Importance of communication influences on a highly collaborative task

Abstract

Social interaction is a key driving force behind any team activity. Real time closely coupled interaction where we quickly see the effect of our actions on others is an important feature of social interaction. For example, smiling and returning a smile, shaking hands and passing a gift or business card are some of the most important interactions in the real world. Collaborative Virtual Environments offers the potential for social interaction between geographically distributed groups. Closely coupled interaction is, however, very difficult in present CVE systems. This is because this medium lags behind real world communication in terms of representation, consistency and responsiveness.

This paper takes a psychological perspective describing how the primary forms of human communication in the real world map to those in the virtual. We discuss how each form of communication relates to the feeling of co-presence giving real examples of behaviour observed an example application. We present detailed results from user evaluation focusing on the perceived importance of these influences on collaboration.

Keywords: social interaction, CVE, collaboration, communication psychology

1 Introduction

The most successful developer of large-scale technologies were those who did not just design devices, but also designed societies into which their devices would fit (Hughes, 1983). We need to create systems that provide us with the same abilities of communication and interaction as in the real world. In this context the terms presence and co-presence are widely used within the virtual reality (VR) community (M.J.Schuemie, 2001). Schuemie concludes that little is known about what interaction has to do with presence. It may be argued that even less is known about effective interaction on common objects as a focus of interest (Greenhalgh, Bullock, Frécon, Lloyd, & Steed, 2001) and co-presence.

Communication is a dynamic process that has been under investigation for many years by both psychologists and sociologists. It can be categorised into four basic forms: verbal and nonverbal communication and the role of objects and the environment in communication (Burgoon, Hunsaker, & Dawson, 1994; Knapp, 1978). Conversational analysis (CA) has been used to study the interaction through communication for over thirty years (Hutchby, 2001). An understanding of the nature of interaction in the real world can help us reason about co-presence and may lead to defining its requirements.

1.1 Psychology of interaction and communication

An understanding of the nature of interaction in the real world can help us reason about co-presence and may lead to defining its requirements. We can distinguish interaction into different categories like verbal and nonverbal communication and the role of objects and the environment in communication.

Verbal communication includes mainly speech and sound, and our body language reckons as non-verbal communication. Communication through/with objects and the environment are also non-verbal, but not person related and we have found them important to co-presence.

1.1.1 Verbal communication

Techniques for analysing human-human interaction may fruitfully be applied to human-machine-human interaction. Conversation analysis (CA) is the study of talk-in-interaction (Hutchby & Wooffitt, 1998). This is the systematic analysis of the kinds of talk produced in everyday, naturally occurring situations of social interaction. Normally a conversation is taped and later on analysed. However, the majority of CA work has focused on telephone conversation, but this method can also be adapted to CA in virtual reality. For example, interaction in virtual environments including audio is similar to a telephone conversation.

The gazebo showed that speech between participants is vital. The lack of cues (gestures, touch, etc.) affords more verbal communication to coordinate and fulfil a task. An example of a simple dialog is questioning: How would we do it without audio using only gestures?

Bob: Give me the plank, please. I like to fix it on the beam.
Jeff: Which plank, the green or yellow one?
Bob: Just give me one!
Jeff: Do you want that I hold it for you when you are going to fit it?
Bob: Yes, please!

Without speech this dialog would take a while and it would be probably even more effective for Bob to get the plank by himself. The advantage of VR is that it can provide us with more cues which again help us to understand and to interact (among other things it increases the feeling of co-presence). Nonetheless, verbal communication plays a vital role in close coupled collaboration and CA can help us better to understand in which way we need to improve our applications. Conversation analysis on talk-in-

interaction is helping us to understand how we change the subject of a conversation, called turn taking. The analysis for turn taking conversations is focused on "What and how do we change the focus of a conversation". The point is to show that mutual understanding or inter subjectivity are publicly ratifiable accomplishments, which are observable in the data of talk-in-interaction. We can observe that interactions are patterned. For example, people have developed systematic, recursive ways of beginning (and ending) telephone conversations that may exhibit differences from their ways of beginning conversations in face-to-face circumstances.

