Breaks in Presence as Usability Criteria

A. Steed, A. Brogni, V. Vinayagamoorthy

Department of Computer Science, University College London
Gower St, London, WC1E 6BT
A.Steed@cs.ucl.ac.uk, A.Brogni@cs.ucl.ac.uk, V.Vinayagamoorthy@cs.ucl.ac.uk

Abstract

A number of authors have noted that usability engineering for virtual environments (VEs) is complicated by the fact that the worlds and tasks are often quite open ended and that the experience is real-time. In particular, with immersive virtual environments (IVEs), the simulation is designed and engineered so that the participant experiences presence in a place other than where they are physically related. In such environments there is often not a single task and the simulations have to be built to encompass a wide range of possible participant behaviours. Informally, our observations of participants in such simulation show that sometimes they are behaving as if the simulation was real, and at other times they are responding to some aspect of the technology. Slater & Steed (2000) introduced the Breaks in Presence (BIP) model which was used as a way of identifying changes (breaks) from the virtual to the real. They constructed a model that could be used to measure presence from the number of BIPs. In this paper we propose studying the BIPs themselves and treating them as events for usability analysis. In a pilot trial we show that by recording the participants experience and then replaying it with a visualization of the occurrence of BIPs experimenters can start to identify aspects of the VE that might need attention and refinement.

1 Introduction

A number of authors have noted that usability engineering for virtual environments (VEs) is complicated by the fact that the worlds and tasks are often quite open ended and that the experience is real-time (Johnson, 1999, Tromp, Steed & Wilson, 2003). The open-ended nature of most VEs means that it can be difficult to identify usability problems and their importance. The real-time nature of the VE experience means that it is difficult to obtain usability information because of the pace of the task and the likely unsuitability of a method such as a talk aloud protocol (Nielsen, Clemmensen & Yssing, 2002).

In related work, many researchers have identified presence, the sense of being in the virtual environment and acting appropriately, as an important phenomenon in virtual environments (Draper, Kaber & Usher, 1998, Sadowski & Stanney, 2002). To examine presence, researchers have typically set up controlled experiments where they have examined how one or two parameters of the world or displays have affected presence. They have typically used a subjective measure such as a questionnaire. Such experiments are very valuable but they do not provide a methodology that can be used in an engineering process because they are summative in nature and laborious to undertake.

Recently Slater and Steed's Breaks in Presence (BIPs) model has provided a method and framework for understanding how presence might vary over time (Slater & Steed, 2000). In that paper, BIPs were self reported, but in more recent work physiological monitoring has been used to provide a non-invasive method for getting some objective data about changes in a participant's experience might over time (Brogni, Slater & Steed, 2003, Slater, Brogni & Steed, 2003).

In this paper we take the self-report BIPs methodology and whilst the participant is in the VE we make a complete recording of the participant and VE behaviour. Experimenters studying a replay of the recording can study the behaviour of the participant around the time that they signalled a BIP. They can pinpoint common issues that might cause BIPs and explore whether these issues might be addressable as usability problems.

We demonstrate this idea with an urban street environment designed for the purpose of studying agoraphobia (Brogni, Slater & Steed, 2003, Slater, Brogni & Steed, 2003, Romano et al. 2002). We study the causes of BIPs reported by five participants in a pilot trial. We then discuss how this method might be expanded and made more rigorous with the aim of developing a novel usability methodology.

2 Related Work

2.1 Usability of Virtual Environments

At one level, examining the process of studying the usability of a VE should be no different from any other form of usability since VEs are computer-generated interfaces. However, unlike many domains, the interface itself and the simulations presented through it vary greatly. Indeed a VE is often built to support many potential tasks, but in achieving this, the application may also have been built with many apparent affordances (Gaver, 1992, McGrenere & Ho, 2000) that are hard to support behaviourally. With any simulation, be it an urban environment in a 3D game such as Sony Computer Entertainment Europe's 'The Getaway', or a high-end flight simulation, the graphics and sounds that are presented will suggest that certain items might be functional but there will limits to the interactivity. This freedom and the real-time nature of the experience mean that applying traditional HCI techniques has been problematic (e.g. Tromp & Steed, 1998).

Interest in this area has been growing (Bowman, Gabbard & Hix, 2002), but still relatively little has been presented that attacks these problems head on. Most usability work to date has focussed on the performance of low-level interaction tasks such as selection, manipulation and locomotion. Typically this has focussed on abstract tasks that represent a number of common interaction behaviours. For example, Poupyrev et al. (1997) was one of the first systematic analyses of manipulation techniques. Bowman et al. (2004) give a comprehensive overview of this aspect of VE design. A number of other authors and teams have tackled the higher-level problem of designing a VE to support a complete task. Gabbard (1997) presents a taxonomy for VE design. Parent (1998) created a VE task analysis workbook for the specific example of creating virtual art exhibits. Kaur, Sutcliff & Maiden (1998) identified several interaction cycles and then designed their application to support these cycles. Hix, Swan, et al. (1999) gave an insight in to the detailed iterative usability design and evaluation of a battlefield visualization.

