Evaluating Collaboration in Distributed Virtual Environments for a Puzzle-solving Task

Ilona Heldal

Department of Technology and Society
Chalmers University
412 96 Gothenburg
Ilona.Heldal@mot.chalmers.se

Anthony Steed

Department of Computer Science
University College London
London WC1E 6BT
A.Steed@cs.ucl.ac.uk

Ralph Schroeder

Oxford Internet Institute
Oxford University
1 St.Giles
Oxford OX1 3JS
Ralph.Schroeder@oii.ox.ac.uk

Abstract

There is a wide range of different distributed Virtual Environments (VEs) in use and their areas of application continue to grow. However there is still very little research about factors that contribute to effective collaboration and about how people collaborate in distributed VEs. Existing studies have concentrated on the technical aspects of usability, mostly for single user virtual environments. The purpose of the paper is to discuss the benefits and limitations of different methods for evaluating the effectiveness and experience of collaboration in distributed VEs for a puzzle-solving task. One method is quantitative and based on performance measures and responses to Likert-scale type questions. The results obtained with this method can provide valuable insights, but may be limited if the aim is to gauge users' experience of collaboration. A second set of methods is qualitative, using, for example, observations of how people collaborate, and analyzing their responses to open-ended questions. The same phenomena can thus be examined from different angles, and this paper presents some of the benefits and drawbacks of using quantitative versus qualitative methods. The results also shed light on the factors contributing to better collaboration for object-focused tasks in distributed VEs with different combinations of technologies.

1 Introduction

New technologies for collaborating at a distance are becoming ever more widely used. Virtual reality and collaborative or shared virtual environments (CVEs) represent one technology for doing distributed tasks, with three types of technologies emerging as the main options: immersive projection technology (IPT) systems (also known as CAVE-style displays, (Cruz-Neira, Sandin & DeFanti 1993)), head-mounted displays (HMDs), and desktop or PC-based systems. Although virtual environments on desktop systems have been used for distributed collaboration in a number of ways, the use of immersive systems (IPTs and HMDs) to work together at a distance is still rare and confined to a few trials (for overviews of collaborative or shared VEs, see Churchill, Snowdon & Munro 2001; Schroeder 2002; Schroeder & Axelsson 2005).

It would be useful to know which of the three types of systems or combinations of systems, IPT, HMD, or desktop, is most effective for spatial tasks. Spatial tasks present a good 'test case' for evaluating collaboration since interactive graphical displays lend themselves particularly to tasks where objects need to be manipulated together. With spatial tasks, the focus is on jointly handling the objects – unlike, say, tasks in which the nuances of interpersonal expression and communication such as facial expressions or non-verbal communication are most important. It has been shown that pairs of users using networked desktop systems have problems doing spatial tasks since they lack mutual awareness, peripheral vision and the ability to use their bodies as reference points and in relation to manipulating objects (Hindmarsh, Fraser, Heath and Benford 2001). It has also been shown that users can work together more effectively using networked IPTs than desktop systems for an object focused task (Heldal et al. 2005).

Many issues such as presence, performance, intuitiveness, and leadership have been identified as important in virtual environments (Draper, Kaber & Usher 1998; Steed, Slater, Sadagic, Tromp & Bullock 1999). Certain features such as presence, performance, and immersion may be more closely associated with one specific type of VE than with another (Sadowski & Stanney 2002; Slater, Linakis, Usoh & Kooper 1996). For example, desktop systems can be more effective than immersive environments for problem-solving when visualizing large molecules (Heldal & Schroeder 2002) and for educational uses (Youngblut 1998) whereas immersive systems can be more effective for visual modelling (Neubauer & Harris 2003) and abstract information visualization (Bowman & Raja 2004). There are also differences when varying some technical factors of an environment; for example, varying latency (Arthur & Booth 1993; Meehan, Razzaque, Whitton & Brooks 2003) or rendering (Steed, Mortensen & Frecon 2001; Zimmons & Penter 2003) results in different performance and presence measures. For single user environments, Bowman and his colleagues evaluated more general issues in relation to the utility and usability of different systems (Bowman, Gabbard & Hix 2002). A recent study evaluated several interaction styles, comparatively, such as selection and manipulation in IPT and HMD type CVEs, and showed how effectiveness depends on the applied interaction style, and on the used technology used (Steed & Parker 2005).

