UNCHECKED PREPRINT

Forthcoming in Presence: Teleoperator and Virtual Environments, 2005
I. Heldal, A. Steed, M. Spante, R. Schroeder, S. Bengtsson, M. Partanan

Successes and Failures in Co-Present Situations

Abstract

Virtual environments systems based on immersive projection technologies (IPTs) offer users the possibility to collaborate intuitively in a three-dimensional environment. While considerable work has been done on examining interaction in desktop-based collaborative virtual environments (CVEs), there are currently no studies for collaborative interaction using IPTs.

The aim of this paper is to examine how immersive technologies support interaction and to compare this to the experience with desktop systems. A study of collaboration is presented where two partners worked together using networked IPT environments. The data collected included observations, analysis of video and audio recordings, questionnaires and debriefing interviews from both IPT sites. This paper focuses on the successes and failures in collaboration through detailed examination of particular incidents during the interaction. We compare these successes and failures with the findings of a study by Hindmarsh et al. (Hindmarsh, Fraser, Heath & Benford 1998) that examined object-focused interaction on a desktop-based CVE system.

Our findings identify situations where interaction is better supported with the IPT system than the desktop system, and situations where interaction is not as well supported. We also present examples of how social interaction is critical to seamless collaboration.

1. Motivation and Aim

Collaboration at a distance has long been an important research goal of networked or collaborative virtual environments (CVEs). The recent emergence of immersive projection technology (IPT) displays such as the CAVE™ (Cruz-Neira, Sandin & DeFanti 1993), promises a new generation of tele-presence and tele-collaboration systems that are more natural to use than desktop-based CVEs or video tele-conferencing systems. Whilst there have been many demonstrations of CVEs using IPTs (Leigh, Johnson, Brown, Sandin & DeFanti 1999) there is little knowledge about which tasks IPT CVEs are more suitable for collaboration than desktop–based CVE systems. Moreover, it would be useful to know more about the specific advantages and drawbacks of collaborating using IPT systems rather than desktop-based technologies.

The complexity and variation in technologies and the huge range of applications makes comparisons between desktop and IPT systems difficult. Desktop CVEs are in widespread use for games and distributed work, whilst IPT systems are relatively rare and are typically used in more specialized applications that do not undergo such rigorous usability engineering. There would be significant benefit to identify aspects, for example in which context and under what conditions one system is more suitable than others. Since object-focused work in distributed virtual environments is quite common, this motivates us to examine object-focused interaction (Hindmarsh, Fraser, Heath & Benford 2000). We might expect to find problems in performing object-focused interaction that are common to both types of systems, and others that are unique to one display system. Understanding the similarities and differences will provide some general criteria for creating effective and enjoyable CVE applications.

We report on a study where participants had to interact with the virtual objects in the space and with each other. Participants had to solve five tasks; here we concentrate on the two most object-focused ones: solving a Rubik's-cube type puzzle, and using Lego-like objects to build a sculpture together. The other three tasks were solving a murder mystery in a set of rooms, a word puzzle task, and a landscape orientation task, see Steed et al. (Steed, Spante, Heldal, Axelsson & Schroeder 2003). The findings of this paper contribute to the understanding of interaction in CVEs by identifying possible patterns of successes and failures. The focus is on differences between immersive CVEs and desktop CVEs in collaboration and object-focused interaction.

We also suggest factors that can be considered in usability to improve the design of future CVEs. It would be useful, for example, to study how different types of IPT systems impact on collaboration with especially in the case of distributed environments. There are also several subjective factors that influence interaction in CVEs: the individuals ability and their willingness to collaborate, and also the actual usage context. By knowing more about individual factors, the context of the collaborative applications, and available technologies, it should be possible to identify critical issues that influence the collaborative interaction.

This paper begins with an overview of the background, concentrating on the findings of a study by Hindmarsh et al. (1998) about problematical interaction scenarios for desktop systems. Section 3 details the study design and Section 4 reviews the findings of the Hindmarsh et al. study as applied to the IPT context. The four following sections describe successes and failures for object-focused collaboration scenarios for immersive systems. Section 5 includes examples based on design issues connected to the IPTs' physical structure and the projected environments, and Section 6 to the actions one can perform within one's own environment. Section 7 contains issues related to collaboration and partner(s) and Section 8 is about the impact of the users' embodiments. Suggestions for how the design of the CVEs can be improved to support better interaction are discussed in Section 9. Section 10 contains the main conclusions.

2. Background

Many issues like presence, performance, intuitiveness or leadership have been identified as important in CVEs (Draper, Kaber & Usher 1998; Steed, Slater, Sadagic, Tromp & Bullock 1999). Studies have also demonstrated that each of these issues might depend on the technology. For example, it has been shown that task performance in networked IPTs can be higher than in an IPT system connected to a desktop system for puzzle solving tasks (Schroeder, Steed, Axelsson, Heldal, Abelin, Wideström, Nilsson & Slater 2001). However, desktop systems can be more effective than the immersive environments for problem solving when visualizing large molecules (Heldal & Schroeder 2002) or for educational uses (Youngblut 1998). Other studies have demonstrated that certain features – such as presence, performance, and immersion – may be more closely associated with one specific type of VE than with another (Slater, Linakis, Usoh & Kooper 1996; Bystrom & Barfield 1999; Sallnäs 2002).

Additional questions concern the level of realism of representation of self (the embodiment), objects, and surrounding space. The aim of virtual reality systems is not necessarily to reproduce physical artifacts or to achieve graphical realism (Stanney 2002), nor is it to reproduce natural interaction in these environments for all situations (Bowman 1999). A realistic embodiment, for example, may support better collaboration, but simple embodiments can also support different types of interaction (Bowers, Pycock & O'Brien 1996b; Bente & Krämer 2002).

While the number of CVE applications is increasing, there are still many problems regarding, for example, user interfaces (West & Hubbold 2001), the flow of interaction (Tromp, Steed & Wilson 2003), and communication modalities and social interaction (Schroeder 2002). Since all CVEs have to consider interaction via technology, a first consideration is the impact of the user interface. Interaction in human-computer interaction is related to how the user can communicate with the system, i.e. the interaction technique (Dix, Finlay, Abowd & Beale 1998). Indeed, interaction techniques have a strong influence on the collaboration in multi-user VEs by affecting how individuals act to reach their goals and disturb the flow of social interaction (Steed et al. 2003). The requirements for social interaction can also make certain demands on the infrastructure of the application. This is shown by Tromp (2001), who also suggests taking into consideration some general collaboration patterns for designing social spaces when building CVE applications.