The approaches mentioned above focus on identification and evaluation of a small number of tasks. However as noted, often the task is not very specific. A completely different approach to the evaluation of VE systems is the study of presence. Presence is a concept that has received a lot of attention, and several attempts at definitions (Draper et al. 1998, Sadowski & Stanney, 2002). One definition of an effective VE might be a VE where the person experiences a sense of presence within that environment and thus acts according to the stimuli received, not the situation of presentation (Whitton, 2003). Some of the most compelling demonstrations of effectiveness are VEs that generate significant and appropriate stress responses to threatening situations (Pertaub, Slater & Barker, 2001, Meehan et al. 2002). Stress situations can be objectively monitored using physiological measures (Meehan et al. 2002, Slater, Brogni & Steed, 2003), but otherwise presence researchers have mainly relied on subjective questionnaires (Slater, Usoh & Steed, 1994, Witmer and Singer, 1998, Lessiter et al. 2001). Questionnaires are often unsatisfying because they are the subjects' post-hoc rationalisations of the experience. There may also be a number of problems with the language that they use (Usoh et al., 2000). Summative measures of presence are themselves not particularly useful for studying usability. Typically they have been used to examine fairly gross choices in the representation or interactivity of an environment. For example, in Slater, Usoh & Steed (1994) two methods of travel between areas within worlds were investigated. This approach would be too expensive to employ in an iterative design cycle because there would be too many comparisons to make.

2.2 Breaks in Presence

The breaks in presence approach was introduced by Slater & Steed (2000). This approach is based on the idea that a participant experiencing virtual reality technology, at any one moment, interprets the stimuli coming from the environment as belonging either to the virtual or to the real world. Thus presence was treated as a gestalt: there are two possible interpretations of the stimulus received. Slater & Steed suggested that the participant switches between the two interpretations throughout the experience, and that a measure of presence could be obtained if the amount of time that the participant spent interpreting the stimuli as coming from the virtual could be estimated. They proposed to do this estimation by looking for "breaks", those times when the participant realised they were in the real world. This only gives one direction of change, but this does allow an estimation of overall presence to be obtained. The process is designed to be lightweight, in that it doesn't change the main task of the VE experience.

In the Slater & Steed (2000) study the times of breaks were gathered participant was asked to say "Now" when they realised they were in the real world. The participants were trained in the elicitation of their change in state on 2D Gestalt pictures that included well-known images such as the ambiguous face-vase illusion.

Having the participants make a verbal report of BIPs raised the immediate issue that the participants might be expecting someone to listen, and thus they could presume that they are able to talk to the experimenters. In more recent variations of the protocol, participants have been asked to press a button on a hand-held controller in order to signal a BIP (Brogni, Slater & Steed, 2003). This has its own drawback in that it is limited to environments where there is no interaction only locomotion. Our experience is that participants easily confuse buttons and this could lead to ambiguous situations. Recent investigations have determined that it may be possible to detect some types of BIP from physiological monitoring (Slater, Brogni & Steed, 2003).

In de-briefing after studies, participants report BIPs in three main categories

- External – related to real events
- Internal – related to virtual events
- Situational – related to personal and environmental situation

External BIPs are caused by the display failing to exclude the real world. This includes interruptions from external sources and interference from the display systems. The former is easier to control if there is freedom to isolate the participant. For example, a common report is of external noises such as phones ringing. The latter is of course impossible to remove with today's technology. In the original Slater and Steed study, one participant reported a BIP when they felt the cable of the HMD brush against their legs. In our recent studies, using a CAVE^{TM}-like device, the trackers are now wireless, but we still can't remove problems caused by tracking or glasses failure or other incidents such as the participants colliding with the walls. External BIPs are thus tackled by changing the form of the display system, careful control of the situation of the display and good protocols for introducing the participant to the display.

Internal BIPs are related to the environment. Typical reports include objects not behaving "correctly" (e.g. a chess piece that floated through the air), or not being consistent such as objects that made no sound when dropped. This doesn't necessarily mean that behaviours have to be realistic, but they should be consistent and unsurprising for the context. These types of behaviours are under the control of the VE designer and implementer to some extent and seem tractable, but in actuality we very quickly reach the situation noted in the introduction: the system cannot support all the affordances that are suggested by the appearance of the objects. This is quite pervasive, and two examples from the original BIPs paper will highlight the difficulties: not all objects appearing to be solid, and a grass texture not feeling like grass when one stood on it. Internal BIPs can thus be tackled by a balance of design and implementation when necessary, and steering the participant so they encounter few surprises.

