SOCIAL PRESENCE AND MODE OF VIDEOCOMMUNICATION IN A COLLABORATIVE VIRTUAL ENVIRONMENT

Mohammed Alghamdi, Department of Information Science, University of Otago, New Zealand, algmo560@student.otago.ac.nz

Holger Regenbrecht, Department of Information Science, University of Otago, New Zealand, holger@infoscience.otago.ac.nz

Simon Hoermann, School of Electrical and Information Engineering, The University of Sydney, Sydney, Australia, simon.hoermann@sydney.edu.au

Tobias Langlotz, Department of Information Science, University of Otago, New Zealand, tobias.langlotz@otago.ac.nz

Colin Aldridge, Department of Information Science, University of Otago, New Zealand, cln.aldridge@gmail.com

Abstract

Collaborative Virtual Environments (CVE) with co-located or remote video communication functionality require a continuous experience of social presence. If, at any stage during the experience the communication interrupts presence then the CVE experience as a whole is affected - spatial presence is then decoupled from social presence. We present a solution to this problem by introducing the concept of a virtualized version of Google Glass™ called Virtual Glass. Virtual Glass is integrated into the CVE as a real-world metaphor for a communication device, one particularly suited for collaborative instructor-performer systems. Together with domain experts we developed a prototype system based on an instructor-performer architecture. In two studies with a total number of 115 participants we showed that the concept of Virtual Glass is effective, that it supports a high level of social presence and that the social presence for the performers is rated significantly higher than a standard picture-in-picture videoconferencing approach used for the performers. We present our experimental system, our studies, and the generalizability of our approach towards future uses.

Keywords: Virtual Reality, Human-Computer Interface, Videoconferencing.

1 INTRODUCTION

A Virtual Environment (VE) can be defined as a three-dimensional simulation of a real or imaginary environment generated by a computer (Cobb, Kerr, & Glover, 2001). Collaborative Virtual Environments (CVE) enable people to meet and interact in a shared, virtual space. Using this design, people in remote areas can work together over networks to share tasks and experiences (Park & Kenyon, 1999). In an effective CVE users experience a sense of presence in various aspects, most notably spatial presence – i.e. the sense of being part of the VE – and social presence – i.e. the experience of being together. To support the experience of presence we must design a believable virtual world, where users can interact with the environment and with each other.

For instance, a Virtual Reality Exposure Therapy (VRET) application for phobia treatment (Carlin, Hoffman, & Weghorst, 1997; Juan et al., 2005; Regenbrecht, Wickeroth, Dixon, & Mueller, 2006) is a classic example of a CVE requiring communication capabilities. Here, a client and a therapist meet in a VE. The client navigates the environment and is exposed to varying degrees of fearful stimuli controlled by the therapist. At the same time the therapist is communicating with the client to encourage him/her to face the stimuli/fears and to accept feedback about the client's experience. For the treatment to be effective (North, North, & Coble, 1997), the environment needs to provide believable stimuli (e.g. spiders) as well as offering a continuous, uninterrupted communication channel between therapist and client. Unfortunately, this communication channel is usually decoupled from the CVE. That is, the video/audio communication is not integrated into the VE. Instead, it is appended to the system, often as a separate picture-in-picture (PiP) videoconferencing window. This leads to breaks in the presence experience when the user has to switch from the immersive CVE to be able to communicate via the videoconferencing channel.

With the Virtual Glass concept introduced here we propose an integrated communication solution combining communication and CVE collaboration in a way that maximises social presence. This solution is suitable for a range of types of CVEs, in particular for systems enabling an instructor (e.g. therapist) and performer (e.g. client or patient) to meet in the same virtual space. We refer to these as instructor-performer systems. Here, the instructor maintains an overview of the CVE and, based on events occurring within it, provides instructions and guidance to the performer. The performer remains immersed in the CVE, following the instructor's guidance to respond to these events.

2 RELATED WORK

For our work it is of most interest how two parties can collaborate in the same virtual space and communicate simultaneously without breaks in the experience of being together. We therefore have a closer look at the concept of presence, and how the richness of the media can support this. We also highlight the specifics of the instructor-performer systems targeted in our research, –which are a combination found in a range of application scenarios. We therefore review related work in presence and instructor-performer systems. Later our studies will show the effectiveness of our Virtual Glass implementation in addressing those aspects.

2.1 Presence and Media Richness

A defining component of virtual environments is presence (Steuer, 1992), i.e. a user's experience that her or his body or body parts are perceived as being part of the VE. Presence is achieved by interacting with the environment by self-movement, interaction with objects of the VE and also, in the case of CVE, by the user's interaction and communication with subjects in the VE (Regenbrecht, 1999). The sense of presence in CVE can be decomposed into the factors of spatial presence, involvement, and the realism of the VE (Schubert, Friedmann, & Regenbrecht, 2001), also called the "sense of being there" (Barfield & Weghorst, 1993), combined with social presence (Biocca, Harms, & Gregg, 2001), or the "sense of being together" (De Greef & Ijsselsteijn, 2001). In a wider sense, when considering all kinds of mediated environments, social presence can be defined according to co-presence (colocation and mutual awareness), psychological involvement (e.g. mutual understanding), and behavioural engagement (Biocca et al., 2001).