There are significant differences between interaction in desktop systems and IPT systems. The key difference is that with an IPT system, the tracking of human body contributes to making the interaction more intuitive because actual body posture and gesture are conveyed immediately rather than having to be expressed through a user interface. We will make comparisons between our experience with collaboration on IPTs with a similar study of desktop environments performed by Hindmarsh et al. (1998). In what follows we refer to this study as the "Hindmarsh study".

Hindmarsh et al. (1998) studied how people collaborate in desktop CVEs and they present observations to illustrate both the problems raised when the technology got in the way and disturbed the user but also situations where the users were able to collaborate successfully. The reasons for disturbed interaction are mainly in visual discontinuities during activities caused by the desktop screen. The Hindmarsh study calls disturbed interaction “fragmented interaction”.

The study involved six trials of two participants and two trials of three participants who spent 10 minutes to get used to the system and half an hour to perform a furniture arrangement task. During the study they examined the role of the embodiments in relation to the users interaction and manipulations with the objects. With a desktop system, the embodiments, objects, or spaces can seldom be presented in a way such that collaborating partners do not experience ambiguities or disturbances. The principal issues were the fragments due the narrow field of view, which forced participants to compensate for the problems by explicitly describing actions and phenomena, and reduced peripheral awareness together with reduced opportunities to co-ordinate actions.

On the basis of these three problematical issues, they identified four key limitations that are typical of desktop systems and are likely to result in fragmented interaction. The first limitation is the limited field of view (FOV), the second is the lack of information about others’ action which is also related to the problem of limited FOV: people working together with the desktop system can be easily disturbed by not seeing their partners and their partners’ actions together in the same view. The third limitation was slow, clumsy movements and slow applications. Finally fourth, there was the lack of parallelism for actions, i.e. people could not do things simultaneously without problems.

These key limitations suggested two possible ways for overcoming them: first, making more explicit representations of others' actions, and second, to implement mechanisms for coordinated navigation.

This paper examines which of the limitations and suggestions discussed for desktop CVEs are applicable – and how they are overcome in collaboration between two users in networked IPTs.

3. Study Design

Our study followed a similar format to that used by Hindmarsh et al. (1998). We studied six pairs of participants in detail. In each pair one person was physically located in an IPT system in London and the other was in an IPT system in Gothenburg. Participants undertook a series of five collaborative tasks over the course of a day. We will report on just two of the tasks that were object-focused and highly collaborative.

3.1. Method

Considering the fragments and limitations defined by Hindmarsh et al. (1998), we first present some key successes and failures in object-focused interaction scenarios in networked IPT CVEs. This is followed by a discussion of the role of the embodiments in this immersive setting. Next we discuss the suggestions for overcoming the limitations of desktop systems made by Hindmarsh et al. (1998), how these have been overcome by developments in virtual reality technology since 1998, and to what extent their ideas are nevertheless relevant for the networked IPT CVEs.

We used a combination of questionnaires (with Likert-scales and written answers), debriefing interviews and video and audio recordings of the sessions. The audio and video recordings were analyzed independently by four of the authors. Each analysis used the categories suggested by the Hindmarsh study. The analyses were done separately to ensure broad coverage of the issues and to enhance validity in the interpretation of the data. The authors then met and brought together typical examples regarding successes and failures in collaboration. The examples we present below are verbatim transcripts.

3.2. Users

The study was exploratory in nature and we consider 6 pairs in this paper and how they collaborated on two different applications in a virtual environment. They were 10 males and 2 females with various backgrounds (two of them were students). Most of them were Swedish, but they were asked to communicate in English for the benefit of the experimenters in London. They all speak English fluently. Apart from the third pair (2 males, who were stopped because both partners experienced nausea), each pair spent at least 210 minutes in the immersive environment. They started with approximately 15 minutes to become familiar with the system. This was followed by the first task, a “Rubik’s Cube Puzzle” application (20-30 minutes). Following this they undertook three other tasks, and had several breaks. The final task was “Modeling World” (50-60 minutes). We will only report on the first and last tasks as these were highly object-centered.

3.3. Applications

This experimental study includes two different applications where the subjects had to manipulate cubes and shapes.

1. Rubik's Cube Puzzle. The task was to do a small-scale version of the popular Rubik's cube puzzle with eight blocks with different colors on each side so that each side would have a single color (i.e. four squares of the same color on each of the six sides, see figure 2).

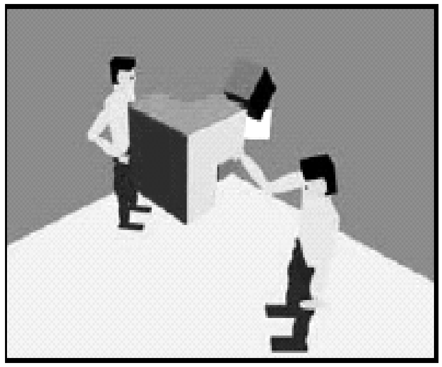

Figure 2. The Rubik's cube puzzle

2. Modeling World. This environment contained 96 shapes (square blocks, cones, etc.) in six different colors. The subjects were told to make a building or model of a building to be entered into an architectural competition (see figure 3). They had to use at least three colors and the building had to be a single object. The result was to be their joint 'architectural masterpiece'.

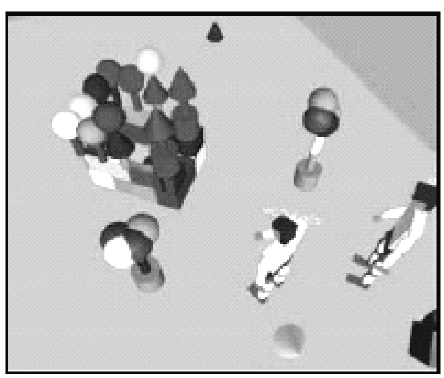

Figure 3. Modeling World.

The first application was quite structured in that it had a definite goal. The second was much more open-ended. For the first application the virtual space was small, being approximately the same size as both IPTs, that is 3m x 3m. Participants did not have to move about in the space. The virtual space of the second application was much larger, 14m x 14m. Thus subjects had to navigate and orient themselves more for this application.

3.4. Interaction

Each subject held a wand that effected locomotion and manipulation. Locomotion was restricted to the ground plane. The wand was tracked and a virtual hand was drawn in the same position as their real hand. The wands at each site were slightly different, in the IPT in London the wand had four buttons and an isotonic joystick, and in the IPT in Gothenburg it had three buttons and an isometric joystick.