There are a number of methods that have been developed to evaluate collaboration in the literature about computer-supported collaborative work (CSCW). VEs, however, are different, in so far as collaboration also involves the interplay between the bodies, gestures and the social dynamic *between* two or more individuals *inside* the environment. There are few methods of evaluation of collaboration in CVEs (Tromp, Steed & Wilson 2003; Heldal 2004) The first of these studies (Tromp et al. 2003) examined social interaction: by defining and examining certain types of human interaction activities that have to be supported by the distributed technology, this study pointed towards more usable applications. The second study (Heldal 2004) examined the interconnection between social interaction and interaction via technical interfaces during problem-solving in order to identify activities that better support collaboration. Both studies also point to the importance of integrating qualitative approaches in evaluating complex collaboration in groups. There is once again a growing interest in qualitative evaluations, both for evaluating computer systems and certain of their features generally as well as for evaluating VEs (Slater 2003).

2 Study Design

2.1 Background

A number of papers have presented other, mainly quantitative, results of the trial discussed here. Some of the results about performance and presence were reported in (Wideström et al. 2000; Schroeder et al. 2001; Heldal et al. 2005;). The first of these studies (Wideström et al. 2000) reported unequal collaboration between the pairs using different (immersive and desktop) systems, despite being unaware of what type of system the other participant was using. The second study (Schroeder et al. 2001) noted that performance in a networked immersive (IPT) setting was almost as good as in a real-world setting, especially compared to a setting linking an immersive and a desktop system. The third study (Heldal et al. 2005) added and compared a variety of settings that had not been previously reported (including IPT system linked to an HMD system, and desktop linked to a desktop system) and explored the quantitative results about collaborative experiences and performance. Quantitative results were particularly useful for analyzing different experiences of immersiveness among participants and also for comparing how this experience differed among participants using different systems. The usability of the different settings was measured in terms of time to perform the task (see 3.1.1 below). The previous studies showed that usability depends to a large extend on effective collaboration. However, collaboration, via estimations given for the Likert-scale type questions in the questionnaires did not show great differences between the five different settings (see 3.1.2 below).

Even though, as we shall see, the performance results were different for the different settings, the self-reports about collaboration were similar and could not discriminate between the settings. In other words, directly asking the subjects about their experiences of collaborating, even when the results are statistically significant, does not necessarily provide any understanding of how well or poorly people collaborated or experienced collaboration. This indicates that quantitative methods for studying differences between successful and failing collaboration are not enough.

2.2 Method

The qualitative results presented in this paper highlight particular hindrances and benefits that participants experienced in the different systems. For this, we analyze answers to open-ended questions from the four best and worst pairs in the different settings combined with observations made by the experimenters (and noted down on observation sheets). We also show aggregate qualitative data from all the subjects to shed additional light on the differences in usability between the different settings. Since the focus in the papers is on collaboration and on improving the methods that can be used to design better and more usable systems for collaboration, we relate the previously obtained quantitative results on performance and experiences to factors obtained by the qualitative data to give insights into collaboration.

2.3 Task and Settings

The puzzle involved 8 separate blocks with different colours on different sides. The pairs had to rearrange the blocks so that each side would display a single colour, i.e. 4 squares of the same colour on each of the six sides (see Figure 1). Therefore, the task was similar to, but less complex than, the popular Rubik's cube puzzle that involves 27 blocks with 9 squares on each side. The squares were 30 cm along each edge both for the physical cardboard blocks and their representations in the VEs.