1.1.2 Nonverbal communication

Non-verbal communication can play a supportive or even predominant role. In interaction a major part takes place via non-verbal communication. These cues help us in face-to-face interactions. Gestures, bodily orientation, eye gaze and so on are a vital part of our everyday chat with others. An array of circumstances for interaction, which can be differentiated on the basis of their degree of "cuelessness", can be found. For example, in face-to-face interaction we have the fullest range of cues, while on the telephone we have the least range (only the pitch of the voice). CVEs fall somewhere in between face-to-face and telephone interactions (e.g. in terms of cuelessness). A desktop application with no gesture support is less informative than one supporting gestures. Immersive displays (HMD, CAVE, etc.) do not rely on predefined gestures, by using their motion tracking system the degree of cuelessness is decreasing. The data from hand and head tracker are matched to the embodiment, which allows other participants to see where you look and with which object you might interact. The hand tracking also allows you some simple gestures (pointing, waving, etc.), by using a tracker for each hand the variety and flexibility is increasing. The gazebo example could show that even simple gestures are helpful. Things like directing someone to a specific tool or place are easier with the ability to point the direction. Our VR system includes only a hand and head tracker and further research is needed to investigate the influence of other cues like facial expressions or a two-hand tracker.

Hutchby argues that the concept of cuelessness is ultimately misleading because there is no evidence that telephone conversation (interaction) is less effective than face-to-face conversations and it has less to do with the coordination of turn-taking and more to do with "psychological distance" (Hutchby, 2001). We have seen that these cues are important for close coupled interaction. Just by telling a person what to do and how is not very efficient e.g. for passing objects, demonstrate or advise people. Cues like showing a direction are necessary for effective interaction. However, Hutchby also suggest that we should focus more on communicative affordances of the technology than cuelessness and psychological distance. Both may be considered worthy of investigation. Technology may be improved knowing the affordances conditional on cues and distance.

1.1.3 Role of objects in communication

One view to look at presence is the ecological view. The basic approaches for this theory is: The environment offers situated affordances, perception-action coupling and tools become "ready-to-hand". The concept of affordances is associated with the work of Gibson in the psychology of perception (Gibson, 1979). For Gibson, humans along with animals (insects, fishes, birds, etc.) orient to objects in their world (rocks, trees, rivers, etc.) in terms of what he called their affordances: the possibilities that they offer for action.

Objects in communication can be person and non-person related. Person related objects is everything subjective the way we look, e.g. clothing and cosmetic can tell other people what we want and where we want to go. For example, when we see a woman dressed in a nice dress with makeup on, and it is in the late evening, we assume that she is going out, but if we see her at the same time in a dirty work suit we assume that she is still working. When we share an object (sequential or concurrent) or hand it over then we interact with this object in a non-person related way. The way in which we do it shows other what we want with it.

In our example applications, objects are the main focus for interaction and we could observe the e.g. the concurrent sharing of an object is difficult. It is a technological and communicational problem. Users are tempted to carry the object unsynchronised, which can confuse the participants. They lose the sense for direction and the use of verbal communication is increasing.

1.1.4 Role of the environment in communication

The success of collaboration depends not only on how we communicate or interact (including objects) with each other. The environment and our perception of it also play a vital rule. Knapp defined different perception of our surrounding (Knapp, 1978):

- *Formality*: our reaction to the surrounding environment
- *Warmth*: nice colours, panelling, carpeting and furniture help us to relax and to feel comfortable
- *Privacy*: an enclosed environment can give us the feeling of privacy and therefore we are more likely to interact closer and personal
- *Familiarity*: in a new environment we normally look for things that are familiar or speak with others about it
- *Constrain*: our reaction an environment is also depend of whether we can leave it or not
- *Distance*: our response to a given environment can also be influenced by the distance we have to other people

Similar rules are used in architecture for the design of buildings, homes or places. They all have to fit in the environment and they need to be suitable for the purpose. It is also required that people working/using them need to feel comfortable with them. Of course, we should follow the same rules by designing the virtual environment for an interactive task. The gazebo example used a deposit for the material and tools. We used textures for wood or lawn. This increased the feeling of presence and made work easier. For example, with a single colour for the lawn we had problems to recognise objects or to pick them up (the contours of the object became blurred). A real looking grass texture solved this problem. Similar things need to be done for the whole environment if needed.