From a technological point of view it is often desirable to maximise the likelihood that users feel present in the CVE by providing the "right" environment and stimuli. For single user VEs this is achieved by offering a believable plot or narrative in combination with appropriate (affordant) possibilities to interact with the environment (Regenbrecht, 1999). In addition, CVE must also offer means of collaboration (or at least co-operation) and communication. If both aspects are combined we can call this a "sense of being there together". The most basic form of such an environment is where two users interact in the same VE and observe each other's actions in the environment without seeing or hearing the other. This form of indirect co-operation can even be asynchronous, i.e. both users experience the same environment but not at the same time (e.g. https://minecraft.net/). This can be, and often is augmented with text chat possibilities, mainly used for quasi real-time communication.

In systems like cAR/PE! (Regenbrecht et al., 2004) or GAZE-2 (Vertegaal, 1999) CVE functionality is combined with videoconferencing features so that audio-video representations are spatially registered and dynamically aligned within the CVE. Here, instead of virtual avatar heads, video representations operate in a space controlled by the participating users. Video and audio streams are captured, transmitted and displayed in the CVE in a spatially correct way. In addition, the pose (position and orientation) of those video-audio representations are controlled by the users themselves. This gives a sense of co-presence and mediated social presence (Hauber, Regenbrecht, Billinghurst, & Cockburn, 2006). While these video-mediated CVEs solve the so-called breaks-in-presence problem, a continuous bi-directional communication between instructor and performer is not guaranteed. Both users are moving around freely in the VE and only if they meet can they communicate via video.

The above approaches provide communication and collaboration capabilities with varying numbers of channels and degrees of richness. According to Media Richness Theory (Lengel & Daft, 1989) this media richness is determined by the system's ability to simultaneously handle multiple information cues, facilitate rapid and immediate feedback, establish personal focus, and utilize natural language (in a wider sense, including non-verbal communication). There is also a link between social presence and the number of (coherent) communication channels, and the perceived "warmth" and "distance" of the medium (Short, Williams, & Christie, 1976). Media richness is also linked to the effectiveness of the communication. However, for more routine tasks with unequivocal information exchange less rich, leaner media are more suitable. For non-routine tasks with more ambiguous information exchange richer media are more suitable (Lengel & Daft, 1989). Hence, applications such as VRET would require richer media than, for instance, simply committing to a meeting date and time.

Face-to-face communication is considered the gold standard (Hauber et al., 2006; Teoh, Regenbrecht, & O'Hare, 2010) in richness and social presence, followed in order by video conferencing, telephone, and letters. Richness and social presence in CVE can arguably be considered in a similar way, which would give spatially aligned, video-mediated CVEs (Hauber et al., 2006) a higher ranking than text chat supported avatar environments. However, some common forms of CVE communication do not necessarily fit into this ranking, for instance the usual combination of standard video conferencing (e.g. Skype™) with a CVE (e.g. SecondLife™) in a PiP configuration. Here, we do have a large number and richness of channels, but still experience a "breaks-in-presence" issue (Slater et al., 2006) as users are either communicating via the video conferencing channel or are part of the VE. This is of particular concern with VRET applications, and for instructor-performer systems in general, where the performer has to stay focused on the task and environment while simultaneously depending on the (virtual) video-audio presence of the instructor.

2.2 Instructor-Performer Systems

Instructor-performer systems are a special type of multi-user, mixed-reality application where two users normally interact, one being the instructor and the other being the performer, with the instructor always having a "What You See Is What I See" (WYSIWIS) view of the performer. This view can be of a real or VE, captured by a real camera (e.g. head-worn), or a virtual camera, controlled by the performer. Therefore the instructor cannot directly control the performer's view, which is always a self-controlled, first person perspective. Usually, performer and instructor communicate via an audio chat channel. Sometimes other channels are used as well.

Classic examples of instructor-performer systems, as mentioned before, are VRET systems, in which the client and therapist meet in the same room. In our case we adopted this VRET approach for a post-natal stress resilience system. In the future, when implemented as a production system, expectant mothers will be prepared for the stressful situations that typically occur after returning home with their new-born babies. A co-located or remote therapist guides the client through a series of tasks and experiences within a VE specifically designed and implemented for this purpose. The client (woman, performer) is immersed in a virtual therapeutic environment intended to systematically train her to cope with certain stressful situations. The therapist (instructor) and client are in a mode of continuous audio and video communication.

Multi-user, first-person shooter computer or video games might be considered as routine task environments for some people, but still require a high level of support for social presence. By their very nature they demand “flow” (Csikszentmihalyi, 1991), i.e. a certain balance of challenge and skill level. Here, when acting together with other players a true "sense of being there together" is desirable and is supported by certain types and levels of interaction, plot, involvement, and realism, combined with the provision of avatars of many types, appearances, and behaviour. Inter-player communication is realized by text chat, audio channels or PiP videoconferencing. These games can be played in an instructor-performer way, with swapped WYSIWIS views enhanced by audio communication, if players allow for followed over-the-shoulder views for other players. A PiP mode decouples the co-presence (the sense of being in one place) from spatial presence, because the players have to switch between inside (game world, VE) and outside (PiP communication) views.

To solve this decoupling problem, systems have been developed which integrate all users within one VE, including the communication channels. For instance for therapeutic purposes, cMRET (Regenbrecht et al., 2006) uses a video-communication-in-space approach to bring the therapist (instructor) into the environment of the client (performer) leading to the desired effect of a shared experience in one place. In addition, the instructor sees a video (web cam) view of the client for therapeutic observation. Because the client is allowed to freely move around in the VE it can't be guaranteed that therapist and client can see each other all the time - the therapist has to "run after" the client to be in constant video contact.