In both IPTs, locomotion was performed with the joystick. Locomotion was nevertheless head centric. Horizontal rotation was around the head position and translation was forward and backward along the direction of gaze.

To manipulate the objects participants needed to touch the virtual hand to the virtual object that they wished to manipulate. They would grasp by clicking and holding the left-most wand button and moving their hand. To carry the virtual object, the subjects first had to grab the object and then use the joystick to move. The design of both wands meant that two hands were required if the object needed to be moved while locomoting at the same time.

For the Rubik cube application the main interaction was manipulation in order to put together the overall cube and the pairs had visual feedback about how close to the end result they were. For the Modeling World application, they had to navigate, orient themselves and manipulate the objects in order to build something that they both liked.

3.5. Technology

The IPT system at Chalmers was a five-sided (no ceiling) 3m x 3m x 3m TAN VR-CUBE . The application was run on a Silicon Graphics Onyx2 Infinity Reality with 14 250MHz R10000 MIPS processors, 2GB RAM and 3 graphics pipes. The participants wore CrystalEyes shutter glasses. A Polhemus magnetic tracking device tracked both the glasses and the wand. The rendering performance was at least 30 Hz in the two applications discussed here.
The IPT system at University College London was a four-sided Trimension ReaCTor with a floor of 2.8m x 2.8m and three 2.8m x 2.2m walls. It was powered by a Silicon Graphics Onyx2 with 8 300MHz R12000 MIPS processors, 8GB ram and 4 Infinite Reality2 graphics pipes. The participants wore CrystalEyes stereo glasses. The head and wand were tracked by an Intersense IS900 system. Rendering performance was at least 45Hz in the two applications discussed here. Both applications were implemented in a customized version of the Distributed Interactive Virtual Environment (DIVE) system (Frecon, Smith, Steed, Stenius & Stahl 2001; Steed, Mortensen & Frecon 2001). DIVE supports collaborative environments by way of a partially replicated shared database that is maintained by a multicast event protocol. DIVE has an abstract model of interaction that allows a wide range of locomotion and manipulation to be implemented.

Each participant was portrayed to the other by the use of a simple avatar with a jointed left or right arm. The participant could not see their own avatar, except for a virtual hand drawn in the same position as their physical hand. Although local tracker updates perform at the fastest rate provided by the tracker driver, updates to the remote avatar are only sent at 10Hz to avoid congestion. The network latency between the two sites was approximately 180ms.

The subjects could talk to and hear each other using a wired headset with microphone as well as earphones. We used the Robust Audio Toolkit (RAT) for audio communication between the participants.

4. Successes and Failures in Networked IPT CVEs

In this section we review the fragmented interaction and limitations discussed in the Hindmarsh study as applied in the context of IPT systems. We give a top-level review of interaction fragments that still cause problems and those have been overcome by using IPT systems because of the nature of the technology. In the sections thereafter we will then give detailed examples from the experimental study to explore these. We are also looking for general patterns of successes and failures that can characterize object-focused interaction for the IPT situation.

Hindmarsh et al. (1998) suggested that some of the fragmented interaction observed for desktop systems could be alleviated by using immersive environments. It is true that a rendering performance greater than 30 Hz that permits near real-time interaction, tracking system that allows more intuitiveness, and the larger FOV all contribute to smoother interaction. However, it is interesting to examine how the limitations and suggestions observed for the desktop systems are modified in IPT systems since there are situations of fragmented interaction for these as well. The first limitation reported for desktop systems, the limited FOV is obviously not a severe issue for IPT systems since an IPT has a FOV close to that of the user's vision. However, questions remain regarding the surrounding virtual environment, such as how many walls an IPT system should have for effective collaboration. The second limitation, lack of information about other's actions, is also alleviated somewhat by the surrounding nature of the IPT. However there are still problems due to the representation of the other user. For example, some actions, such as subtle movements and eye gaze, are not transmitted by the medium. The third limitation, slow clumsy movements and slow applications, still occur with IPTs, as does the fourth, the lack of parallelism for actions. However, as we will show, these limitations take a different form in IPTs.

For IPT systems, these limitations can be translated into four somewhat different types of potential difficulties: the first is the problem caused by the IPTs not being completely immersive environments because they are usually not totally surrounding. The IPT system in Gothenburg has no ceiling and the one in London also has no ceiling and only three surrounding walls. The second type of difficulty is a result of the existing physical form of the IPT systems and the parts of the other system that are not visible. For example, experiences can be disturbed by the dark corners of the IPT, the physical form of the glasses or joystick, and the tethering of wires. Moreover, immersive environments often require a helping person to monitor the system the system, assist with the technology, and ensure the safety of the participant. The presence of another person may occasionally interrupt the subject (Bowman, Gabbard & Hix 2002). The third type of difficulty concern the participants' use of the medium, and misunderstandings about which of their actions are conveyed at the other end. Here we can mention gestures, for example when subjects used both of their hands to indicate or describe something even though only one hand was tracked. The fourth type of difficulty is concerning avatar representations. In this case we will present observations about how the virtual representation of the users enables effective collaboration. The following four sections tackle each of these in turn; Section 5: Issues related to the (physical and projected) environment, Section 6: Issues related to actions, Section 7: Issues related to collaboration and partner(s), and Section 8: Issues related to embodiment.

5. Issues Related to the Environment

In this section we describe how the design of the real and virtual environments has an impact on interaction. The IPTs used in this study are not completely surrounding. Thus we might encounter a few instances of the first limitation from the Hindmarsh study, the limited horizontal FOV. We might also expect to encounter the second limitation, missing information about others' action, caused by the limited FOV. However, for this networked setting the incomplete surrounding projection did not cause major observable problems.

In Gothenburg the virtual environment was projected on the floor and on four walls while the London IPT had projection on the floor and on three walls. In the analysis of video recording we could see that there were occasions when a person in London facing into the IPT made a backwards glance and did not see the virtual model because there was no wall behind her or him. However in the next second he or she was looking forward again and working further without comment. Example 1 shows a situation when the person in Gothenburg disappeared from the sight of the person in London. The Gothenburg participant locomotes directly through the representation of the London participant who was facing the middle wall. When the person in London looked behind himself, he did not see anything or anybody because of the missing wall and projections in that direction. Perhaps the most severe case is reported in example 1, yet the participants did not consider this problematic.