1.2 Communication in CVEs

We have talked so far over things we like to do and how we psychologically react, but there are many problems that need to be looked at before we can interact in a VE like in our own world. VE allow us to overcome problems of remoteness and bring us together. Systems able to do that are called collaborative virtual environments (CVE), but we are still far away from recreating the real world. To some extent there are some forms of social communication and interaction in collaborative virtual environments. These are partly capable of supporting real world engineering kind of tasks. At the moment you can get together, chat with people or perhaps go in some game and have little of adventure, but there is nothing that supports close coupled collaboration that companies could use. Research in this field is trying to develop a working system, but there are still certain problems which need to be solved like network latencies and communication problems. Many systems are at the moment just like single user systems rather than multi-user interaction applications. They are connected in a multi-user environment but the interaction between them is very limited and non-verbal language like gestures are sometimes implemented (predefined) but seldom used by the users.

At the moment, we are capable of creating an immersive VE including sound for communication. Additionally, we can choose a human like embodiment and create environments needed for the interactive task (physical behaviour is/can be implemented). To reproduce gestures we can use motion tracking which could be implemented for the whole body, but these systems do not work in real time, which is an essential condition for social interaction in VR.

Earlier work on human interaction looked at the way in which people use their viewpoints and react to gestures (J. Hindmarsh, Fraser, Heath, Benford, & Greenhalgh, 2000). Hindmarsh et al. observed that a desktop user, when directed to an object by gesture and verbal comment, tend to visually locate the user and then follow his gesture to locate the object. This can lead to confusion when the directing user is changing his/her position or the gesture. However, this work did not focus on close interaction or immersion. Immersive displays place a user in a spatial social context allowing natural first person observations of remote users interacting with objects. This improves the work within such an environment and when connected with other non-immersed users it can be observed that the immersed user adopts a leadership role (Slater, Sadagic, Usoh, & Schroeder, 1998; Steed, Slater, Sadagic, Tromp, & Bullock, 1999).

We will now introduce a CVE application that will be used as a basis for discussion through the document, the Virtual Gazebo (Roberts, Wolff, & Otto, 2003). This has been purposely derived to examine a set of distinct forms of interaction within one structured task. Tools and materials have different physical properties (weight, gravity, etc.) and only by using them in the right order it is possible to fix beams to each other (using screws and joints).

2 Experimentation

We have designed the structured task of building a gazebo in order to examine distinct scenarios of sharing the manipulation of an object. To complete the Gazebo, tools and materials must be shared in various scenarios of shared object manipulation, distinct in the method of sharing attributes. The two main scenarios were moving a beam to a distinct position and to fix the beam vertically. A detailed description of the gazebo application is included in our earlier work (Roberts et al., 2003). Verbal and non-verbal communication are required to archive the task and to agree on the work-sharing. All these forms of social communication should create the feeling of presence for the participants in this environment and enable co-working with others. However, the user behaviour and manipulation of objects may be affected by the different display devices provided.

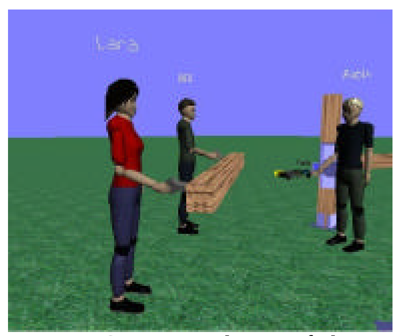

Figure 1. Concurrent sharing of object through the same attribute.

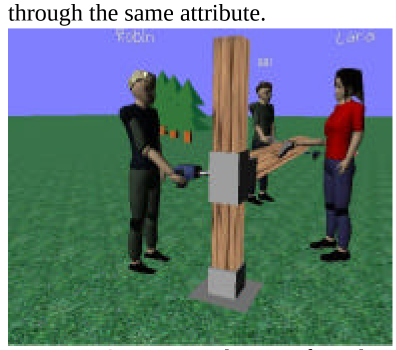

Figure 2. Concurrent sharing of an object through distinct attributes.

The first scenario of moving a beam, which is too heavy for a single user to lift alone due to simulated gravity, need's people to collaborate closely to synchronise their action. They have to agree on the objective and contribute in the same way to the task (Figure 1). After successfully dragging the beam to the desired position, the users have to decide in the second scenario who is holding the beam in place to prevent it falling while the other is fixing it (Figure 2).

The users must synchronise their activity using any appropriate selection of forms of communication in order to move the beam to the desired position and to fix it. It is up to the users to decide how they approach the task. They can use social communication and talk about how they proceed. In addition non-verbal communication can be used, such as gestures, to point where to go or which beam to take. When a user picks up one end of the beam with the carry tool, this end will be surrounded by a coloured aura, indicating to everybody that the user is now ready to drag the beam. The same happens when the second user picks up the other end. This is helpful for the synchronisation of the two users' actions. When the beam is fixed, the aura around the beam, joiner and screw disappears.

