports exist [5]) do not disrupt the basic course of thinking [6]. But it is decisive that the retrospective think-aloud takes place immediately after the termination of the primary task.

There is no doubt that the method has its limits, e.g. subjects can only report those thoughts that are expressible in language. Images or feelings have to be translated from nonverbal to verbal code, which can lead to changes and corruption. Furthermore, thinking aloud requires more time than thinking without verbalization. This can lead to a selective verbalization or to a slowing down in the handling of the primary task. Moreover, it is possible that thoughts are not reported because they involve fast, automatic and therefore unconscious processes. Thus, verbal reports can underestimate what is thought during understanding [6] [7] [9]. But think-aloud offers the possibility to obtain process-oriented data which would otherwise be inaccessible. As Crutcher puts it [6], “the advantage of verbal reports as protocol data” lies in “providing a sequence of observations over time rather than just a single observation at the end of a process”. And this should be decisive for the use of think-aloud in presence research as the construction of (spatial) presence is a cognitive process. Which of these processes one can finally assess by applying think-aloud in the end has to be answered empirically (see section 7).

Over the years, the status of verbal report methodologies has improved considerably. Verbal report data is no longer only considered useful in generating hypotheses, but also in evaluating hypotheses or in testing psychological models. The method itself has been improved, and so has the analysis and interpretation of verbal data [6]. Nowadays, analysis and interpretation of verbal data is analogous to quantitative content analysis which is one of the most important research techniques in quantitative social research, respective mass communication research. As Riffe, Lacy and Fico put it [8], "quantitative content analysis is the systematic and replicable examination of symbols of communication, which have been assigned numeric values according to valid measurement rules, and the analysis of relationships involving those values using statistical methods, in order to describe the communication, draw inferences about its meaning, or infer from the communication to its context, both of production and consumption." (p. 20) Many important aspects are mentioned in this definition which show the strengths of the method, primarily replicability and quantification. That means: Findings should not underlie researchers beliefs or hopes as far as the results are concerned, findings should be objective. First, a code scheme with significant categories along which the material should be analysed has to be developed in an iterative process using both theoretical assumptions and empirical findings. Secondly, code scheme, used operations etc. have to be exactly described and fully fixed in a so called “codebook”. Thus, other researchers can read about the procedure and are able to repeat the operations. If there is more than one researcher analysing the material, the coders have to be trained to make sure that all coders understand definitions and treat categories and variables in the same way. Another important thing is to check the reliability, that is, the consistency of codings of individual coders across time or of different coders analysing the same material. As numeric values are accurately assigned to represent each category and variable in the codebook, huge amounts of data can be brought into manageable form. The quantification of data “… permits researchers to assess the representativeness of their samples, and thus use powerful statistical tools to test hypotheses and answer research questions.” (p.26) [8]

3. The use of think-aloud method in social sciences

Think-aloud has turned out to be a useful method in a lot of fields, amongst them educational and clinical psychology or usability testing. There are also a lot of areas especially in media psychology and communications in which the use of think-aloud is fruitful. To mention only a few examples, the method was used to study selection processes while navigating the WWW [9] as well as to investigate psychological processes of TV viewers during reception to explain reception strategies such as frequent channel switching and fragmentary viewing [10]. Furthermore, memory operations involved in text comprehension were analysed using this research technique [11]. The method is also used in a few studies related to VR environments and presence research (at least in its broadest sense). To give insight in how the method was used so far in these fields, a brief overview is presented. Marsh [12] discusses if conventional methods for the evaluation of usability of the conventional two-dimensional Graphical Use Interface (GUI) such as think-aloud are also appropriate for the evaluation of three-dimensional virtual reality systems. Whereas for example Johnson [13] doubted the effectiveness, Marsh refers to results of a preliminary study which seem to indicate that users are able to think aloud while evaluating desktop VR.