L (London participant) is standing in the middle of the Cave, with his back to the camera. G (Gothenburg participant) is facing him. G is trying to find out how to move.
G: "Ok, how can I do that...If I move like this...Ok!"
When G succeeds he goes straight through L and disappears behind L, where the 4th wall would be if there was one. L turns his head, first to the right, then to the left, looking for G and asking him:
L: "Uh, are you, where are you now?"
G: "Yes, I'm right behind you...Huhuhu..." [*sounds like a ghost*]

Example 1. An incident of one user not seeing the other because of limited FOV.

The second type of potential difficulty concerned disturbances with other parts of the physical construction of the IPTs. We saw no severe disturbances due to the corners, walls or how the projected image varied in brightness between the corners and the center of the wall. However, some subjects in Gothenburg occasionally complained about blurred projections close to the floor.

We have to stress here the impact of the task characteristics on the representations and on the collaboration. Both applications used an open virtual space and objects placed around this space. There were no virtual hindrances or additional obstacles in the space that restricted the view. The representations were simple cubes and Lego like shapes. For these types of applications, the projection was enough for collaboration. However an application for house building, or with more and bigger objects for example, or when people would need to build up walls in all directions and when their embodiments might be shielded from each other – it might be different. In the Modeling World, for example, couple number 6 built a house with a roof. The person from Gothenburg went into the house and she wished to see the roof from inside by looking upwards, but looking upwards she will only have been able to see the physical parts of the IPT and the physical environment.

According to the Hindmarsh study, hidden perspective problems are those problems that are caused by others' embodiments, or activities that cannot be seen because of the structure of the physical environment or the virtual domain. During all the sessions in the immersive environment we observed this only in one case, when a subject was out of sight. The following example will illustrate how it can be difficult to show an object when a person holding it is not in the view. In this case the portal on entranceway that allows travel to another virtual world was located behind the person so he could not see this portal at first. The person from London, who was not in view for the person from Gothenburg, intended to grab an object, but accidentally he grabbed the portal instead. He was surprised and showed what he grabbed to his partner (they called the portal the "hub").

L: "Look what I've got!" [*pause*]
L: "he, he. Look what I've got! He, he..."
G: "Where?"
L: "No, no...look behind you!"
G: "What?"
L: "Look what I've got, what I'm holding...The hub...the..."
G: "Wha?"
L: "This one."
G: "What are you holding?"
L: "Humm...I'll show you." [*L starts to back off so that the portal will be in front of G.*]
G: "Because, I can't see what you are holding"
L: "See what I'm holding?"
G: "Ha, ha. OK!" [*G sees what L is trying to show him.*]
L: "He, he"

Example 2. The person from London has taken the portal.

This situation could be interpreted as similar to situations that were highlighted in the Hindmarsh study when participants could not assess the others' view. The Hindmarsh study described an example when one person was talking about a fireplace but the other could not recognize the object "fireplace". In the IPT CVEs similar misinterpretations could happen. However, another reason for this type of misunderstanding could be to do with concentrating on one’s own work and not really being available for social interaction; i.e. the focus of attention is directed towards object manipulation to such an extent that it takes all the individuals’ attention.

Another disturbance caused by the design of the environments is the missing information about the system status (changing batteries for the glasses, problems with devices, cables, etc.) For all applications and in both IPT systems, an instructor was present at beginning of the application to help the subject with the devices and explain the task. While the instructor was in the IPT, the participant in the other IPT would not necessarily be aware that there was an instructor. Occasionally the participants would tell the other person that they were receiving assistance. At other times, they would overhear a conversation between their partner and their partner’s instructor. Since instructors were usually only involved at the beginning of the trials, these peripheral conversations did not usually interfere with collaboration because both sites would be busy preparing at the same time. However in certain situations, e.g. helping with a device because of discomfort, caused temporary misunderstandings.

The verbal interaction provided the most important channel to describe those activities that partners cannot see. Because our study took place over the course of a day when the subjects worked through long hours in the IPT systems, there were several interruptions to replace batteries and the like. A person from the other end of the CVE could not see interruptions of this kind. The only way for the partner to know about these problems were via audio. If the audio was problematical at one site the partner did not know what is happening. To compensate for this lack of feedback, most of the subjects used talk to inform their partner if they had problems with the technology. This did not seem to cause any severe disturbances for the subjects. However, one pair seldom informed their partner when things like this happened and this was considered one explanation of their rather poor collaboration.

The third type of difficulty concerns the failure to convey indicative gestures. At this point we are only considering how the environments convey gestures - not how gestures are tracked. In the Hindmarsh study, as in other desktop CVEs, the facilities for making gestures were crude. However even in an IPT where gestures are easily expressible with the trackers, they can still be missed. For example, occasionally people used the non-tracked hand to indicate directions or referred to objects in the virtual environment by their non-tracked hand. Usually subjects realized immediately that they had pointed with the non-tracked hand and swapped it to the tracked hand to continue their interaction. This is further discussed in Section 6.

6. Issues Related to Actions

In solving the “Rubik’s cube” a person had to manipulate objects intensively in order to solve the task, and in the “Modeling World” they had to manipulate, navigate and orient themselves at the same time. To find and follow a correct strategy was more important for the first application than the second, which was open ended. In this section, activities like manipulation, navigation and orientation are examined. These are related to the third and fourth limitations caused by slow movements and lack of parallelism for the desktop CVE systems in the Hindmarsh study.

Because of the real-time nature of the IPT CVEs and the tracking system, people did not have to perform several physical commands or perform parallel actions, such as moving and simultaneously pointing around, grasping and looking, or grasping and moving – as with the desktop systems. These activities can be more intuitively performed in IPT CVEs. Nevertheless, the virtual hand is still clumsy to handle in these systems. To grab an object a person had to (1) put the virtual hand in the object – that could be quite difficult in the beginning since the movements of the joystick and how the virtual hand followed the joystick had to be gotten used to (2) push a button, and (3) move it in accordance with a particular trajectory that one has thought about.

A very important and frequent form of interaction was coordination. In this case subjects often used gestures to refer to items. There were several cases, as in the next example, when one person used the non-tracked hand to point with, but in the next second changed to the tracked hand. This kind of behavior generally did not create a particular problem for the other person; however, it has to be mentioned since it happened quite often and occurred throughout the sessions from the beginning to the end.

L and G were working together with the puzzle, standing almost opposite to each other. G asks L what colors she has got on her side and L answers and points with her non-tracked and tracked hand, and again with her non-tracked and tracked hand.