Situational BIPs fall in to four sub-categories: task, personal, attention or spontaneous. Task BIPs relate to complexity or confusion in instructions. Participants will often be confused by instructions and this is not helped by the relatively novel form of display being used. Personal BIPs arise because of self-awareness of the situation. An example would be becoming aware that they were being watched by the experimenters even though the experimenters would not usually be visible. Attention BIPs arise because the task is not interesting enough and the world does not afford any further activity. Finally some participants report BIPs spontaneously with no particular reason.

3 Background to the Street Experiment

One of the most promising applications of IVEs is in virtual reality exposure therapy (VRET) for treating certain types of phobia (Hodges et al., 2001, Pertaub, Slater & Barker, 2001). From the discussion in the previous section, one measure of success of a VE is that the participant behaves towards the VE in a way that is consistent with the expected behaviour in similar situations in the real world. The rationale behind VRET is that if the person reacts to the VE simulation of a stress-inducing situation, then that simulation can be used as part of an exposure protocol.

In some of our recent work we have focussed on agoraphobia. Agoraphobia is a term that covers a broad range of phobias, but it is mainly concerned with a stress response to outdoor environments or the objects that might be found in them (Salovskis & Hackmann, 1997). We have focussed on what is potentially a subset: agoraphobia within urban environments. Thus, we have built several urban simulations and exposed people to these environments. Our interest is two-fold: we are interested in the simulations as a tool for VRET; but also we are interested in what makes people feel present in such environment.

Simulations of environments that might provoke a stress response are a good example of the difficulty in defining a task that was mentioned in the previous section. We do not have a good specification of what the environment should contain or do. We had previously done a study that compared six very different environments that varied in

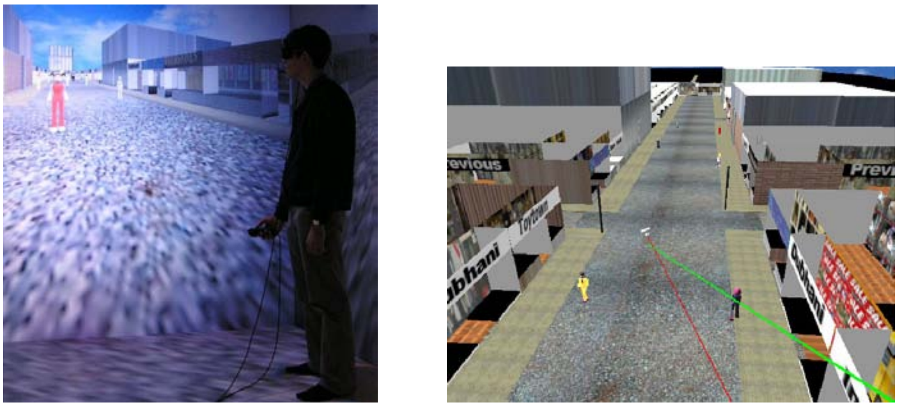

Figure 1: Left: A participant in the UCL ReaCTor experiencing the street environment. Right: A view of the street environment from above. The participant can be seen within it (see Section 4.2).

type of space, geometric complexity and population density (Romano et al., 2002). The task in these environments was simply to explore and to get a feel for that place. Thus the success of the task could only be defined by whether the participants had an experience where they reacted to the world as if it were real. This would include feeling stress on exposure to heights or people.

In a previous study (Vinayagamoorthy et al. 2004), we had investigated whether there needs to be consistency between the levels of realism of different elements within the representation of an urban scene. In that paper, this meant whether the level of realism of the buildings needed to be consistent with the level of realism of the characters populating the environment. Our main interest was to try to understand where to devote the effort in building such an environment: did the appearance of the buildings or the appearance of characters inside the environment matter more? We hypothesised that less repetitive textures in the scene and more lifelike characters would enhance the participants' presence experience in the virtual world. We varied visual realism of the VE 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 type had a face that was texture-mapped from real human faces. Figure 1 shows example views of the world used in that study and in the current study. The results of experiment showed that if the avatars were relatively realistic, but the world was not, subjects reported a lower sense of presence compared to other combinations. This suggests that there needs to be some consistency of representations.

4 Recording and Replaying Sessions with BIPs

Section 2.2 summarised the variations of the BIP methodology that we have employed in previous study. As discussed in that section, regardless of the manner in which a BIP is reported, the BIPs highlight some part of the participant's immediate response to the environment. Previously we had simply recorded the BIPs as they happened and then asked the participant to reflect upon the probable causes of the BIPs after the completion of the experience. This was done because it would be inappropriate to interrupt the participant during the experience, as this would potentially cause further BIPs. However this meant that it was impossible to track the exact circumstances under which a BIP occurred.