3 User Evaluation

The gazebo application was put to two trials and users were asked to complete a comprehensive evaluation. Some of the results concerning contribution and collaboration were used in a previous paper (Roberts et al., 2003), while this article concentrates on the communication (Figure 3) during the building scenario (Figure 4). Over fifty undergrad students were split into teams of three for each test. Within every task, each user interacted through a distinct display device. Users' perceptions of the test were ascertained through a questionnaire. At the time of writing a second pilot trial tried to resolve some of the question rose by the first evaluation and used ten students, which filled out the some questionnaire in the same way as for the first trial.

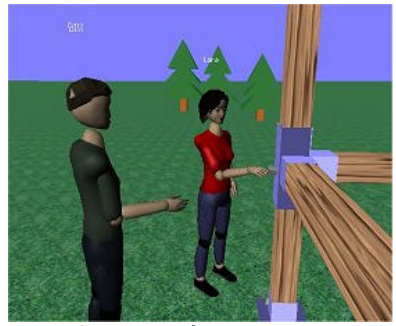

Figure 3. team work instruction

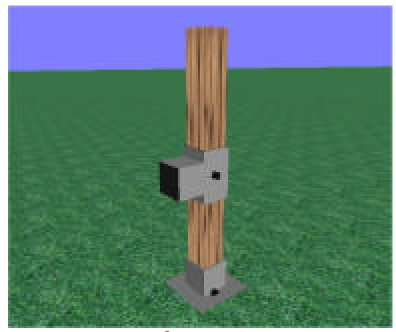

Figure 4. a simple structure

3.1 Display configurations

The tests involved four different display configurations, each different in its ability to support interaction with the other two participants as shown in Table 1 and Table 2.

For the IPT_T1 and DT1_T1 user in the first trial, verbal and non-verbal communication were only supported through the use of technology. This was in contrast to the DT1_T1 and DT2_T1 where the user shared the same physical space although each of them viewed the virtual environment through their own display from the perspective of their avatar. They could therefore communicate verbally without the restrictions of technology, but were facing in opposing directions and did not communicate non-verbally except through the interface. The second trial supported audio only through the use of technology but using headsets. This was in contrast to the first trial were the technology was not directly visible for the user.

Table 1 Display configurations of 1st trial

Name	Display Type	Audio	Input	Embodiment
IPT_T1	Walk-in	Speaker & Microphone	Tracked Wand	Motion tracking
DT1_T1	Desktop	Speaker & table microphone	Keyboard & Mouse	Low realism
DT2_T1	Desktop	No*	Keyboard & Mouse	Medium realism

* within calling distance to DT1 (ca. 2m) and able to use text chat

Table 2 Display configurations of 2nd trial

Name	Display Type	Audio	Input	Embodiment
IPT_T2	Walk-in	Headset	Tracked Wand	Motion tracking
DT1_T2	Desktop	Headset	Keyboard & Mouse	High realism, static
DT2_T2	Desktop	Headset	Keyboard & Mouse	High realism, static

3.2 Conditions: Display combinations

The different combination of the display configurations created altered test conditions (Table 3) which have been borne in mind while the participant filled out his/her questionnaire. This means that the questioned user had to give quotes for the different users from their own perspective.

Table 3 Test conditions

Condition	Questioned user	User 2	User 3
C1	IPT	DT1	DT2
C2	DT1	IPT	DT2
C3	DT2	IPT	DT1

3.3 Questionnaire

Over 20 related questions concerning the perception of collaboration both generally and for each specific task were asked. The questions were based on that of Usoh and colleagues (Usoh, Catena, Arman, & Slater, 2000) and sets of related questions attempted to reduce error from misinterpretation. Errors arising from a user's misinterpretation of a question were reduced by asking sets of related questions. The answer could be given on a scale of 1-7 where 1 represented agreement to a very small extent and 7 to a very large extent. The questionnaire included questions concerning how subjects interacted with the object in the different scenarios, but also on how they perceived the interaction with the remote users. Some of the questions were asked twice, separated for the different scenarios. Furthermore, the questionnaire focused on how the display device influenced the users' interaction.