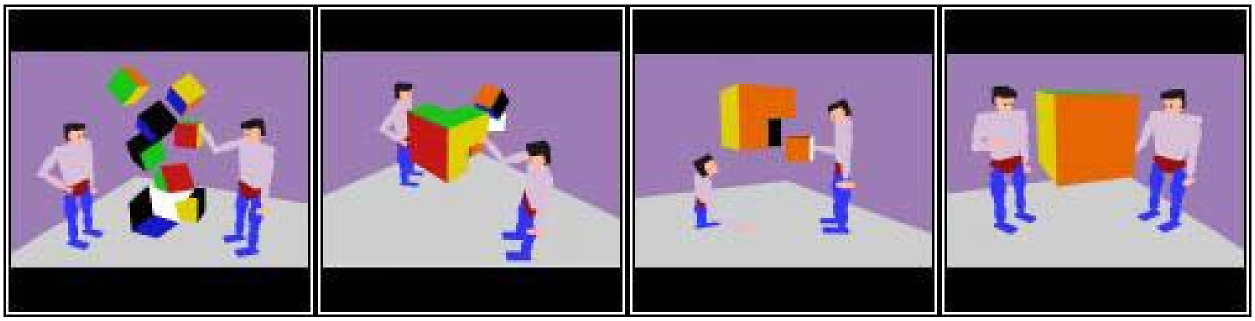

Figure 1: Four stages of solving the Rubik's Puzzle

The technology used in the trials is described in detail in the previous papers mentioned above (Heldal et al. 2005). There were 22 pairs in each setting; in other words, 220 subjects in total (5x2x22) and in this paper we focus mainly on the four best and worst pairs (for each 5x2x4 = 40 subjects). The best pairs were chosen according to their best performance result, i.e. the shortest time for solving the tasks. The four worst were those that failed and those with the worst results for the rest for the FtF and the I – I setting. For the I – HMD, I – D, and D – D settings, we examined randomly chosen couples who failed from the total amount of couples who failed (8 for the I – HMD, 14 for the I – D, and 10 for the D – D setting). Table 1 below presents the different settings and the abbreviations that will be used in this paper.

Table 1: Abbreviations and settings

	The Five Settings
FtF	Face to face collaboration with cardboard blocks
I – I	Gothenburg IPT connected to London IPT
I – HMD	IPT connected to HMD
I – D	IPT connected to Desktop system
D – D	Two connected Desktop systems

3 Results

3.1 Quantitative Evaluation

3.1.1 Performance

We first present the results for performance in the four settings (Figure 2). It is worth observing that the curves for FtF and I—I setting are quite similar, with a few groups completing the task very quickly. This suggests that the learning threshold for the IPT is low, in contrast with I—D and D—D, where no group finishes until 11 minutes have passed.

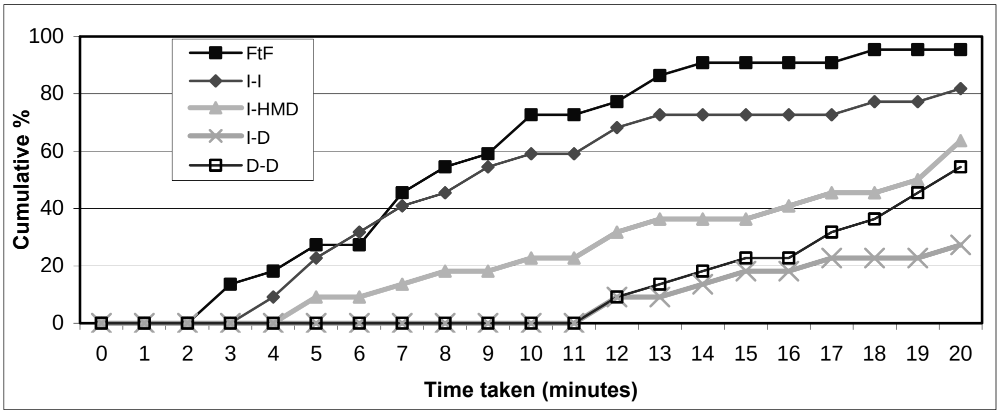

Figure 2: Cumulative percentage of the pairs completing the tasks during the available time.

3.1.2 Collaboration

We asked several questions to determine how individual participants estimated their contribution to the collaboration. The questions were: “To what extent did you have the experience of collaborating with your partner?” and “Think back about some previous time (before today) when you enjoyed collaborating with someone. To what extent did you enjoy collaborating with your partner in today’s task?” The participants marked responses on a scale of 1-5, where 1 was to a very small extent and 5 was to a very large extent, and ANOVA was used to find statistical significance. There was no significant difference ($p > .05$) across the five different conditions for any of the questions. The mean values for *all* three questions and for *all* five settings were in the interval of 3.9±0.4. Therefore, we report below only the estimations for the first question.

The differences between the mean values for the best and the worst pair and all pairs are quite similar. As we will present qualitative reports of these pairs later, we present the results for these pairs in table 2. We can see here some contrasts between the best and worst pairs in terms of how they assessed their collaboration, especially for those who used asymmetrical settings, i.e. I – D and I – HMD.