We propose to combine the “best of many worlds” into our Virtual Glass concept, which we now present and evaluate.

3 VIRTUAL GLASS

In February 2013 Google released the developer version of a head-worn, optical see-through display with an integrated camera, microphone, headphone and a built-in computer controlled by simple gestural commands and voice. Of particular interest for our project are Google Glass's audio and video capabilities, i.e. the combined use of (1) the in-built miniature camera capturing (a portion of) what is in view of the user, (2) the microphone and headphones (bone conduction or otherwise) as communication devices, and (3) the miniature display as a potential video display device. The integrated use of these three components can lead to new forms of videoconferencing: (a) allowing for hands-free, on-the-move remote communication and (b) replacing a "see each other" with a WYSIWIS approach.

While Google Glass is arguably not a suitable device for Virtual Reality or Augmented Reality applications its concept can be used in collaborative environments to implement an audio-video communication channel with social presence-supporting properties. In instructor-performer scenarios and systems an asymmetrical setup allows for a continuous communication between instructor and performer while enabling an instructor WYSIWIS interface. In addition, there is no break in spatial or social presence for the performer.

The basic principle of the Virtual Glass device is that it is intended to replicate for the performer the functionality of a Google Glass within a VE. The performer sees a first-person perspective view of the VE, enhanced by the Virtual Glass showing a web camera view of the instructor. The instructor sees a web camera view of the performer presented on screen as a PiP window. In addition the instructor sees a first person perspective view (if selected) of the performer's Virtual Glass camera.

We have developed the Virtual Glass concept to allow for a continuous experience of social presence for the performer while immersed in the VE and communicating with the instructor at the same time.

With the following two user studies we will demonstrate the effectiveness of the Virtual Glass concept, particularly in relation to the experience of social presence.

4 FIRST USER STUDY

To test whether Virtual Glass is a feasible concept for instructor-performer systems and whether social presence can be improved over standard PiP-enhanced CVE systems we implemented a prototype system.

Based on our experiences and what we learned from investigating a real Google Glass device, several aspects of it were applied to the Virtual Glass in the prototype: (a) the design and dimensions; (b) the display and its features; and (c) a first-person view perspective, all intended to provide the perception of wearing a Google Glass in real life. The displayed size (85 mm x 55 mm on monitor) of the Virtual Glass prism corresponds to what a Google Glass user would see in a real world setting. The PiP video for the instructor was set to the same size for reasons of appropriate comparison (Figure 2).

4.1 The Test System

A fully furnished average-size family house is used as a 3D environment. Using the game engine Unity 3D™, the development follows the concept of multiplayer games involving two networked users (i.e., performer and instructor) who interact in the CVE. Two interfaces share the same virtual world (i.e., one view for the performer and one view for the instructor) and are synchronized with each other. Each interface provides its user with functionalities that suit their tasks.

Both instructor and performer use an almost identical hardware setup, although the instructor operates the system with a standard computer mouse while the performer uses a standard gaming joystick.

A simple avatar entity is added to the world to represent the performer and to enable her/him to move, navigate, and interact within the environment (i.e., picking up or putting down objects).

The instructor is limited to viewing the virtual world from different perspectives, including in particular the first-person perspective of the performer. If the instructor switches to a non-first-person view of the performer then the avatar representation can be seen moving through the environment.

The focus of our study is on using remote communication to evaluate the Virtual Glass technique. A videoconferencing communication library (Video Chat, developed by Midnight Status) is integrated into the system to facilitate video/audio communication between the two users. The video and audio streams are processed and queued, then transmitted over the network in point-to-point streams between two networked users.

In order to start the communication, the performer dons the Virtual Glass that is located on the top of the kitchen table in the virtual house (a 3D model of Google Glass™ acquired from Google 3D Warehouse). This is done using the performer’s input device (the joystick). The Virtual Glass slowly moves in an animated sequence from the top of the table towards the performer where it finally appears to reach his/her face (Figure 1). When the sequence is complete, the glass prism (display) and a small part of the glass's temple are shown at the top right of the performer's screen (simulating the real Google Glass™) (Figure 2 (right)). This animation is a means of introducing the experience of putting on and then wearing the virtual equivalent of a Google Glass device. The main purpose here is to introduce the Virtual Glass in an understandable way ready for CVE communication and to make the user perceive it as a natural, credible and sensible metaphor.

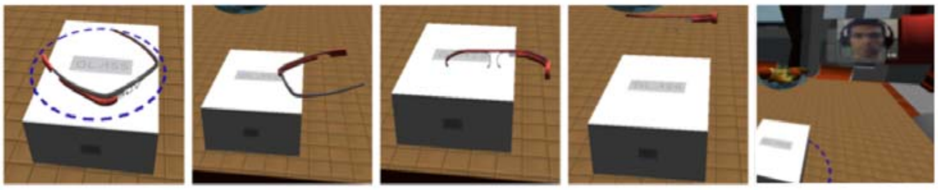

Figure 1. *Animated sequence of putting on Virtual Glass as a means of introducing the metaphor*

Once the client is wearing the Virtual Glass, the audio-video communication channel is established, and the performer now has video communication with the instructor through the glass's prism, and audio communication through their headset. At the other end, the instructor has for communication a common PiP videoconferencing window, along with their own headset. Figure 2 below shows screenshots of the graphical user interface for both the instructor and the performer.