4 Results

This section focuses on the results of the user evaluation. The results come under scrutiny of an analysis of variance (ANOVA), which is a common method to check the significance of results. Questions concerned aspects of verbal and non-verbal contributions to the task. In relation to the Verbal communication, Non-verbal communication, Shared objects and the Environment the participants gave us their opinion on "To what extent did each of the following contribute to task performance?". The evaluation with ANOVA showed that there is a significant difference between the participants using the different display types, table 4.

The results of table 4 show clearly that the participants found verbal communication significantly more important than non-verbal communication (figure 5). In addition, users were found to be ambivalent towards the contribution of shared objects with an average mean value of M=55. However, for all conditions the interviewed user perceived the environment as similarly important (M=64).

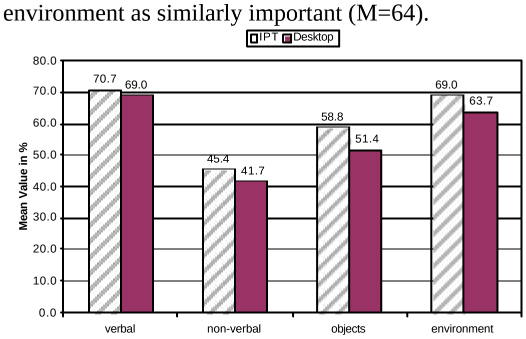

Figure 5. influences first trial

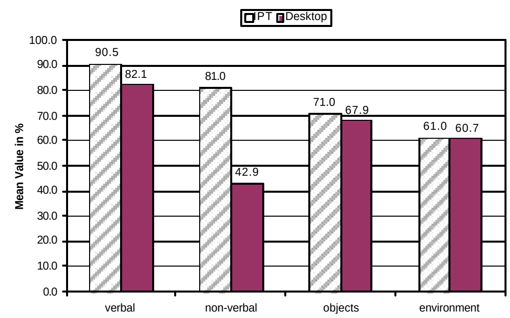

Figure 6. influences second trial

We observed that all participants partly ignored the ability to talk to each other for synchronising their action, even though they knew each other. Although the observers tried to encourage them to use this facility, a lively discussion was seldom observed as one person pointed out: "[There was a] Lack of verbal communication despite equipment provided". These behaviours were observed for the two different methods of communication provided for the participants as on the one hand a technology restricted microphone connection and on the other a non-restricted direct communication. We observed that this

Table 4 ANOVA results for Verbal and non verbal contribution 1st trial

ANOVA results ($\alpha=0.05$)	posthoc test (Tukey) shows differences between:
$F(2,48)= 3.78$, $MS_W=3.10$, $p=0.014$	verbal (M=70.7, SD=28.4) non verbal (M=45.4, SD=30.0)
$F(2,34)= 5.06$, $MS_W=1.96$, $p=0.005$	verbal (M=76.2, SD=22.2) non verbal (M=45.7, SD=17.6) shared obj (M=52.9, SD=25.2)
$F(2,30)= 3.46$, $MS_W=2.05$, $p=0.025$	verbal (M=61.9, SD=20.5) non verbal (M=37.7, SD=9.6)

reluctance to verbal communication is counter productive to the highly collaborative task. For this reason a seconded trial attempted to resolve this question by providing communication technology to all participants.

During the seconded trial it could be clearly observed that the degree of communication between the participants was significant higher than during the first trial (figure 6). The desktop user (DT1_T2 and DT2_T2) found verbal communication was significantly important (M=82.1 SD=7.1) than non-verbal communication (M=42.9 SD=26.1). In contrast the immersed user IPT_T2 found verbal communication very important (M=90.5 SD=8.25) as well as non-verbal communication (M=81 SD=16.5). With more communication the participants also interact more with each other and therefore they found the contribution of shared objects to the task higher (M=67.9 SD=7.1) as during the first trial. However, the environment was found similar important (M=60.7 SD=13.7) for both trials. In case of this second study, the sample set was insufficient to justify an ANOVA. Nevertheless, the observation and the gathered results show a clear trend for higher use of communication when provided with technological support in form of a audio headset.

Desktop users can manipulate distance objects through space without moving their own position whereas the immersed user can only manipulate the object by moving to it and have it within his/her reach. This desktop behaviour is a common feature for CVE's and designed to reduce the movements of the avatar, but also to simplify the object handling. However, as found in previous studies (Jon Hindmarsh, Fraser, Heath, & Benford, 2001) this makes it more difficult for other users to see the relationship between the acting user and their object of interest. The same conclusion can be drawn for such a task as the virtual gazebo. People became confused about who was doing what. It could be observed that participants were surprised when another desktop user interacted with an object of their own interest. They started to ask questions such as "Are you taking the metal joiner? I thought I should take it".