G : "So...what color do you have on on thesssse...the other side
there?"
L :"Oh, the thing is...ehhh...Well, this...[*points with non-tracked hand*]
is blue here, then we've [*points with the tracked hand*]
got white here and here [*points with non-tracked hand*]
it's orange. However, ahmm [*sits down and looks at the cubes from beneath*]
...we can't leave the two [*points with non-tracked hand*]
like that because we've got yellow and ahmmm...yellow [*points with tracked hand*]
here and red [*rises, stands up*]
on the other side so that's already violated."
G:"Aaah!"

Example 3. Alternating gestures with the non-tracked and tracked hands.

Many subjects realized immediately if they pointed with the non-tracked hand and changed it to the tracked hand to continue their interaction.

Some comments from the subjects showed that they were aware of the necessity to use the tracked hand to point if they wanted the other person to see their gesture. One example is when couple number 5 is working on the Rubik's Cube puzzle and the person in Gothenburg wanted help from his partner to move an object. Trying to get his partner's attention, he snaps with his fingers and also points with his non-tracked hand. He soon realizes that this gesture cannot be seen and verbally comments on that .

> G: "Here, here...against me...Do you see my...Of course, you can't see my hand...Damm, I'm waving with my real hand...Hmmmm".
> *L makes G aware of the need to use the tracked hand.*
> L: "Sure, yes...If you point with the joystick hand I'll see it!"

Example 4. Using the non-tracked hand to point.

Pointing with the non-tracked hand, and the fact that the gestures were not transmitted through the media caused little disturbance, but perhaps the more interesting point is that people made the same mistakes over and over again, and in the course of several hours. It is hard to make any generalizations about failure to adapt since the subjects only experienced the environment over the course of one day, but it is worth mentioning that people adapted to other technical difficulties such as watching out for the cables that were in the way and handling the audio easily and using the joystick intuitively.

We also have several examples of misunderstandings when one person was pointing (correctly) to an object but the partner thought that the reference was to another object nearby. These situations did not cause major problems because they were solved quickly, possibly as a result of the nature of the application and the intuitive use of gestures. In contrast to the Hindmarsh study, a function allowing pointing at a distance was not needed, (they implemented a laser-type pointer). In our study, the virtual workspace was not very big and the amount of virtual objects quite limited. For a very large construction space or if there are a large number of similar objects, it might nevertheless be useful to have a method for pointing at a distance.

7. Issues Related to Collaboration

Subjects could understand their partner's strategy for approaching the problem better if they saw and also heard the others reasoning at the same time. The combination of physical interaction with speech was an aid in problem solving. Subjects were often seen to use two different channels (verbal communication and non-verbal communication using indicative gestures) to make a strategy explicit and demonstrate it quickly.

In the Hindmarsh study subjects often similarly needed to use social interaction to compensate for some ambiguity in the situation. They called this 'compensating interactions'. The relevant situations here were mainly when people did not see an activity clearly or had difficulties in performing composite actions (for an example moving and grasping). Therefore they verbalized instead. The Hindmarsh study also found that on many occasions problematical interaction was compensated for by more explicit talk - for example, pointing at objects was more difficult to perform than describing it in words.

In the IPT situation, subjects often used verbal confirmations to establish a mutual understanding of where objects were or where they should be put in relation to the other's embodiment or to other virtual objects. For example, they helped each other *both* by pointing to a cube and describing its colors or its position in relation to the embodiment of the other person: “See the blue cube on your left side”.

However, in contrast to the Hindmarsh study, we found that instances of such utterances were usually not compensating for ambiguities, but rather they supported the general flow of conversation. Also, subjects used social interaction not *instead* of pointing but in parallel with it. This is probably because in the desktop systems pointing would take a relatively long time, whereas in the IPT systems it is a simple gesture. (It can be noted that on a desktop system, pointing involves several uses of the keyboard and mouse that take a longer time.)

In general, we observed that in the IPT system it is relatively easy to point, and it is easy to time pointing to coincide with verbal emphasis and give a rapid indication of one’s focus. Even though there are misunderstandings, as could be seen in example 2, people could solve these quickly. The subjects rarely needed to use verbal communication to make the implicit explicit. Rather they seem have been expressing their intention simultaneously with talk and gesture in order to make their intention clear.

Example 5 shows the importance of understanding - or trying to understand - the intentions of the other person , and the good will resulting from the attempt. The pair in the following example had problems to reach a solution but at least they tried to do it together, and so they maintained the social interaction by helping each other.

> L: "Ok what's here...can have...yellow...green...two red sides here...”
> G: "and blue"
> G: "perhaps there...eh...ah...do you have a...white side on that cube your are holding?"
> L: "I have a white one here"

G: "Should we try..."
L: "Yeah, it is white on one side here."
G: "Put them..."
L: "white on...Yeah, Ok...Oh no. I can't...I can't here at least. Because then I would...Can you swap them, because I've got black on my left side?"
G: "OK"
L: "I put it there and you put yours..."
G: "Yeah"

Example 5. Solving the task together.

Solving the tasks there is the important issue how the couples can *collaboratively* work on the problem, how they understand each other's strategy, and how they help each other (Heldal 2003). Example 6 illustrates a subject asking his partner for a missing piece during collaboration.

L: ”Which side do you want to have yellow at?”
G: ”Yellow towards me here.”

Example 6. Helping each other.

The next example also shows a couple helping each other, but here the subject from Gothenburg has difficulties in grabbing objects (example 7). The partner from London asks if he should help, but he also fails at the first attempt. After that, the person from London succeeds and hands it to his partner from Gothenburg.

L: "Should I help you?"
G: "Yes, please!"
L: "Wait, maybe once it 's in another position you can well, this one, I can't grab this one either."
L tries to give G an object but fails to grab it.
L: “Try to grab this one.”
L tries with another object and succeeds in grabbing it.
G: "OK"
L: "Try this one..."
L hands G the object.
L: "This one is a lost case". [*L is referring to the first object he tried to grab*.]
G: "Eyaä..."

Example 7. Helping the partner.

This helping activity is very significant. One limitation in object-focused manipulation with the IPT CVE system is that two people cannot hold the same object at the same time. That means that to give to each other an object, one person has to let it go and then the other pick it up. Handover would be almost impossible in a virtual environment with simulated gravity. The following example 8 wouldn't be possible if gravity would have been implemented; it was only possible because the cube hung in the air during the time between the subject from Gothenburg letting go and the subject from London taking it. This limitation of the system can be seen as a success for problem-solving because people believe that they can hand over cubes to each other in this way.