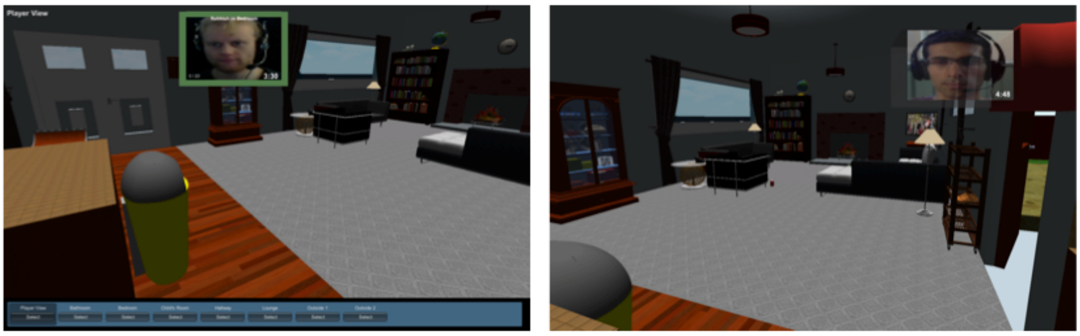

Figure 2. *User interfaces: instructor (left) and performer (right)*

The interface for the performer shows a first-person view of the virtual world and includes the glass prism of the Virtual Glass as a semi-transparent overlay (as in real life). Text messages can also be overlaid on the Virtual Glass display for additional functionality.

Because of the difference in responsibilities, the instructor interface has a different functionality to that of the performer. It provides the instructor with more control over the system. When the application starts, the instructor has a first-person view from the performer's perspective. This means the instructor sees exactly what the performer sees (WYSIWIS). A control area is located at the bottom of the interface for changing the views inside the house. For communication, the interface has a standard videoconferencing window that displays the video stream from the performer.

4.2 Experimental Design

It was hypothesised that the Virtual Glass would promote higher levels of social presence among participants compared with the standard video conferencing interface (PiP).

The main dependent variables included the users' perceived social presence and their understanding of the Visual Glass metaphor, as well as attention, video quality, audio quality, and the question of whether the communication medium would be seen by the participants as helping them to solve the task. The experiment followed a between-subjects design (Standard (PiP) Videoconferencing x Virtual Glass).

4.2.1 Participants

A total of 67 adult participants (aged 16 and above) took part in the experiment in 46 instructor-performer sessions (some participants took part in two session with different roles). Two participants were excluded from the analysis because of missing data. Participants were pseudo-randomly assigned to one of two conditions: 32 participants (14 females and 18 males) to condition A (performers), and 33 participants (16 females and 17 males) to condition B (instructors). The age of the 65 participants ranged from 16 to 71 years (M= 34.55, SD= 12.93).

4.2.2 Apparatus, Conditions and Tasks

The experiment was conducted in one physical room but with two booths set far enough apart that the instructor and performer could not see or hear each other, except via the system. The physical setup for each side was exactly the same (apart from the input device as explained earlier).

The instructor and performer were requested to work collaboratively to solve a shared task inside the CVE. The activity required tidying up the mess in the virtual house (Figure 3) in no more than 5 minutes. The instructor tells the performer where to find items and to where to place them. The performer executes the requested tasks.

In order to have continuous communication and collaboration between the instructor and the performer, the clutter in the house (e.g., rubbish, food, shoes, clothes, toys) was organized in sequence. This meant that there was only one item of clutter in the house at a time, and when the performer dealt with it, another item would appear elsewhere and the performer would need to deal with that one next. The instructor is notified via text instructions on the video screen of the nature and location of each next item, and can use the multi-viewing functionality to observe and guide the performer.

The performer commences his or her task by “picking up” and “putting on” the Virtual Glass, which enables the video communication channel. The glass display for the performer shows a 5-minute countdown timer to track the progress and encourage faster performance. The counter is also displayed for the instructor. The instructor helps the performer to find the items in the house. This is done by reading the text instructions from the videoconferencing window and providing the performer with appropriate information. The text instructions include the type and location of the item and how the performer should deal with it (e.g., take the rubbish to the rubbish bin). Based on these instructions, the performer keeps performing the tidying activity by using the joystick to navigate in the environment and to pick up and relocate clutter items.

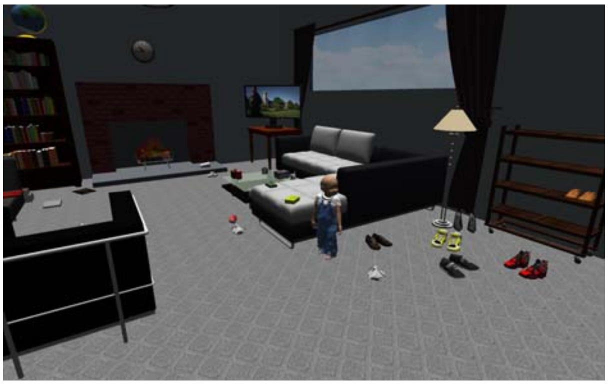

Figure 3. Example room with multiple items of clutter (during the experiment only single items would be presented)

For the data collection, a paper questionnaire was used for both sides. Demographic information requested included gender and age. To ensure better control of the experiment, participants were asked if they knew the other partner and if they had previously used either a videoconferencing system or Google Glass. A seven-point Likert-like scale (from 1 "strongly disagree" to 7 "strongly agree") was used to measure the dependent variables understanding, attention, video quality, audio quality, and the degree to which the communication medium helped the participant to solve the task. Social presence was measured using an unmodified scale by Short, Williams, & Christie (1976). This instrument has been successfully used in numerous studies before, e.g. (Hauber et al., 2006).