5 Conclusion

Combing IPT and CVE technology provides, for the first time, the possibility of supporting the full range of social human communication between geographically remote people. One would have expected verbal communication between remote users to become more natural when the technology is transparent, that is when the microphone and speakers are hidden. However, we observed a significant increase in verbal communication when the user is constantly aware of a familiar communication device, that is, a headset with microphone and earphones. When this was done, the team worked together more successfully and each participant made greater use of the remaining communication influences. For example the perceived importance of shared objects by experienced users increased by 20% when using a headset. Verbal communication was perceived to be of the greatest importance. Little difference was perceived in the importance of the other influences. The current state of technology is still some way from providing natural social human communication between remote participants. Improvements must be made in interface, representation and underlying communication. We should not, however, address this in a adhoc manner. Understanding real world social interaction and communication is key to emulating it. The classic taxonomy adopted in this paper is well accepted for co-located groups and we propose that it is useful for reasoning about the requirements and effectiveness of CVE technology. This paper has demonstrated such reasoning.

5.1 Future Work

We are now continuing this work on a number of fronts. We want to extend the seconded trial to a number which allows us to use statistical methods on the results. Furthermore we are about to repeat the test with multiple immersive displays including tracking system to increase the concentration of the participants the task and not the interface. We are also about to implement the gazebo application to the alternative platform CAVERNsoft, which allows us to introduce different message handling as well as new communication methods. Furthermore, we are devising a system to recognise, in real-time, a set of user behaviours typically exhibited during collaborative work. Once classified, only the "name" of the behaviour will be sent over the network and the communication requirements between distributed locations should be dramatically simplified.

6 References

Burgoon, M., Hunsaker, F. G., & Dawson, E. J. (1994). *Human Communication* (3rd ed.). London: SAGE Publications.

Gibson, J. J. (1979). *The ecological approch to visual perception*. Boston: Houghton Mifflin Co.

Greenhalgh, C. M., Bullock, A., Frécon, E., Lloyd, D., & Steed, A. (2001). Making Networked Virtual Environments Work. *Presence: Teleoperators and Virtual Environments, 10*(2), 142-159.

Hindmarsh, J., Fraser, M., Heath, C., & Benford, S. (2001). Virtually Missing the Point: Configuring CVEs for Object-Focused Interaction. In A. J. Munro (Ed.), *Collaborative Virtual Environments. Digital Places and Spaces for Interaction* (pp. 115-139). London, UK: Springer Verlag.

Hindmarsh, J., Fraser, M., Heath, C., Benford, S., & Greenhalgh, C. (2000). Object-Focused Interaction in Collaborative Virtual Environments. *ACM Transactions on Computer-Human Interaction (ToCHI), 7*(4), 477-509.

Hughes, T. (1983). *Networks of Power: Electrification in Western Society*. Baltimore, MD: Johns Hopkins University Press.

Hutchby, I. (2001). *Conversation and Technology*: Cambridge, UK.

Hutchby, I., & Wooffitt, R. (1998). *Conversation analysis*: Cambridge, UK.

Knapp, M. L. (1978). *Nonverbal Communication in Human Interaction* (2nd ed.): Holt, Rinehart and Winston.

M.J.Schuemie, P. v. d. S., M.Krijn. C.A.P.G.van der Mast. (2001). Research on Presence in VR: a Survey. *Cyberpsychology and Behavior, 4*(2), 183-202.

Roberts, D. J., Wolff, R., & Otto, O. (2003). Constructing a Gazebo:Supporting team work in a tightly coupled, distributed task in virtual reality. *Presence: Teleoperators & Virtual Environments, 12*(6).

Slater, M., Sadagic, A., Usoh, M., & Schroeder, R. (1998, June). *Small Group Behaviour in a Virtual and Real Environment: A Comparative Study.* Paper presented at the BT Workshop on Presence in Shared Virtual Environments.

Steed, A., Slater, M., Sadagic, A., Tromp, J., & Bullock, A. (1999, March). *Leadership and collaboration in virtual environments.* Paper presented at the IEEE Virtual Reality, pp. 112-115.

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