Roussos et al. [14] used think-aloud to enlighten learning effects in children using NICE (Narrative, Immersive, Collaborative/Constructivist Environment), a virtual world in which children can grow plants. Think-aloud here was used to examine cognitive conceptual changes and the development of new skills. Presence itself, however, was measured by analysing the children’s behaviour as well as the amount of time it took for the participants to feel comfortable and immersed in the mediated environment.

Another study [15] also deals with learning processes. Hedden examined flow and learning with think-aloud and questionnaires while subjects played computer games. Un-

Note that there are possibilities to increase validity, e.g. a short warming-up exercise [5].
fortunately, as the study is only briefly described on the internet, no more details can be provided concerning the concrete use of think-aloud.

Turner et al. [16] describe their project BENOGO as follows: It “seeks to re-create real places using photorealistic immersive virtual reality technology and in so doing investigates the nature of presence and sense of place in such environments.” (p. 1) In a first experiment, they explored the sense of place in the real world Edinburgh Botanical Gardens and in its virtual equivalent. They tried to find out which elements are inevitable to create a virtual reality which enhances the feeling of sense of place. Participants had to think aloud while exploring the real or the virtual environment. Together with results of structured interviews, Turner et al. could clearly notice missing or wrong elements in their current VR. A few participants even reported experiencing some sense of place (e.g. they wished to pick a leaf).

A further step in BENOGO was to investigate the importance of soundscapes (the term is derived from ‘landscape’) in creating a sense of place in virtual environments [17]. Here, the authors got interesting results as to their experimental procedure e.g. as far as contradictory cues are concerned: “I feel a bit lost sitting still when everyone else is moving around getting on with their stuff…” (p. 4).

Sure enough, by most of these studies using think-aloud presence itself is hardly assessed. The method was rather used e.g. to test for virtual environments and research conditions which enable the construction of presence. But it should have become clear that the method turned out to be useful especially for that the insights revealed by it could not have been made visible with any other method. Therefore, it could be fruitful to apply this method to the measurement of the experience of Spatial Presence.

4. Applying think-aloud method on exploring Presence

Although the think-aloud method is a method used more for explorative research, at least basic theoretical assumptions are necessary for its application. For what the subjects say is normally analysed by applying a code scheme. According to methodological guidelines of content analysis [8] such a code scheme is developed in an iterative process using both the verbal comments of subjects and theoretical considerations. As a theoretical framework we used the MEC model of Spatial Presence [18], verbal comments were received in an exploratory study (see section 5). We now shortly introduce the model and explicate the different constructs.

4.1. The MEC model of Spatial Presence

The MEC model of Spatial Presence has been proposed as a unifying theoretical approach in explicating the development of feeling present in a mediated environment. In integrating both media and user factors it is applicable for not only virtual environments but for less interactive and immersive media settings as well. Moreover, in introducing processes of attention and perception it integrates mental mechanisms which enable humans to feel spatially present. In the following we will provide a short overview over the most important aspects of MEC’s model of Spatial Presence.

According to the MEC model, Spatial Presence is considered to be a specific part of the experience of presence. It arises when media users think that they are put in an environment offered by the medium. The model suggests that Spatial Presence emerges in two steps (cf. figure 1). In the first step, the user constructs a spatial situational model (SSM) of the mediated environment (cf. [19]). Decisive for the construction of the SSM are automatic and controlled attention processes. They result in the user’s fading out of environmental stimuli and his focusing of the mediated environment. The user perceives the spatial cues of the mediated environment and constructs a mental model out of them. It is assumed that individual user characteristics such as spatial ability [20] and absorption [21] influence the form of the SSM. The second step of the model comprises perceptional processes that are based on the SSM and guided by hypotheses [22]. In the course of hypothesis testing the user scrutinizes the assumption that the mediated spatial environment is the primary ego-reference frame [23] [24]. If this assumption is affirmed, according to the model the user feels spatially present. Within this process of hypothesis testing, different user characteristics are considered to be important. The model here concentrates on suspension of disbelief [25] and processes of cognitive involvement [26].
4.2. Explication of Constructs

The experience of Spatial Presence requires attention. Attention in the context of media usage means that stimuli of the non-mediated environment are faded out, whereas stimuli of the mediated environment are focussed. Attention may be the result of the user’s motivation to engage intensively and deeply in and concentrate on the mediated stimulus (controlled attention) or may be caused by novel stimuli or rapid and surprising changes in the stimulus (orienting reactions [27]) and features of the media environment that meet more or less latent needs, motives or interests of the user (automatic attention).