4.2.3 Results

Most of the participants in both groups (the Virtual Glass in group A, and standard PiP videoconferencing in group B) rated a good understanding of the concept of the communication mediums and how they worked. The responses were significantly above the scale midpoint ($p < .05$; mean difference 1.906 and 1.576 respectively).

Examining the level of understanding of the communication mediums between the two groups (A: performers with Virtual Glass, versus B: instructors with PiP interface) a two samples Mann-Whitney U test indicates that there is no significant difference between the two groups ($p > .05$). So, despite the introduced novelty of the Virtual Glass, the understanding of the concept was rated equally well.

Knowing the other collaborator in the task had no influence on any of the dependent variables (social presence, attention, video quality, audio quality, and solving the task). A one-way MANOVA showed that there was no significant difference between participants who knew or did not know the other collaborator (Pillai's trace test, $p > .05$).

The performers group (A) rated social presence and video quality significantly higher than the scale midpoint ($p < .05$, mean difference .813, $p < .05$, mean difference .625 respectively). Responses to all other factors were close to the midpoint with no significant differences.

To test the hypothesis that there is a statistical significant difference between the Virtual Glass (performers group) ($N = 32$, $M = 4.81$, $SD = 1.355$) and standard PiP video conferencing interface (instructors group) ($N = 33$, $M = 3.70$, $SD = 1.262$) in terms of social presence, an independent sample t-test was performed. In addition to the normality assumption, the assumption of homogeneity of variances was tested and found to be satisfied using Levene's F test, $F(63) = .195$, $p = 660$.

For our main focus on social presence, the independent sample t-test showed a statistically significant effect, $t(63) = 3.437$, $p = .001$. Thus, the mean social presence of the Virtual Glass was significantly higher than that of the standard videoconferencing interface (PiP). Social presence was measured using multiple-item construct scales with very high reliabilities (Cronbach's Alpha (α=.961 (performers), α=.956 (instructors)).

The Virtual Glass (performers group) and standard PiP video conferencing interface (instructors group) showed no statistically different means compared with respect to the other dependent variables (attention, video quality, audio quality, and solving the task).

4.2.4 Limitations of the first study

First, some participants were allowed to take both roles in different sessions (performers and instructors). The effect of this confounding variable was not examined during the analysis, which might have influenced the results. The difference in the interfaces may also have affected the results leading to different social presence ratings. The instructors' and performers' interfaces were slightly different. The control area located at the bottom of the instructor's interface for changing the views inside the house added functionality. More importantly, both instructors and performers had different tasks. Instructors guided performers to tidy up the virtual house by instructing them where to find the task items and where to put them. Performers executed the task "tidying up the house" that required them moving around the house, finding the items, picking them up, and then putting them in the appropriate places. This might have introduced a bias, because the performer and instructor had to act in different roles with different tasks. In addition, the difference in the tasks required the participants to use different input devices. The instructors operated the system with a computer mouse while the performer used a gaming joystick.

Another potential shortcoming of the first study was not investigating more immersive wearable display device technologies (i.e., head-mounted display (HMD)). With the Virtual Glass mechanism, the participants wear the Virtual Glass in the VE: how would participants react if this action was accompanied with wearing a glasses-like device in real life? For example, a more immersive HMD? Could this enhance the impact of the Virtual Glass concept?

Based on these limitations, it was difficult to draw a definite conclusion as to whether the significant change in the level of social presence was caused by the new Virtual Glass concept or due to the influence of one or more of the above factors.

5 SECOND USER STUDY

This study was designed to overcome the limitations from the first study and to provide a deeper understanding, analysis and interpretation of the results. To eliminate the stated limitations, the new focus was entirely on the performer's side. Data were only collected from the performers, and the same instructor worked as a partner for all performers, which resolved the first mentioned limitation. In addition, two identical interfaces (Figure 4) were designed for the performers where a participant used both the Virtual Glass (VG) and PiP designs to perform exactly the same task in two sessions within the same time frame using identical hardware. This eliminated the other limitations.

Furthermore, a HMD condition was implemented to examine the possible influence of using a more immersive display on the Virtual Glass concept. Besides the use of a standard monitor, the Oculus Rift DK2 was configured with the system to examine if it could provide higher levels of presence.

Figure 4. Performer interfaces: PiP (left) and VG (right)

5.1 Experimental Design

The experiment followed a 2 (communication mechanism) x 2 (display type) factorial design. This was based on two groups generating four conditions in which the communication mechanism was a with-in-subject variable and the display type was a between-subject variable. Each variable consisted of two levels: the communication mechanism (VG x PiP) and the display type (Monitor (M) x HMD).

This study hypothesised that:

Communication hypothesis:

- Users' social presence and sense of presence would be higher with the VG mechanism regardless of the display types (M x HMD).

Immersion hypothesis:

- Users' social presence and sense of presence would be higher with the HMD regardless of the communication mechanism (VG x PiP).

Matching hypothesis:

- Users' social presence and sense of presence would be higher with the HMD and VG mechanism.

5.2 Counterbalancing

Based on the mixed design, every participant in each group (M x HMD) repeated the same task twice (VG x PiP). To control for any possible learning effects, participants were balanced in randomised order. A participant was pseudo-randomly assigned to one of the two groups (M or HMD). Then, randomly assigned either to the VG or PiP first.