> *G and L are standing side by side; G is to the right, working with the cubes.*
> L: "Do you have green and orange somewhere?" [*G looks for it amongst the cubes he has on his right side.*[
> G: "Aah, I think I saw it somewhere here...Hm, green and yellow... Uh..."
> *G finds the cube he is looking for, and hands it to L who takes it.*
> G: "Here we have green and orange...Um, orange. Here you are."

Example 8. Handing over the cubes to the partner.

In the study there was one pair that had problems collaborating in the beginning. They also understood the technology and the task at hand differently and in addition they had language problems. Neither had English as their native language and the person in London had difficulties to understand the strong Swedish accent of the person in Gothenburg. For this study we do not have statistically significant results about performance, but it can be mentioned that this couple did not solve the puzzle during the given time (40 minutes). However, for the puzzle application we can compare this result with another study where statistically significant results were calculated for 22 couples who solved the puzzle in the same networked CVE, with much better results (mean time 8,00 minutes) (Schroeder et al. 2001). In other world, it seems likely that the poor performance of this couple had a lot to do with their poor collaboration.

Another observation here is the “gluing” function of small social phrases. There are several examples when subjects were working independently, but from time to time they would ask each other “What are you doing?”, “Everything is all right?”, “Are you there?”. This, together with the possibility to quickly take a glance at one’s partner (as in example 9) contributed to the maintenance of the flow of interaction.

> *The third couple is working together on the modeling application. The subject from G is in the background picking up an object. His partner passes G in the foreground at a distance, carrying an object. L backs up a bit, quickly turns his head to glance at G in the background, and then continues working. None of them says anything during this episode.*

Example 9. Subjects glancing at each other to coordinate.

In the example above, the enlarged FOV together with the coordinated actions between embodiments and objects contributed to collaboration and control over the situation. The activities in the Modeling World application were fluent, seamless and the people collaborated well on building with the colored shapes. For this application the subjects often worked in parallel, quite silently, but they also followed each other’s movements during the time, as example 9 shows. Similar quick moments for the Modeling World could be observed throughout the sessions. In the course of collaboration, people would sometimes take a glance at their partners and continue without necessarily making evident that they have seen their partner.

Considering how closely the problem solving has an impact on object-focused collaboration, it can be seen that the limitations of not being able to see the other’s activities and the parallelism of partner’s actions are interrelated.

8. Issues Related to Embodiments

The Hindmarsh study argues for the necessity of humanoid - like embodiments as a part of the virtual environment in order to support interaction with objects. Their argument follows the line taken by Bowers et al. (Bowers, O'Brien & Pycock 1996a; Bowers et al. 1996b) that embodiments are necessary to relate actions within a CVE to the users' intentions. In this section we extend this to discuss the importance of the humanoid – like shape to support object-focused interaction in IPT CVEs.

The participants in the IPT system had a nearly 180 degrees FOV. Such a wide FOV made it possible to see the partners' embodiments in relation to the objects much better than would have been possible on a desktop system. According to our observations, the embodiments were mainly used to inform the partners about (1) one's position, (2) direction of movements and gaze direction - but also (3) to be used as a reference object.

The first point to make here is that for the IPT situation, the user does not see their own virtual representation in the virtual model, only a virtual hand that changed shape to indicate grasping and locomotion. This virtual hand had to be larger than the participant's own hand because otherwise their real hand would obscure it. The fact that both users could see the embodiment of the other, should help in problem solving because they could refer to it, it indicates the presence of the other user and her or his location in the application and also shows some of their body actions. During the first introductory minutes the partners commented on what the other looked like, such as a "Lego-man", a person with "Elvis hair", or as one asked "Are you really as tall as this robot? So am I so tall too?" After this brief period they did not pay attention to the appearance of the embodiment.

We observed that people made often references to a body part of the partner's embodiment. For the Rubik's cube task, when people wanted to make sure that their partner understood what object they meant, they would often describe the object by its position as being close to a part of the embodiments such as legs, arms, head, or shoulders. People also made reference to the size and position of their partner's embodiment. These were used to indicate where to build the Rubik's cube but also to refer to a special smaller cube. And, even though the cubes had different colors on their sides it was easier to refer to one as "that one by the side of your leg" than to indicate the colors which would not necessarily have been the same as the partner's (because of the different view).

The following example shows a pair that had begun to build the Rubik's cube very low towards the ground, so that it was difficult to see what colors the bottom-sides of the cubes had. In the Gothenburg VR-Cube, tracking was not very accurate close to the floor so this made it even more difficult for this participant to look underneath low objects. In the following situation they decided to move the objects and build up at the chest level.

G has some difficulties with placing objects close to the ground.
G: " Ok L, for technical reasons we have to start the building at chest level."
L: "Ok..."
They start to build and L offers to fill in missing pieces underneath if G wants him to. L then places a block at approximately chest level and says:
L: "I'll put things where you can reach them then..."

Example 10. Using embodiment as reference.

The subjects often referred to the other’s avatar when they placed objects. In this particular example we can also see that it had a positive influence on the collaboration since they could create a situation where they could both participate in the object manipulation. This is clearly shown when the person from London is saying: "I'll put things where you can reach them then." This statement shows a level of knowledge of body-relative distance and scale that would be very difficult to achieve in a desktop situation. The statement also suggests that subjects believe that the avatar properly represents the potential actions of their partner. Finally, it is further evidence of how the participants strive to enhance their opportunities for working together in the immersive system and adapt their behavior to the constraints given by the technology.

Even though there was a slight confusion concerning about how to interpret left and right, the example below shows how participants related the position of the object they where referring to to their avatar. Again, this strategy was often used to create a mutual understanding of what object they where talking about.

> *L and G are facing each other, and thus have right and left on different sides, therefore the misunderstanding.*
> L: "Now we have blue on your left side. Blue on your left side."
> G: "You mean right...oh...yeah, towards..."
> G: "Your left?
> L: "The left side!"
> G: "Ok, ok, sorry."

Example 11. Referring to the partner’s specific side.

Our observations in the IPT systems also confirmed that “the embodiment should broadly reflect the appearance of physical body” as in the Hindmarsh study (1998: 218). By using human-like embodiments, it is easier to make reference to body position and scale.