The user’s situational motivational stance is crucial for the focused attention allocation. This motivation may not only depend on situational factors, but also on the user’s general interest in the portrayed topic. Following Krapp [28] we then define domain specific interest as the relationship between an individual’s motivational dispositions and the content or issue of an object.

Having directed its attention towards the media environment, the individual can construct a spatial situational model (SSM) of the depicted space. An SSM is a mental model [29] [30] of the spatial environment which the individual constructs based on (1) spatial cues she/he processes and (2) relevant personal spatial memories and cognitions [19].

Different user characteristics affect the construction and quality of the SSM. Most important is the users’ spatial visual imagery (SVI). This skill belongs conceptually to the construct of spatial ability [31] and is defined as the capability to produce vivid spatial imaginations. It may support the formation of Spatial Presence, as it increases the cognitive salience of spatial structures and makes it more easy to ‘understand’ the spatial quality of the mediated environment, which fosters and accelerates the formation of an SSM.

The existence of an SSM does not automatically guarantee the sensation of Presence. In order to go from an SSM to Spatial Presence, additional cognitive processes have to occur. The model here focuses on suspension of disbelief and processes of cognitive involvement. Cognitive involvement is regarded as higher forms of information processing, i.e. the intense cognitive engagement with a media environment that can be observed in an active and intense processing of the world presented by the media (e.g. appraisal, elaboration, evaluations, and mental explorations) [26]. In contrast, suspension of disbelief is defined as the intentional elimination of external stimuli and internal cognitions that (might) contradict the assumption that the mediated spatial environment is the primary ego-reference frame. It can therefore be understood as an act of volition in which the individual actively pursues the goal of not being distracted from the current experience [32].

Beside these cognitive processes the user trait absorption may take effect during the formation of Spatial Presence experiences by activating suspension of disbelief. Absorption refers to an individual’s motivation and skill in dealing with an object in an elaborate manner [33]. 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. Absorption is therefore considered to be correlated positively with suspension of disbelief and involvement.

Spatial Presence, finally, is defined as the subjective experience of being in the mediated environment. Following considerations to “embodied cognition” [34] [35], it consists of two aspects: (1) The classic description of presence, i.e. the sensation of being physically situated within the spatial environment portrayed by the medium (self-location), and (2) perceived possibilities to act in the mediated environment (possible actions).

Having proposed the theoretical considerations for developing a code scheme for the think-aloud method we now turn to an exploratory study that was conducted to provide impressions which of these theoretical modelled constructs individuals would address when thinking aloud. In doing so, the iterative approach of developing a content scheme was used.

5. Exploratory Study

5.1. Design

10 subjects (six female, four male, aged from 20 to 38, different educational and professional backgrounds) participated in the exploratory study. Participants viewed two films which had carefully been selected and tested in advance. We had to make sure that the stimulus material would enable subjects to think aloud while watching it, and that there would be spatial cues so as to enable the construction of an SSM. Thus, the documentary “Louvre – the visit” and a narrative stimulus “Halloween” were chosen. The non-narrative stimulus was chosen with regard to the “House of Learning” stimulus that will be used in the main study (see section 6); the narrative one was chosen with regard to the MEC project’s goal of finding a way to measure presence in different kinds of media presentations. Both stimuli have strong spatial aspects: The non-narrative stimulus imitates a visitor walking through a museum, the other one for large parts takes the perspective of one of the film’s characters. Both films are suitable for concurrent think-aloud, as the non-narrative stimulus could be shown without the narrator’s voice and the narrative stimulus contains only few dialogues, all of which permit subjects to make remarks while still paying attention to the film. Subjects viewed parts of the two films (about 10 minutes each, seven subjects first viewed the non-narrative film) in a completely dark, sound-proof room at the University of Munich. The films were played using a laptop DVD drive and additional loudspeaker, the picture was projected onto a projection screen. A camera was positioned next to the subject recording the picture on the screen and subjects’ voices. Participants were instructed to think aloud as proposed in section 2 and did a short warming-up exercise. After each study the data were transcribed.
5.2. Results