5.3 Participants

A total of 50 adult participants took part in the experiment as performers (25 participant in the M group and the same number for the HMD). One of the paper's authors played the role of the instructor in the four conditions with all participants. In the M group (9 females and 16 males) the age of the participants ranged from 18 to 32 years (M = 20.44, SD = 2.87). The participants' age of the HMD group (11 females and 14 males) ranged from 18 to 33 years (M = 21.36, SD = 3.61). All participants had normal or corrected-to-normal vision. From the total number of the participants, only two had tried a real Google Glass and HMD (same two participants from the HMD group) before. On the other side, all 50 participants had tried using videoconferencing systems such as Skype and FaceTime.

5.4 Ethics, Procedure and Tasks

An ethical approval was obtained from University of Otago. Participants were provided with information sheets prior the experiment, and a written consent form was signed and obtained from each participant.

The same tasks and procedures from the first study were applied in all four conditions along with changing the display device, and the communication mechanism. Each participant performed the task twice based on the display device that to which they were initially randomly assigned (M-VG/M-PiP, or HMD-VG/HMD-PiP).

5.5 Measures

For the data collection, a paper questionnaire was used. Questions were completed after the five minute task for all four conditions, which included the following:

- Seven-point Likert-like scales (from 1 "strongly disagree" to 7 "strongly agree") to measure the variables understanding, attention, video quality, audio quality, and the degree to which the communication medium helped the participant to solve the task.
- Perceived social presence: the Short, Williams, & Christie, (1976) construct was used for measuring social presence.
- Perceived sense of presence: to measure sense of presence in the VR environment, the Igroup Presence Questionnaire (IPQ) was used (Schubert et al., 2001).

For the HMD conditions, two more questions were asked to assess the participants' well-being and discomfort using the HMD:

- Simulator sickness questionnaire (SSQ): the SSQ developed by (Kennedy, Lane, Berbaum, & Lilienthal, 1993) was used to measure simulator sickness.
- Subjective discomfort: the level of discomfort using the HMD was rated by a modified questionnaire adopted from (Peli, 1998).

5.6 Results

First, the data distributions of all variables were assessed using graphical methods (histograms) and numerical methods using descriptive statistics (Shapiro-Wilk test). Results showed that the data of the variables social presence and sense of presence were normally distributed. However, the rest of the variables departed from normality (i.e., understanding, attention, video quality, audio quality, the degree to which the communication medium helps the participant to solve the task, and the subjective discomfort). Based on this result, parametric tests were used to examine social presence and sense of presence (the proposed hypotheses), and non-parametric tests for the other variables.

Because both social presence and sense of presence are constructs comprising several items, reliability tests were performed to ensure that these items measure the same construct (i.e., Cronbach's Alpha). The results indicated that both constructs were reliable showing high and strong internal consistency among their items (higher than 0.7).

5.6.1 Social Presence and Sense of Presence

The means and standard deviations of all four conditions for social presence and sense of presence are shown in Table 1 below. The total of means for both groups are also provided (within-subject and between-subject).

			Communication mechanism		
			VG	PiP	Total M
Social Presence	Display type	Monitor	M = 4.92 SD = 1.13	M = 3.71. SD = 1.00	4.32
		HMD	M = 5.01 SD = 1.23	M = 3.95 SD = 0.94	4.48
		Total M	4.97	3.83	
Sense of Presence	Display type	Monitor	M = 0.38 SD = 0.89	M = 0.25 SD = 1.05	0.32
		HMD	M = 0.58 SD = 0.96	M = 0.54 SD = 1.04	0.56
			0.48	0.39	

Table 1. Means and standard deviations of social presence and sense of presence

To examine the proposed hypotheses focusing on social presence, a two-way, repeated-measures ANOVA was conducted. The results supported the first hypothesis that there was a significant difference between the VG and PiP regardless of the display types as shown in Figure 5, $F(1,48) = 41.25$, $p < .001$, $\omega 2 = 0.46$. For the immersion hypothesis, there was no significant difference between HMD and standard monitor regardless of the communication mechanism, $F(1,48) = 0.41$, $p = 0.52$, $\omega 2 = 0.009$. ANOVA also revealed that there was no significant interaction effect between the display type and communication mechanism (users' social presence was not significantly higher with the HMD and VG mechanism), $F(1,48) = 0.19$, $p = 0.66$, $\omega 2 = 0.004$.

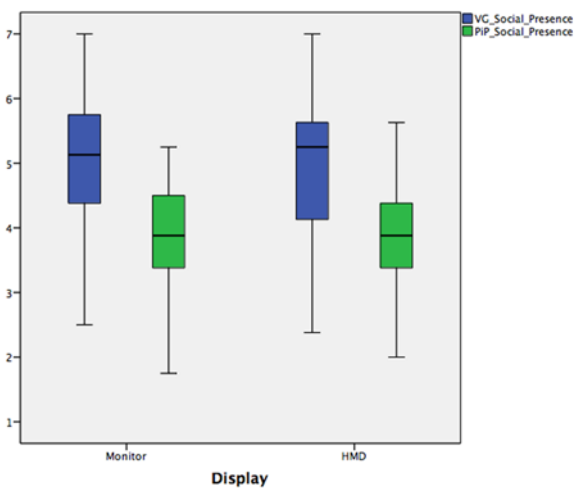

Figure 5. Boxplot showing the four conditions of social presence

Another two-way, repeated-measures ANOVA was performed concentrating on sense of presence. The results showed that all proposed hypotheses of this variable were rejected. There was no significant difference between the VG and PiP regardless of the display types, $F(1,48) = 0.75$, $p = 0.38$, $\omega 2 = .01$. Also, no significant difference were found between HMD and standard monitor regardless of the communication mechanism, $F(1,48) = 0.85$, $p = 0.35$, $\omega 2 = 0.01$. Last, there was no significant interaction effect between the display type and communication mechanism, users' sense of presence was not significantly higher with the HMD and VG mechanism, $F(1,48) =0.21$, $p = 0.64$, $\omega 2 = 0.005$.