Table 2: Collaboration reported for the pairs who performed best and worst in relation to all

	Best: Mean / SD	Worst: Mean / SD	Best & Worst: Mean	All: Mean / SD
FtF	3.62 / 1.06	3.37 / 1.30	3.74	3.52 / 0.90
I – I	3.87 / 0.64	3.25 / 1.28	3.56	3.81 / 0.97
I – HMD	4.37 / 0.51	2.75 / 1.16	3.29	3.52 / 1.17
I – D	3.87 / 1.12	2.87 / 0.64	3.33	3.65 / 1.01
D – D	3.12 / 0.64	3.37 / 1.18	3.56	3.72 / 0.90

We also asked questions about their share in contributing to the task with three questions where the answer had to add up to 100% for the two partners. For the symmetrical settings (FtF, I – I, D – D), answers were close to 50% for each partner, whereas for the non-symmetrical settings there was a greater deviation (I – HMD: 50±45 and for I – D: 50±35). Interestingly, on this quantitative measure of contribution, there were no differences in the answers for contributing to the task between the subject population as a whole and the best and worst pairs.

3.2 Qualitative Evaluation

The quantitative data allow us to identify the best and worst four pairs in the first place, but the qualitative data then allow us to examine the key aspects of collaboration that emerge from these two groups. We begin with the responses of the best and worst pairs to an open-ended question about how they experienced collaboration.

3.2.1 Best versus Worst on Collaboration

Many of the pairs who performed best or worst mentioned collaboration and cooperation explicitly by answering to the open ended question regarding their experiences. On some occasions, as when they referred to a ‘cool’ or ‘engaging’ experience, it is not clear whether they mean their joint or their individual experience. For the two non-symmetrical cases (I – HMD, I – D), the collaboration was more extensively commented upon and comments were more focused on communication. The users of desktop systems commented in particular on the use of a microphone. For example, in Table 3, the last person comments that using the ‘phone’ made the experience in the desktop system more ‘lifelike’. Apart from this, the comments about collaboration are overwhelming positive for the four best pairs (Table 3), and we summarize their responses and and those from the four worst pairs (Table 4) below. Table 3 contains the answers to the question “Mention the things about the task that you thought were pleasant.” The comments are translated from Swedish and some of the answers are abbreviated for clarity and brevity’s sake (quotation marks indicate direct quotes, blank spaces indicate no response).

Table 3: Comments about collaboration from the four best pairs

	Person 1	**Person 2**
FtF	To work together.	Collaboration.
	Playful contact between us.	-
	It was cool to work with a stranger.	-
	Collaboration.	-
I – I	“This is the right sort of human condition for enjoyable collaboration.” Friendly partner.	“To help my partner.”
	To collaborate. “Especially considering someone who was abroad.”	-
	Helpful partner.	Collaboration.
	-	“My partner. It was nice to laugh a lot together.”
I – HMD	Pleasant to discuss the environment with another person not at the same physical location.	Collaboration. He had a pleasant voice.
	I enjoyed the experience.	“New experiences are always pleasant. A bit hard to communicate. Probably because of misunderstanding.”
	“Collaborating with another person and cooperating in a strange world.”	Cooperation. To be able to speak and see at the same time that the other is moving accordingly.
	Cool experience.	Communication with another person and see him performing the task.
I – D	Cool to work together with something new.	-
	-	To work together.
	-	To work with someone.
	We did it! It was nice to do it together with someone!	To work together with someone.
D – D	-	Communication.
	The experience.	Collaboration. “Talking into the phone and pointing into the world in a way to select objects that the other person could see.”
	To collaborate.	Cooperating and competing against time.
	“The phone call during the task made a life-like experience. “	To solve a problem with someone else is always doubly pleasurable (even if it can also be a double disappointment).

It can be added that those who performed best were more positive towards collaboration in their answers than people who performed worst. Table 4 contains the views about collaboration from the answers to the questions: “Did your effort differ from that of your partner? If so; please describe how in your own words”, “Mention the things that you think prevented you from accomplishing the task”, and “Mention the things about the task that you thought were unpleasant”.