The process we are developing and have tested in pilot trials is to make a complete recording of the IVE session so that it can be replayed. The recording will contain all of the behaviours of elements of the VE and the tracking information and avatar representation of the participant. This record will be sufficient to replay the session exactly as was presented to the participant, or to take a 3rd person view of the participant in the environment. If we also record the occurrences of BIPs, we can then replay the VE around that time so that we can identify what the participant saw and potentially identify environmental causes of particular BIPs.

We have utilised and extended a record and replay mechanism for the DIVE software (Frécon et al., 2001). DIVE (Distributed Interactive Virtual Environment) is an Internet-based multi-user virtual reality system in which participants can locomote in a shared 3D space and interact with each other. DIVE was originally designed for desktop environments, but has been extended to support IVEs such as CAVE^{TM}-like (Cruz-Neira et al., 1993) immersive projection technology (IPT) systems (Steed, Mortensen & Frécon, 2001). The immersive version of DIVE hides all the complexity of the immersion support and allows an immersive client to interact with a collection of other desktop and immersive clients connected over the Internet.

DIVE is a peer-to-peer system and clients that join a multi-user session fetch the current state from one of the peer group. Afterwards all participants are kept up to date by distribution of events that represent change in state. Distribution of events is done using multicast. The peer-to-peer nature of DIVE has two important implications: a participant joining must identify the world they are joining so that they can be introduced to the peer group sharing that world; and the first person to join the world gets to create it as they wish, though they will usually be reading and instantiating a world from a file.

4.1 Record and Replay

The record and replay mechanisms exploit the multi-user nature of DIVE. Figure 2a illustrates the record mechanism and Figure 2b the replay mechanism. In the following text numbers in parentheses refer to the labels in that figure. A "diveserver" process is initialised. A diveserver's role is not to distribute events; it only matches requests for worlds to the multicast group so that peers can then contact any other relevant peers. The world is started on a desktop client ("vishnu") or immersive DIVE client ("spelunk"). The client fetches the multicast group to use from the diveserver (1). The client then loads the world description. Recording is done by a "record client". The record client is started by with the name of a world to join, and the diveserver will give it the same group (2). Once it detects a peer, the record client receives a complete copy of the world state from that peer (3). It serialises this to disk in a binary format (4). It then listens for all events on that world (5), and records then in an ASCII log file (6).

Replay takes advantage of the fact that the first peer gets to decide how a world with a particular name gets populated. Replay is done by a "player client". The player client asks the diveserver which group to join (1). The player client will have given the name of a binary serialisation of a world. It loads this file, rather than the world's source files, when it starts (2). It then starts replaying the event in the ASCII log file at a controllable rate (3). At the current time, playback can only be forwards in time, but this limitation will be removed soon. One or more vishnu or spelunk clients can then connect to the same world (4). They get a copy of the world from the player client (5) and receive any subsequent events (6). They will see the original avatar(s) moving around the world, but cannot interact with them.

4.2 Recording BIPs

DIVE has a scripting language (DIVE /TCL) built-in so retrieving buttons presses is a simple case of registering an appropriate DIVE /TCL handler. We are only using buttons presses to signal BIPs in this environment, so any button press will register a BIP. We can print out a message to the console when this happens, but we want to then correlate this with the performance of the participant in a replay. We can aid the replay by exploiting the subjective views facility in DIVE: (Frécon et al., 2001). With this facility visual properties of the world are made dependent on the participant's name and role group. For example, this allows for a part of the world to be made invisible for a set of participants. We routinely use this mechanism for avatars that represent IPT participants. In an IPT, it is not usually possible to see one's own avatar. However an avatar should exist so that other participants in a shared session can see a representation of the IPT participant. This is simple in DIVE: the IPT participant's avatar is made subjectively invisible to themselves. We can extend this very easily. When a BIP is signalled, we make an otherwise invisible yellow ball attached to the feet of the avatar flash visible for two seconds. This ball is subjectively invisible to the IPT participant, but will be visible to any viewers of the replay. We also augment the avatar with two rays that indicate which way the participant is looking and which way they are pointing. Again these are subjectively invisible to the participant.

5 Pilot Trials

5.1 Methodology

On arrival for the study the participants were asked to sign a consent form that 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 they were asked to agree not to drive or operate complex machinery for at least 3 hours after the conclusion of the study.