5.7 Additional Variables

As the rest of the variables did not fulfil the normality assumption, non-parametric tests were employed to examine them (i.e., understanding, attention, video quality, audio quality, and the degree to which the communication medium helped the participant to solve the task). Before performing any tests, the means of all variables in all conditions were checked and found to be above midpoints (i.e., 4).

First, a Mann–Whitney U test was used to test each variable across the two groups / between-subjects (e.g., the variable understanding was compared across the display type groups, M x HMD). All results showed that there was no significant difference between the variables across the display type, supporting the null hypothesis.

For testing the same variables with the related samples / within-subjects, a Wilcoxon signed-rank test was used (e.g., the variable understanding was compared within the same group across the communication mechanism, VG x PiP). This was performed for the two groups, M and HMD. Again, no significant differences were found in the test results for all variables and the null hypotheses were retained.

5.8 Simulator Sickness and Subjective Discomfort

To assess simulator sickness that could impact the participants' performance during the experiment, the simulator sickness questionnaire (SSQ) was used. In the current experiment, only participants in the HMD group were asked to answer the SSQ. The analysis of the SSQ focused on the symptom severity scores and not the total scores. Firs, the scores of all symptoms were summed for each participants and then the overall means of all participants were assessed from both conditions and compared with the categorization of symptoms provided by (Kennedy et al., 2003). Results showed that the means of the symptom scores for the two conditions were HMD-VG: $M = 8.56$; and HMD-PiP: $M = 9.28$. Both scores fitted in the scores category 5-10, which reflected minimal symptoms. This result was expected as none of the participants' abilities to complete the experiment were affected.

Participants reported a low subjective discomfort using the HMD for both conditions (HMD-VG: $M = 1.05$, $SD = 0.92$; HMD-PiP: $M = 1.10$, $SD = 1.03$). In addition, a Wilcoxon signed-rank test showed no significant difference between VG and PiP in terms of subjective discomfort.

In summary, the two studies showed that the concept of Virtual Glass is effective, that it supports a high level of social presence and that the social presence for the performers is rated significantly higher than a standard picture-in-picture videoconferencing approach used for the performers.

6 DISCUSSION AND FUTURE WORK

Overall, the results show that the Virtual Glass concept was well understood by the majority of the participants. They were clearly comfortable with this mode of communication and it seems to have been no less comprehensible than a standard video conferencing interface. This is quite surprising, as only six of our participants had previously tried a Google Glass or similar interface. It is even more surprising considering that awareness of the Google Glass interface technology is not wide-spread.

Despite the fact that both tested interfaces provided the same type of communication (video/audio), the interfaces performed significantly differently with respect to perceived social presence. The level of social presence was rated significantly higher when using the Virtual Glass than with the standard videoconferencing interface regardless of the display types, thereby supporting our main social presence hypothesis. We argue that the increased social presence results from (a) accepting the concept of the Virtual Glass, as well as from (b) continuously maintaining task performance in the VE, whilst (c) simultaneously communicating with the remote collaborator. Perhaps the Virtual Glass was seen as less "artificial" and therefore affected social presence positively (Hauber, Regenbrecht, Hills, Cockburn, & Billinghurst, 2005).

While the Virtual Glass had a significant effect on social presence, there was no corresponding effect on the sense of presence, surprisingly, even when using the HMD. Our matching hypothesis of higher social presence and sense of presence using the HMD with the VG was also not supported. These results occurred even though no significant effect appeared on any of the other dependent variables: understanding, attention, video quality, audio quality, and solving the task (across the display types and the communication mode).

Further research is needed to prove the general applicability of the Virtual Glass concept, but for certain application types, like collaborative VRET systems or VR multi-player computer games, Virtual Glass seems to be an appropriate, social presence-enabling or -increasing component. Future research could also explore to what extent our results would be improved when people have tried or are users of a physical Google-Glass-like interface.

Another possibility could be to exchange CVE interfaces, such that the instructor adopts the Virtual Glass interface, allowing the performer to become an observer of their actions, allowing them to switch between the instructor’s first-person perspective via Virtual Glass and the other available viewpoints into the VE. This would then be a "What I See Is What You See" (WISIWYS) system (rather than WYSIWIS), and which could facilitate broadcast of instructor demonstrations to numbers of performers. It would be an Instructional Virtual Environment and could be useful in situations where the observers would benefit from the instructor’s enhanced feeling of social presence, such as in sports training or other kinaesthetic activities.

Future studies might investigate the influence of communication metaphors with appropriately implemented modes on performance or other characteristics. With our findings we can demonstrate how much the metaphor (here Virtual Glass) and how little the immersion technology (here the HMD) contributed to the experience of being together (social presence).