9. Discussion

The use of immersive technologies, with larger FOV and tracked avatars, seems to have a positive impact on the collaboration in networked virtual environments. As we have seen, the major hindrances identified by the Hindmarsh study for object-focused interaction have little impact on IPT CVEs. By describing collaboration in IPT CVEs, we have identified a number of ways in which technology supports a “problem-solving situation” and successful social interaction - as well as some problematical ones.

9.1. Reflections on the Suggestions from the Hindmarsh Study

Although the four key limitations do not affect the collaboration to the same degree in the IPT type systems as in desktop systems – they still exist. The limited horizontal FOV and problems relating the slow or clumsy actions are mainly overcome in the IPTs. However, the lack of information regarding others’ actions is still an issue – the partners are more informed, but the information could be improved. The limitation relating to the lack of parallelism of others’ actions also needs to be considered for IPTs. Based on the application used in their studies, the Hindmarsh study proposed two different areas to improve: making more explicit representations of others’ action, and implementing mechanism for coordinated navigation.

The first suggestion is intended to avoid fragmented interaction for networked desktop systems. For the tasks presented in our study, a pointing function as used in the “Furniture world” in the Hindmarsh study would not necessarily have been useful. In the IPT situation, the ability to make natural gestures alongside verbal instructions combined with the wide FOV means that the meaning of indicative gestures is less missed. Furthermore, in our study, the distance over which indicative gestures were made was not great, although over greater distances inaccuracies due to tracking errors, avatar misalignment, and depth perception might be more significant.

There are actions and activities in networked CVEs that are not transmitted because of system limitations. The CVEs considered here could not convey facial expression, eye gaze, or much of the body posture. The most notable example of this was the subjects' repeated use of their non-tracked hand to make gestures, with the result that the gesture was not conveyed to the partner. Video avatars are a promising technology that aims to include detailed representations of participants into the world (Gross, Würmlin, Naef, Lamboray, Spagno, Kunz, Koller-Meier, Svoboda, Gool, Lang, Strehlke, Moere & Staadt 2003). However, to our knowledge there is currently no implementation that allows seamless interaction between participants and between participants and virtual objects.

The second suggestion in the Hindmarsh study was coordinating navigation in relation to the other's actions. In neither of the tasks presented in this paper did the subjects have problems to navigate. Indeed, in the puzzle task, they almost did not need to navigate. This is partly due to the closed nature of the task, but also because the subjects could move within the tracking area and did not need to re-orient themselves by means of an interaction metaphor to see nearby objects.

9.2. Advantages of IPT Technologies

The main suggestion that the Hindmarsh study proposed to tackle the problems of fragmented interaction was to increase the FOV by using desktop methods to visualize larger horizontal FOV, or using "radical solutions to these problems [which] might involve the use of immersive technologies such as head-mounted displays, wide-screen projection interfaces and CAVEs"

(1998: 215). In this paper we have confirmed their main hypothesis: interaction does appear more seamless in the networked IPT systems, though there are other classes of hindrances to be overcome.

The IPT situation seems to better support collaborative interaction since it allows parallel actions. The subjects were able to handle the lower level interaction tasks intuitively through the devices, and this allowed them to concentrate on collaboratively solving the problems with the partner in the application. Moreover, the subjects frequently manipulated the environment whilst speaking to each other. For the desktop CVE systems, social interaction was needed to compensate for interaction via the interface. In the immersive CVEs, the two were much more interwoven. Here we have to mention how people used small social phrases or took glances at each other (minor interactions due the intuitive technology) to maintain the flow of collaboration. The responses to these small social or technical activities the partners were often none or negligible. However, people used these frequently, especial for the open-ended task in the Modeling World application. These activities seem to have a positive influence on collaboration and they deserve further study. The role of social interaction in supporting collaboration has also been documented elsewhere (Spante, Heldal, Steed, Axelsson & Schroeder 2003).

As shown in the previous section, immersive IPT type CVEs diminish certain problems found in desktop CVEs. At the same time, however, other problems become more visible. For example, although tracking information was available, participants would often forget that it was not complete in the sense that it did not record their full movements, but only the movements on the wand. Here we can mention that there were cases when a person reminds their partner to point with their “right” (tracked) hand, even tough this person could not see that the partner was pointing. The person could guess the non-tracked hand movements of the partner after hearing her or him explaining “this one here” or “ that one there” without seeing that she or he pointed somewhere. Consequently, many of them concluded that the partner was pointing with the non-tracked hand.

Occasionally we observed situations were verbal interaction was paramount, notably when the subjects had to interact with their local environment. Thus during times, for example, when batteries had to be replaced, or when they were interrupted by the experimenter, the collaborators were temporarily detached from the virtual environment, even though their avatar remained. This led to a number of minor confusions where they had to verbally explain to their partner what was happening at their end.

Conclusions

Our main conclusion is that for these applications and with CVE IPT system, we see very few disturbances of the nature reported by the Hindmarsh study (1998). The IPT systems seem to support object-focused interaction in a very efficient manner. We suggest that this improvement is due to two main factors: the larger FOV and the active embodiments due to the tracking system. We have presented several examples of task situations that would not be possible in desktop systems.

Overall, for IPT CVEs we have seen that for object-focused interaction, collaborative tasks proceed quite efficiently without many disturbances due to misunderstandings of reference or action. Thus, the verbal communication was much more in parallel with gestural or physical interaction rather than having to compensate for misunderstandings. While for the desktop CVE study, speech was used *instead* of the gestural interaction to describe activities, for the immersive environments it supports continuing activity.

On the whole, the communication oriented to the collaborative process for this study was not to any large extent fragmented by technological disturbances. Accordingly, the main limitations influencing interactions have a different character for the immersive CVEs. While the obstacles regarding limited FOV and slow movements are diminished, the collaboration is supported by the possibility of shifting the focus towards the partner and supporting parallel actions.

We have also shown that some minor characteristics of the technology could matter for effective interaction. For example the differences in the usability of the devices should be more highly prioritized. For these applications, the use of a more complex IPT system with an additional wall in Gothenburg did not lead to a considerably better performance for Gothenburg participants, nor did it seem to make a difference to the participants' ability to collaborate. In fact the only technology difference that could be noticed was the simpler wand device in London, which turned out to be more easily usable on the whole. This points to the need for of further studies on the impact of the complexity of the technology on collaboration.

Another future direction that should be more closely examined is the role of the social interaction in supporting effective collaboration. It would be useful to know how the requirements for social interaction could be enhanced by the underlying design of the environment and interaction techniques. This does not mean that the next improvements in CVEs should not involve technical changes, but these improvements should result from direct requirements of the users and the application context. This need is even more urgent in the case of CVEs, where a major part of the interaction is social, and the focus needs to be more on supporting collaboration and not on supporting the *individual* users' interactions with the environment.