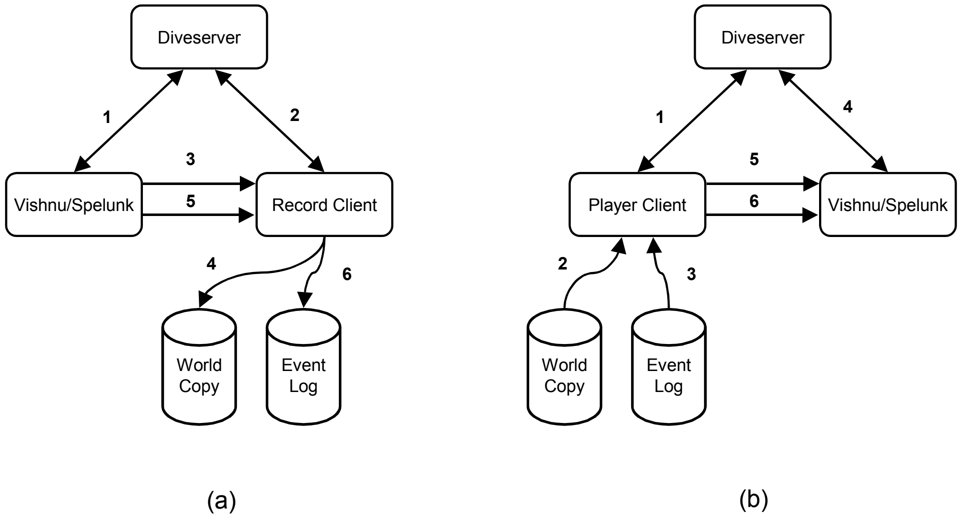

Figure 2: Configuration of DIVE processes for recording (a) and replaying sessions (b). The numbers are explained in the main text.

The participant was then given a questionnaire asking them to provide information about their background for demographic purposes. The participant was then given a short training to help in understanding the concept of BIPs. In this training they were asked to look at four Gestalt pictures and switch their focus from one perceived image to the other. An example picture was the classic visual illusion that can be interpreted as two faces or a vase. The transitions they experienced in the viewing exercise were equated to transitions they might feel to the real world of the laboratory while in the virtual environment. Throughout the study, BIPs were referred to as *transitions to real*.

The participant was then invited to step into the ReaCTor, don the head tracker and glasses, and hold the controller with joystick for locomotion and buttons to signal BIPs. They first saw a virtual training room containing three-dimensional numbers. The experimenter showed the participants how to move through the environment by 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. They were reminded to signal any *transitions to real* that they felt by pressing a button on the hand-held device that they used for navigation. At this stage, the experimenters left the participant in the ReaCTor and opened the virtual door leading to the street. Whilst the participant was being trained, a second experimenter started the record client. At the end of 3 minutes, the global lights in the street were switched off and the participant was assisted to remove the interface devices and step out of the ReaCTor. They then filled in a questionnaire. This included 5 questions from the Slater-Usoh-Steed (SUS) questionnaire (Slater, Usoh & Steed, 1994) and 6 from the ITC Sense of Presence Inventory (ITC-SoPI) (Lessiter et al., 2001). Typical questions, which would be answered on a seven point Likert scale include "There were times during the experience when the street became the reality for me." (SUS) and "I felt that I was surrounded by the displays" (ITC-SoPI). After the questionnaire they were also asked to reflect upon what, if anything, caused them to report BIPs. They were not shown the replay as an aid to reflection.

5.2 Display System

A CAVE™-like (Cruz-Neira et al., 1993) IPT system, a Trimension ReaCTor, was used to generate our IVE. The ReaCTor 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 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.

5.3 Software

We have used the one of the conditions from the previous street experiment as a basis for this trial (Vinayagamoorthy et al. 2004). An example scene is shown in Figure 1. The DIVE software, introduced in Section 4 was used to generate the world. In order to create a small, animated crowd of people, we used the PIAVCA

Figure 3: A subject signalling a BIP as an avatar approaches. The subject is represented by the large sphere and their hand by a small spheroid. Both have protruding rays. The timing of the BIP is represented by the appearance of a yellow hemisphere on the ground near the participant's head. From left to right: As the avatar approaches. Subject turns and signals a BIP. Another BIP is signalled as another avatar approaches.

(Platform Independent API for Virtual Character and Avatars) software. PIAVCA is a character animation library 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. 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. The avatars used were quite cartoon-like in appearance.

The world was created and textured in 3D Studio MAX. In the model used in this pilot, we used a world where textures were reused a small number of times throughout the world. In Vinayagamoorthy et al., (2004), a world with a larger variety of texturing mapping was an alternative in one of the experiment conditions.