We now give a short overview which theoretical constructs of the MEC model have been addressed and thus can further be evaluated using think-aloud:

**Attention allocation**: Subjects’ remarks indicate to what they pay attention to and thus make attention allocation manifest. The data show that we can refine the construct ‘attention allocation’ of the MEC model and distinguish between spatial, contextual and action-related attention allocation (see section 6).
- Example for spatial attention allocation: “Now that’s a different building than before, different museum probably.” (participant 5)
- Example for context attention allocation: “Ah, this is a typical US small town.” (participant 2)
- SSM: Subjects think aloud about their associations with certain aspects of the stimulus and try to organize their interpretations into a framework. Therefore, the process of the construction of an SSM can clearly be shown.
  - Example: “Now we’re probably going one floor up and like that we do the entire museum, one room after the other.” (participant 5)

**Suspension of disbelief**: The method makes visible distracting factors [16] [17] which reveal the mediated nature of the experience. But it also seems to become visible when subjects totally forget about such factors and thus suspend disbelief. E.g. rooms and landscapes are no longer seen as parts of a film set, but as ‘real’ rooms or landscapes. Additionally, the data reveal a stage in which subjects are peripherally aware of factors revealing the mediated nature but are not distracted by them (imminent suspension of disbelief).
- Example for disbelief: ”And the child has a dubbing voice which is used all the time…for girls and for boys.” (participant 2)
- Example for imminent suspension of disbelief: “The music shows that something is about to happen.” (participant 4)

**Involvement**: Since all remarks which show attention processes or suspension of disbelief implicate a kind of involvement, only statements revealing higher levels of involvement are assessed.

**Spatial Presence (desired or possible actions)**: Once subjects have constructed an SSM some are able to anticipate actions or movements in the form of specific, short-term predictions and show the desire to act in the mediated environment. In one case, a subject had the feeling of being able to act in the mediated environment.
- Example for desired action: “It would just be nice to be there now, to have time to look at everything really slowly … (camera moves away from female statue) … to stay with her.” (participant 9)
- Example for possible action: “As if I could choose myself where in that building like … or where in that whole area I want to look at something … “ (participant 7)

**Spatial Presence (‘being there’)**: In three cases we found hints for the Spatial Presence experience of ‘being there’.
- Example for ‘being there’: “You can really imagine how that smells down there…the temperature is probably cool.” (participant 6)

It seems, that with think-aloud we can show the different steps by which Spatial Presence is constructed. However, it is decisive to interpret the data very carefully and not to over-interpret the statements of the subjects. One can say, that if the subject is spatially present the person will by definition forget to think aloud. Nonetheless, whether this holds true is a question that has to be answered in (1) using more cases, and (2) combining think-aloud with other methods of measurement to secure validity (see section 7). The exploratory study presented here was intended to prepare such an investigation by allowing to develop a code scheme that could be used in the main investigation (see section 7).

6. Instruments

The purpose of this section is to introduce the conceptional code scheme we developed to be finally applied in the methodological main studies of MEC (see section 7). We developed and executed our scheme on the one hand regarding the theoretical considerations of the MEC model (see section 4.1.) and on the other hand regarding the experiences of the exploratory study (see section 5). In the following, we first present the conceptional development of the code scheme while afterwards the operationalization of our variables to measure Spatial Presence will be explained.