References

Bente, G. & Krämer, N. C. (2002). Virtual gestures: Analyzing social presence effects of computer-mediated and computer-generated nonverbal behaviour. *Fifth Annual International Workshop PRESENCE 2002*, Porto, 233-44.

Bowers, J., O'Brien, J. & Pycock, J. (1996a). Practically accomplishing immersion: Cooperation in and for virtual environments. *Computer Supported Cooperative Work (CSCW'96)*, 380-89.

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

Bowman, D., Gabbard, J. & Hix, D. (2002). A survey of usability evaluation in virtual environments: Classification and comparison of methods. *Presence: Teleoperators and Virtual Environments*, 11(4), 404-24.

Bowman, D. A. (1999) Interaction techniques for common tasks in immersive virtual environments. Design, evaluation, and application. (PhD Thesis, Georgia Institute of Technology, 1999).

Bystrom, K. E. & Barfield, W. (1999). Collaborative task performance for learning using a virtual environment. *Presence: Teleoperators and Virtual Environments*, 8(4), 435-48.

Cruz-Neira, C., Sandin, D. & DeFanti, T. (1993). Surround-screen projection-based virtual reality: The design and implementation of the CAVE. *Proceedings of SIGGRAPH 93*, New York, 135-42.

Dix, A., Finlay, J., Abowd, G. & Beale, R. (1998). *Human-computer interaction.* (Second ed.), Glasgow: Prentice Hall.

Draper, J. V., Kaber, D. B. & Usher, J. M. (1998). Telepresence. *Human Factors*, 40(3), 354-75.

Frecon, E., Smith, G., Steed, A., Stenius, M. & Stahl, O. (2001). An overview of the coven platform. *Presence: Teleoperators and Virtual Environments*, 10(1), 109-27.

Gross, M., Würmlin, S., Naef, M., Lamboray, E., Spagno, C., Kunz, A., Koller-Meier, E., Svoboda, T., Gool, L. V., Lang, S., Strehlke, K., Moere, A. V. & Staadt, O. (2003). Blue-c: A spatially immersive display and 3D video portal for telepresence. *ACM Transactions on Graphics (Siggraph'03)*, 22(3), 819-27.

Heldal, I. (2003). The treatment of collaboration in the usability evaluation models for collaborative virtual environments. *The 10th International Conference on Human Computer Interaction (HCII2003)*, Crete, Greece, 914-18.

Heldal, I. & Schroeder, R. (2002). Performance and collaboration in virtual environments for visualizing large complex models: Comparing immersive and desktop systems. *The 8th International Conference on Virtual Systems and Multimedia (VSMM2002)*, Gyeongju, Korea, 208-20.

Hindmarsh, J., Fraser, M., Heath, C. & Benford, S. (2000). Virtually missing the point: Configuring cves for object-focused interaction. In Churchill, E. F., Snowdon, D. N. et al. (ed.) *Collaborative virtual environments.* London: Springer, 115-42.

Hindmarsh, J., Fraser, M., Heath, C. & Benford, S. J. (1998). Fragmented interaction: Establishing mutual orientation in virtual environments. *Computer Supported Collaborative Work (CSCW'98)*, 217-26.

Leigh, J., Johnson, A., Brown, M., Sandin, D. & DeFanti, T. (1999). Visualization in teleimmersive environments. *IEEE Computer,* December, 66-73.

Sallnäs, E. L. (2002). Collaboration in multi-modal virtual worlds: Comparing touch, text and voice and video. In Schroeder, R. (ed.) *The social life of avatars presence and interaction in shared virtual environments.* London: Springer, 172-87.

Schroeder, R. (2002). Social interaction in virtual environments: Key issues, common themes, and a framework for research. In Schroeder, R. (ed.) *The social life of avatars presence and interaction in shared virtual environments.* London: Springer, 1-19.

Schroeder, R., Steed, A., Axelsson, A. S., Heldal, I., Abelin, A., Wideström, J., Nilsson, A. & Slater, M. (2001). Collaborating in networked immersive spaces: As good as being there together? *Computers & Graphics,* 25(5), 781-88.

Slater, M., Linakis, V., Usoh, M. & Kooper, R. (1996). Immersion, presence, and performance in virtual environments: An experiment with tri-dimensional chess. *ACM Virtual Reality Software and Technology (VRST)*, 163-72.

Spante, M., Heldal, I., Steed, A., Axelsson, A. & Schroeder, R. (2003). Strangers and friends in networked immersive environments: Virtual spaces for future living. *Proceeding of Home Oriented Informatics and Telematics (HOIT)*, Irvine.

Stanney, K. M., (Eds.). (2002). *Handbook of virtual environments. Design, implementation, and applications*. London: Lawrence Erlbaum Associates.

Steed, A., Mortensen, J. & Frecon, E. (2001). Spelunking: Experiences using the dive system on cave-like platforms. In Deisinger, J., Bullinger, H. (ed.) *Immersive projection technologies and virtual environments,*. Wien: Springer-Verlag, 153-64.

Steed, A., Slater, M., Sadagic, A., Tromp, J. & Bullock, A. (1999). Leadership and collaboration in virtual environments. *IEEE Virtual Reality*, Houston, Texas, 112-15.

Steed, A., Spante, M., Heldal, I., Axelsson, A.-S. & Schroeder, R. (2003). Strangers and friends in caves: An exploratory study of collaboration in networked IPT systems for extended periods of time. *ACM SIGGRAPH 2003 Symposium on Interactive 3D Graphics*, Monterey, California, 51-54.

Tromp, J. (2001) Systematic usability design and evaluation for collaborative virtual environments. (PhD Thesis, University of Nottingham, 2001). Nottingham.

Tromp, J., Steed, A. & Wilson, J. (2003). Systematic usability evaluation and design issues for collaborative virtual environments. *Presence: Teleoperators and Virtual Environments,* 10(3), 241-67.

West, A. & Hubbold, R. (2001). System challenges for collaborative virtual environments. In Churchill, E. F., Snowdon, D. N. et al. (ed.) *Collaborative virtual environments digital places and spaces for interaction.* London: Springer-Verlag, 44-55.

Youngblut, C. (1998) Educational uses of virtual reality technology, Institute for Defense Analyses 1801 N. Beauregard St. Alexandria.