6 Results

This was a pilot trial with only five subjects, but the results are encouraging. All of the subjects were male, aged 24-35 and studying at post-graduate or doctoral level in computer science. We have reports of BIPs in the same categories that were reported in a previous study: external, internal and situational (refer to Section 2.2). Out of the five participants one signalled no BIPs, two signalled only 2, one signalled 8 and one signalled 22. This range of numbers of BIPs is similar to that found in previous studies. When questioned afterwards, participants reported external causes for BIPs such as seeing the walls of the ReaCTor and hearing background noises. The latter is an example of something that could be controlled by more careful procedure: this world had no audio so it would have made sense to use noise-cancelling headphones. The former is an example of something that is trickier for IPT developers: occasionally participants notice the walls or the joints between them. One participant reported two situational BIPs but these were arguable cases. He reported feeling "strange as time went by" and "feeling uncomfortable" when it turned dark. These are, perhaps, appropriate responses to the environment, so they might be classified as noting changes in self-awareness rather than BIPs.

Most of the BIPs were related to internal causes. Here we can use the replays to study occurrences of BIPs and try to associate them with types of BIP. The main finding from observation is that participants will often fixate on a target before signalling a BIP. This might be interpreted as them trying to understand what they are looking at, or noticing it and deciding that it is out of place somehow. For example, subject 5, who signalled 22 BIPs, signalled a BIP on several occasions as the avatars walked past. Figure 3 from the replay shows how, as an avatar walk towards him, the subject turns towards the avatar, and signals a BIP as the avatar gets within about 2m. This behaviour repeats on a few more occasions: as the subject fixates on the avatar and the avatar approaches, he signals a BIP. When he is not fixating on an avatar, that avatar can pass by without causing a BIP to be signalled.

In contrast to this, subject 2, who only reported 2 BIPs, reported one of them when he couldn't get a response from one of the avatars. He "stalks" an avatar down the street, even passing through it, but then stands back and signals a BIP as it walks away (see Figure 4). This is interesting because it suggests that the subject was testing the environment to get a response and leads us to speculate that some level of interactivity is missing.

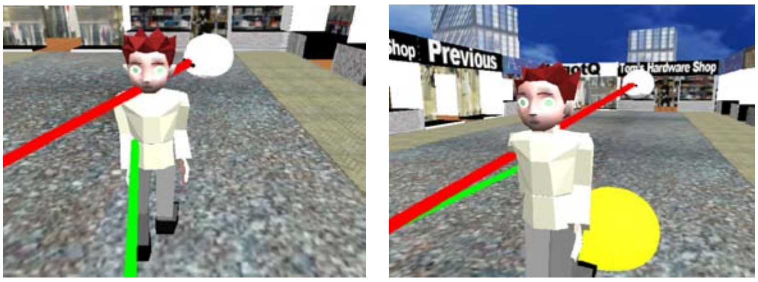

Figure 4: Left: A subject "stalks" an avatar down a street. Right: The subject signals a BIP after failing to get a response from the avatar.

Figure 5: Left and Right: Two different subjects signal BIPs when looking at the textured signs on shops.

One internal BIP that was reported by two subjects (3 and 5) was spotting a texture repetition. This is interesting as it supports the formative hypothesis in Vinayagamoorthy et al. (2004) which that study was unable to confirm conclusively: higher levels of texture realism would support a higher level of presence. In Figure 5 subjects 3 and 5 are looking at repeated shop signs when they signal BIPs.

Some of the BIPs are more difficult to associate to verbal reports afterwards. Some of the reports are quite vague, and refer to pervasive details. In Figure 6 we have shown two occurrences of BIPs from subject 5. One occurs when the subject has ventured into a shop. We can tentatively associate this with their reporting that they signalled a BIP because a room felt claustrophobic. This might be an appropriate response or it might be a facet of the presentation causing the room to feel smaller than it was. This subject also reported that a BIP was caused by far away objects looking blurry. Again we speculate by associating this with their looking at objects in the far distance. Note that the ray from the head in Figure 6 passes slightly above the true line of sight because pitch of the head tracker is not usually calibrated accurately for the ReaCTor display due to pitch not being that essential in order to estimate eye position accurately for the stereo renderings on the walls in the ReaCTor. The comment might have arisen because of the lack of detail in the far texture or because of resolution limits of the ReaCTor displays.

7 Conclusion and Further Work

We have presented the beginnings of a novel approach to evaluating the usability of VE systems: making full VE recordings of the participant and world behaviour and then replaying this augmented with a visualisation of the participant signalling breaks in presence. From a pilot trial we found a number of types of BIPs and some causes that might be common. In this particular world two common causes seem to be the lack of interactivity of the avatars and the repetitive nature of some of the texture mapping. Because of the range in number of BIPs, it is hard to claim that with five subjects that we covered a large number of types of BIPs, so future work needs to look at how many participants will be required to get a good range of usability problems.