7 ACKNOWLEDGMENTS

The first author of this research is funded and sponsored by the Saudi Ministry of Education. The authors wish to thank all of our participants in the experiment and the members of the HCI group for their help and support during this research. Mike Goodwin helped the first author a lot with this study and in general. Mike died last year - we will always remember him. Also, special thanks to Elgar Storm and Abdulaziz Alshaer for their help and support during this research.

References

Barfield, W., & Weghorst, S. (1993). The sense of presence within virtual environments: A conceptual framework. *Advances in Human Factors Ergonomics, 19*, 699-699.

Biocca, F., Harms, C., & Gregg, J. (2001). *The networked minds measure of social presence: Pilot test of the factor structure and concurrent validity.* Paper presented at the 4th annual International Workshop on Presence, Philadelphia, USA.

Carlin, A. S., Hoffman, H. G., & Weghorst, S. (1997). Virtual reality and tactile augmentation in the treatment of spider phobia: A case report. *Behaviour Research and Therapy, 35*(2), 153-158.

Cobb, S., Kerr, S., & Glover, T. (2001). The AS Interactive Project: Developing virtual environments for social skills training in users with Asperger syndrome. *Robotic and Virtual Interactive Systems in Autism Therapy, Communications of the Adaptive Systems Research Group, University of Hertfordshire (Report no 364).*

Csikszentmihalyi, M. (1991). *Flow: The psychology of optimal experience* (Vol. 41). New York, USA: Harper Perennial.

De Greef, P., & Ijsselsteijn, W. A. (2001). Social presence in a home tele-application. *CyberPsychology & Behavior, 4*(2), 307-315.

Hauber, J., Regenbrecht, H., Billinghurst, M., & Cockburn, A. (2006). *Spatiality in videoconferencing: trade-offs between efficiency and social presence.* Paper presented at the Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work, Banff, Alberta, Canada.

Hauber, J., Regenbrecht, H., Hills, A., Cockburn, A., & Billinghurst, M. (2005). *Social presence in two-and three-dimensional videoconferencing.* Paper presented at the Proceedings of 8th Annual International Workshop on Presence, London/UK, September 21-23, 2005. pp. 189-198.

Juan, M. C., Alcaniz, M., Monserrat, C., Botella, C., Banos, R. M., & Guerrero, B. (2005). Using augmented reality to treat phobias. *Computer Graphics and Applications, IEEE, 25*(6), 31-37.

Kennedy, R. S., Drexler, J. M., Compton, D. E., Stanney, K. M., Lanham, D. S., & Harm, D. L. (2003). Configural Scoring of Simulator Sickness, Cybersickness and Space Adaptation Syndrome: Similarities and Differences. *Virtual and adaptive environments: Applications, implications, and human performance issues*, 247-274.

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology, 3*(3), 203-220.

Lengel, R. H., & Daft, R. L. (1989). The selection of communication media as an executive skill. *The Academy of Management Executive (1987-1989)*, 225-232.

North, M. M., North, S. M., & Coble, J. R. (1997). Virtual environments psychotherapy: A case study of fear of flying disorder. *Presence: Teleoperators and Virtual Environments, 6*(1), 127-132.

Park, K. S., & Kenyon, R. V. (1999). Effects of network characteristics on human performance in a collaborative virtual environment. *Virtual Reality, 1999. Proceedings., IEEE*, 104-111.

Peli, E. (1998). The visual effects of head-mounted display (HMD) are not distinguishable from those of desk-top computer display. *Vision research, 38*(13), 2053-2066.

Regenbrecht, H. (1999). *Faktoren für Präsenz in virtueller Architektur [Factors for the sense of presence within virtual architecture].* (Doctoral thesis (Dissertation)), Bauhaus University Weimar, Germany.

Regenbrecht, H., Lum, T., Kohler, P., Ott, C., Wagner, M., Wilke, W., & Mueller, E. (2004). Using augmented virtuality for remote collaboration. *Presence: Teleoperators & Virtual Environments, 13*(3), 338-354. doi:10.1162/1054746041422334

Regenbrecht, H., Wickeroth, D., Dixon, B., & Mueller, S. (2006). *Collaborative mixed reality exposure therapy.* Paper presented at the Cyberworlds, 2006. CW'06. International Conference on.

Schubert, T., Friedmann, F., & Regenbrecht, H. (2001). The experience of presence: Factor analytic insights. *Presence:Teleoperators and virtual environments, 10*(3), 266-281.

Short, J., Williams, E., & Christie, B. (1976). *The Social Psychology of Telecommunications*. London: Wiley.

Slater, M., Guger, C., Edlinger, G., Leeb, R., Pfurtscheller, G., Antley, A., . . . Friedman, D. (2006). Analysis of Physiological Responses to a Social Situation in an Immersive Virtual Environment. *Presence: Teleoperators & Virtual Environments, 15*(5), 553-569. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=iih&AN=23240136&site=ehost-live&scope=site

Steuer, J. (1992). Defining virtual reality: Dimensions determining telepresence. *Journal of communication, 42*(4), 73-93.

Teoh, C., Regenbrecht, H., & O'Hare, D. (2010). *Investigating factors influencing trust in video-mediated communication.* Paper presented at the OZCHI, Brisbane, Australia.

Vertegaal, R. (1999). *The GAZE groupware system: mediating joint attention in multiparty communication and collaboration.* Paper presented at the Proceedings of the SIGCHI conference on Human Factors in Computing Systems, New York, NY, USA.