6.1 The conceptional development of the code scheme

In developing the code scheme it was our aim to be able to measure the particular elements of the MEC model of Spatial Presence with this instrument. As code schemes have to fit for the specific stimulus which is intended to be analysed [8], we based the development of our code scheme on a virtual/hypertext museum showing the life of Mozart – the so called House of Learning (HoL). This stimulus will be used in further studies to test the ability of the think-aloud method to measure Spatial Presence (see section 7).

In the HoL, the user is able to navigate through different stages of Mozarts’ life (like childhood, adolescence, etc.) which are presented in different rooms. Because the media environment permanently changes, the feeling of presence depends on the directly available environment. Besides, also the amount of time the participant is engaged in the media environment can enhance the construction of presence experiences [14]. Following the MEC model, different cognitive processes must have taken place before individuals feel present. Thus, it is e.g. assumed that the user feels presence not until after a certain period of time. Consequently, it seems to make sense to consider not only what the user verbalizes but also when he makes the com-
ments. Accordingly, the code scheme is to be composed of three different levels which are to be linked in the analysis. These three levels are (1) the point of time of the verbal comment, (2) the experienced media environment and, last but not least, (3) the verbal comment itself.

In so doing, we are able to analyse (1) when (point of time) the users comment particular aspects of the model, (2) which media characteristics activate which aspects of the model (e.g. music activates attention allocation), and (3) if there are differences – temporally or based on the stimuli – between individuals.

To facilitate the analysis of the levels of the code scheme, the variables of note have to be encoded on the basis of the recorded verbalization of the participants. Consequently, the case unit/definition of the code scheme is based on single verbal comments. Each unit of the code scheme includes the complete operational situation which is maintained during an object-related comment. An object-related comment is defined as a verbal and paraverbal expression which the participant makes during the experimental session concerning objects in the analysed media environment (e.g. certain subjects in or subjects and persons outside the media environment). Thus, it is possible to encode each verbal comment by taking into account the immediate media environment at the same time. The comment not only gives information about the recognizability of the different components of the MEC model and the experiences of the users, but they also enable us to separate the particular expression units from each other.

Besides the just described variables, other variables have been developed in order to measure the construction of Spatial Presence appropriately: these are (1) variables concerning the activities of the investigator (e.g. the number of the participants, the experimental conditions, the date and duration of each experimental session or disturbances during the procedure), (2) variables of the media environment (e.g. room and object reference, available by the users’ recorded navigation through the media environment) and (3) contextual variables (e.g. different forms of disturbances during the experimental sessions as the breakdown of the display screen presentation or external sound sources).

6.2 Operationalization of variables of the MEC model of Spatial Presence

Linking the results of the preparatory study (section 5) with the theoretical definitions of single constructs (section 4.2), the final variables and their categories reflect the MEC model of Spatial Presence.³

As people can only talk about attention processes if they are in some way aware of them, automatic attention processes can not be taken into account when developing the concrete variables and categories. The results of the exploratory study have demonstrated that individuals distinguish between spatial, contextual, and action-related attention allocation. As action-related attention allocation refers to the plot of narrative stimuli and as the virtual Mozart museum is a non-narrative media environment, this aspect of attention has not further been taken into account. Thus, two variables have been formulated to grasp attention processes: spatial attention allocation and attention allocation regarding the media design. While the first variable includes comments that demonstrate people’s attention to spatial aspects of the virtual museum, the second variable comprises attention processes centered to characteristics of objects of the media environment and the virtual museum itself. Spatial attention, again, is distinguished in implicit (i.e., not reflected description of spatial aspects) and explicit (i.e., location of spatial aspects within the general spatial structure of the stimulus) attention. Attention allocation regarding the media design consists of attention focused e.g. on the interactivity of objects within the media presentation.

Regarding the SSM it was differentiated between a rough (i.e., subjects tell that they try to imagine the location of objects and rooms) and a detailed (i.e., subjects show clear picture of the composition of objects and rooms) SSM.