There are some immediate improvements to the process that could be made. Heldal et al. (2005) have taken video of participants in collaborative IPT trials and used this to study interaction between participants. The record and replay of the session itself would make this type of analysis easier. Heldal's method mirrors a common process in HCI of retrospective commentary on an experience. Also a video would be a useful addition to the system record and replay process because it would allow us to start identifying potential external causes of BIPs such as noises. Another extension would be to extend the visualisation of the avatar to represent the IPT itself so that a reviewer could see

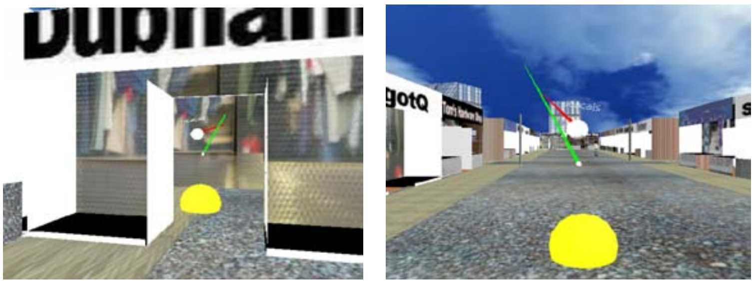

Figure 6: Left: Subject 5 signals a BIP when inside a room. Right: Subject 5 signals a BIP when looking at distant objects.

where the person was inside the display and detect certain effects such as tracking errors, wall collisions or perhaps participants looking directly at the corners of the screens.

Our immediate next step is to replay the session to the participant themselves. This will allow them to reflect on the experience so that they can attribute the BIP to a cause themselves. This will be facilitated by the ability to replay the session from a 1st person or 3rd person perspective and to immerse the participant in the IPT again if beneficial. Retrospective commentary on an experience has been shown to generate rich commentary, though obviously the ability of participants to reflect accurately will vary. Nielsen, Clemmensen & Yssing (2002) provide an overview of the issues with retrospective commentary and think aloud protocols in general. It would be interesting to compare record and replay with BIPs with a variety of think aloud protocols. However others have noted that think aloud might impose a high cognitive overhead (Preece, 1994). With the BIPs method we have strived to make the effort on the participant's part small so that it doesn't disturb presence too much. However think aloud protocols deserve exploration in the VE context as part of a formative usability process.

Acknowledgements

This work was supported by the UK projects, EQUATOR Interdisciplinary Re-search Collaboration (EPSRC Grant GR/N15986/01) and Advanced Grid Interfaces for Environmental e-science in the Lab and in the Field (EPSRC Grant GR/R81985/01). Thanks to David Swapp for support on the UCL ReaCTor, to Marco Gillies for the PIAVCA software and its integration into DIVE and to Amanda Oldroyd of BT Plc for supplying the cartoon-like characters.

References

Bowman, D., Gabbard, J.L. & Hix, D. (2002) A survey of usability evaluation in virtual environments: classification and Comparison of Methods. *Presence: Teleoperators and Virtual Environments*, 11(4), pp404-424

Bowman, D., Kruijff, E., LaViola, J. & Poupyrev, I. (2004) 3D user interfaces: theory and practice, Addison-Wesley

Brogni, A., Slater, M. & Steed, A. (2003) More breaks less presence. In *Presence 2003, 6th Annual International Workshop on Presence, 2003*

Cruz-Neira, C., Sandin, D.J., & DeFanti, T.A. (1993) Surround-screen projection-based virtual reality: The design and implementation of the CAVE. In *Proceedings of Computer Graphics (SIGGRAPH 1993)*, pp. 135-142.

Draper, J. V., Kaber, D. B. & Usher, J. M. (1998) Telepresence. Human Factors, 40(3), 354-375.

Frécon, E., Smith, G., Steed, A., Stenius, M. & Stahl, O. (2001) An overview of the COVEN platform. *Presence: Teleoperators and Virtual Environments*, 10(1), pp. 109-127.

Gabbard, J.L., (1997). A taxonomy of usability characteristics in virtual environments, MS. Thesis, Department Of Computer Science and Applications, Virginia Polytechnic Institute and State University.

Gaver, W.W., (1992). The affordances of media spaces for collaboration. In *Proceedings of CSCW'92*, Nov. 1992, Toronto, Canada, ACM Press, pp. 17-24.

Heldal I., Bråthe L., Steed A., Schroeder R., (2005) Interaction fragments influencing collaboration in distributed immersive virtual environments. To be published in: *Avatars at Work and Play: Collaboration and Interaction in Shared Virtual Environments*, Schroeder R., Axelsson A.S. (eds.), London, Springer-Verlag.

Hix, D., Swan, J.E., Gabbard, J.L., McGee, M., Durbin, J. & King, T. (1999). User-centered design and evaluation of a real-time battlefield visualization virtual environment. In: *Proceedings of IEEE Virtual Reality 199*, pp. 96-103.