Applying the theoretical dimensions of presence and at the same time the empirical findings, two variables were developed to detect Spatial Presence: the first consists of the three categories anticipated, desired, and possible actions, the second refers to comments of the subject feeling him/ her implicit or explicit physically present.

Besides these process factors, the code book includes user characteristics and actions. The individuals of the preparatory study suspended disturbances of media content and form as well as of the viewing situation. Therefore, several variables were created to assess different forms of suspension of disbelief. The first asks for the existence of technical disturbances (e.g., interfering noises, breakdown of VR, flickering) and in how far the subject has recognized them and is bothered by them. The second aims at disturbances of the realism of the presentation, its credibility regarding the Mozart theme, and perturbances caused by missing interactivity of the virtual museum.

Following the definition of involvement, this variable has been differentiatied in categories containing evaluations, appraisals, and elaborations referring to the media environment (i.e., temporal or associative comments referring to the Mozart’s topic and its matching to the user’s expectations) and to the user’s own life.

The MEC model of Spatial Presence differentiates between a more stable form of motivation, i.e. the domain specific interest, and situational motivation. These two aspects of motivation were included as well in developing the variables general motivation and situational motivation.

The last construct of the model that is taken into account in the code book is visual spatial imagery. It has been operationalized as spatial imagination, i.e. the ability to produce vivid spatial imaginations of depicted scenarios.

³ When reading the descriptions of the different variables and categories, please keep in mind that every variable contains not only the explicated categories, but also a “no”-category. This “no”-category has to be assigned if the subjects does not make any comment about the construct assessed by this variable.
As it is difficult to divide this ability into multiple nuances, only the categories bad and good spatial imagination have been created.

Absorption was not included in the code book as it is a stable user characteristic which can be measured best by applying well-proven personality scales [21].

7. Conclusions and implications

Throughout this paper we have reported the development of a code scheme for analysing the processes which lead to the experience of Spatial Presence. Currently, the codebook presented here is used in investigations comparing different approaches in measuring presence experiences, i.e. think-aloud method and questionnaires. The major aim of these studies is to empirically compare the ability of the two different methods in assessing Spatial Presence—thus testing among others common claims regarding the usefulness of think-aloud in presence research. Although the results of these studies may shed some light on the question which method may assess presence best, both methods are subjective ones. However, presence researchers intensely discuss whether objective measures like physiological ones (e.g. heart rate, skin conductance, skin temperature [2] [36] [37]) or behavioural measures (e.g. postural responses [38]) assess presence better, alike or worse than subjective measures do. Furthermore, it is a sketchy point whether subjective and objective measures assess the same aspects of presence or even the same phenomenon at all. The rationale of this paper is not to replace other measurements of presence with think-aloud method, but to complement the established techniques with this method. What is the advantage of this strategy? First, think-aloud method provides us with exhaustive subjective state-ments about processes and stages accompanying or leading to feelings of presence. If applying the think-aloud method concurrently it gives insight into the dynamics of presence. Second, if using the same stimulus material, the results can be compared to or combined with objective process data like physiological measures of presence, fMRI [38] or behavioural assessments [39]. As Draper, Kaber, and Usher put it: “Correlative studies of objective measures with subjective assessments have not been conducted but would be valuable” [40] [37]. Each of these methods has its disadvantages: Subjective measures are limited to conscious experiences and are susceptible to response sets [41]. Objective measures sometimes are obtrusive. Above all, they do not reveal what subjects think or what the meaning is of the stimuli. Because of these constraints, indeed a joint measurement approach seems to be the best choice.

Acknowledgements

The work presented here was conducted within the Information Society Technologies program of the European Commission (EC), project no. IST-2001-37661, and has been granted by the European Commission and the Federal Office for Education and Science of Switzerland. We thankfully acknowledge the support of the European Commission and the Federal Office for Education and Science of Switzerland.

References

Not a valid question or task.