Hodges, L., Anderson, P., Burdea, G., Hoffman, H., & Rothbaum, B. (2001). Treating psychological and physical disorders with VR. *IEEE Computer Graphics and Applications*, 21(6), 25-33.

Johnson, C. (1999) Evaluating the contribution of desktop VR for safety-critical applications. In *SAFECOMP 1999*, pp. 67-78

Kaur, K., Maiden, N., & Sutcliffe, A. (1999). Interacting with virtual environments: An evaluation of a model of interaction. *Interacting with Computers*, 11(4), 403-426.

Lessiter, J., Freeman, J., Keogh, E. & Davidoff, J. (2001) A cross-media presence questionnaire: The ITC-Sense of Presence Inventory. *Presence: Teleoperators and Virtual Environments*, 10(3), pp. 282-298.

Meehan, M., Insko, B., Whitton, M. & Brooks Jr., F. P. (2002). Physiological measures of presence in stressful virtual environments. ACM Transactions on Graphics 21(3): 645-652. (Proceedings of ACM SIGGRAPH 2002, San Antonio , Texas).

McGrenere, J., & Ho, W. (2000) Affordances: clarifying and evolving a concept. In *Proceedings of Graphics Interface 2000*, pp. 179-186.

Nielsen, J., Clemmensen, T. & Yssing, C. (2002) Getting access to what goes on in people's heads? Reflections on the think-aloud technique. In *NordiCHI*, October 19-23, 2002, pp101-110.

Parent, A., (1998). A Virtual environment task analysis workbook for the creation and evaluation of virtual art exhibits, Technical Report NRC 41557 ERB-1056, National Research Council Canada.

Pertaub, D-P., Slater, M. & Barker, C. (2001) An experiment on public speaking anxiety in response to three different types of virtual audience. *Presence: Teleoperators and Virtual Environments*, 11(1), 68-78.

Poupyrev, I., Weghorst, S., Billinghurst, M. & Ichikawa, T. (1997) A framework and testbed for studying manipulation techniques for immersive VR. *Proceedings of the ACM symposium on Virtual reality software and technology*, Lausanne, Switzerland, Septemer 1997, pp.21-28, ACM Press.

Preece, J. (1994), Human-Computer Interaction, Addison-Wesley, England.

Romano, D., Brogni, A., Steed, A., Slater, M. & Barker, C (2002) Agoraphobia research progress report, September 2002. Retrieved 20th February 2005 from http://www.cs.ucl.ac.uk/research/equator/papers/

Sadowski, W., & Stanney, K. (2002). Presence in virtual environments. In K. M. Stanney (Ed.), *Handbook of virtual environments: Design, implementation, and applications*, pp. 791-806, Lawrence Erlbaum Associates, Inc.

Slater, M., Brogni, A., & Steed, A. (2003) Physiological responses to breaks in presence: A Pilot Study. In *Presence 2003, 6th Annual International Workshop on Presence*, 2003

Slater, M., & Steed, A. (2000) A virtual presence counter. *Presence: Teleoperators and Virtual Environments*, 9(5), 413-434.

Slater, M., Usoh, M. & Steed, A. (1994) Depth of presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, 3(2), 130- 144.

Salkovskis, P.M. & Hackmann, A. (1997). Agoraphobia. In Phobias, A Research Handbook of Theory Research and Treatment, G.C.L. Davey Ed., John Wiley & Sons Ltd, Chichester, New York, pp 27–61.

Steed, A., Mortensen, J. & Frécon, E. (2001) Spelunking: experiences using the DIVE system on CAVE-like platforms, In *Immersive Projection Technologies and Virtual Environments*, (B. Frohlicj, J.Deisinger, and H.-J. Bullinger Eds. Wien: Springer-Verlag), pp. 153-164.

Steed, A. & Tromp, J. (1998). Experiences with the evaluation of CVE applications. *Proceedings of Collaborative Virtual Environments* 1998.

Tromp, J., Steed A. & Wilson, J. (2003) Systematic usability evaluation and design issues for collaborative virtual environments, *Presence: Teleoperators and Virtual Environments*, 12(3), 241-267.

Usoh, M., Catena, E., Arman, S., & Slater, M. (2000). Using presence questionnaires in reality. *Presence: Teleoperators and Virtual Environments*, 9(5), 497-503.

Vinayagamoorthy, V., Brogni, A., Gillies, M., Slater, M. & Steed, A. (2004) An investigation of presence response across variations in visual realism. In *Presence 2004: The 7th Annual International Workshop on Presence*.

Whitton, M. C. (2003) Making virtual environments compelling. *Communications of the ACM*, 46 (7), 40-47.

Zimmons, P. & Panter, A. (2003) The influence of rendering quality on presence and task performance in a virtual environment. *Proceedings of IEEE Virtual Reality*, pp. 293-294.