Table 4: Comments about collaboration from the four worst pairs

	Person 1	Person 2
FtF	Each of us tried to do the tasks for ourselves without actually collaborating.	-
	I did not know my partner so I was reserved from the beginning. Anyway it came out all right.	"Sometimes you think something and the other only interrupts you: because she changed things around."
	"She (and pure luck?) moved the cubes in a way that we put them together in a haphazard way."	-
	-	"I wanted to think aloud and find a solution, but my partner was not interested; she just used a trial-and-error strategy."
I – I	I tried to find a strategy to solve the puzzle but my partner did not.	Communication. Understanding your partner.
	He just did not move away.	Difficulties to see what my partner was doing and saying.
	You couldn't really communicate with the other person, agree on a strategy and collaborate effectively.	-
	I did not like the fact that she paid more attention to the cubes than to the other person.	-
I – HMD	It was good to cooperate with another person. However it was too difficult task for me.	-
	The problems with my partners equipment prevented me to solve the task.	I couldn't see properly. I didn't feel that I had a partner.
	"The communication and time. It would be easier to solve when talking to each other."	Prevented: Seeing the colours and communicating (interruptions). "I participated passively."
	"Maybe I wanted a more active collaborator that could help me solving the problem."	-
I – D	We had bad prerequisites for communication.	Bad verbal communication.
	-	"My partner was more active. I thought that I could speed up later on, but it was too late then."
	Work with someone else.	-
	Maybe I took the cubes and moved them closer to me?	Too "dominant partner".
D – D	Lack of coordination: "body and buddy synchronization".	Hindrances: rules to navigate. the spoken language.
	Lack of interaction between partners. Handling the cubes.	"Getting a good overview".
	"It prevented me that we have focused only on one side."	-
	"I realized at the end that I didn't truly understand what WE had to do together. My partner seemed to be in a good mood."	It was hard to see. Navigation. However, my partner was good at it and we worked well together.

3.2.2 *Hindrances and Benefits: The Experience of Collaboration*

Table 5 provides examples selected from among all 220 subjects where *both* in the pair mentioned problems with collaboration:

Table 5: Responses from pairs where both mention difficulties in collaboration.

	Person 1	Person 2
FtF	"I moved the cubes more. I did not know my partner so I was reserved in the beginning. Anyway it was easy to communicate."	"I often thought aloud. Sometimes I was thinking and was being interrupted by the other person who changed the cubes. But it was pleasant to solve the problem together with somebody else."
I – I	He didn't move further…"We were too considerate and too social with each other in the beginning."	"I usually start with thinking in a rather passive way. Then I speed up. He did it the other way around. It was difficult to see what my partner was doing and saying."
I – HMD	It was a difference in method: "He moved himself around a lot while I turned the cubes more to see."	"We also communicated too little. It took a while to develop a clear strategy."
I – D	"Communication with my partner. I was moving and investigating the cubes more often. Maybe I also took the cubes away from him. In this case I was naughty. I guess."	"He had it easier. I did not know at the beginning if it was a solution or not. My partner dominated a bit."
D – D	"We focused on a separate side each."	"I had to move my half-finished combinations near to his."

The desktop system was generally regarded as more difficult for working together. Still, even pairs using the desktop system mentioned that the task was challenging and the experience therefore exciting. For example: “I felt more real than I had thought. The task was challenging. It wasn’t easy!” But one person who solved the puzzle (together with his partner) on the desktop system in 15 minutes said: “It was not real! I couldn’t touch the cubes, smell them, feel their weight, fabric etc. I was willing to solve the task but not as curious as I would have been if it were real. I lost my patience.” It was not always easy to become used to the desktop system: One person commented, “Given so little time to practice using the control I didn’t really feel I was in charge of the situation.”

To give an indication of the range and frequency of the problems experienced in collaboration, we provide a summary that adds up the number of times different problems were mentioned by all pairs in Table 6. Note that there are more problems mentioned than there were participants since no limit was set on the number of problems that participants could mention and several different items (e.g. glasses, cables, phone line) could be mentioned in one category (ie. physical devices).

Table 6: Summary of usability problems mentioned by all 220 subjects.

Problems	FtF	I – I	I – HMD	I – D	D – D
Software problems: e.g. GUI, slow, not a continuous model	-	8	4 – 5	9 – 13	4
Manipulation, marking, control with steering	-	9	7 – 10	5 – 36	26
Navigation	-	1	1 – 2	1 – 9	7
Orientation	-	1	0 – 1	1 – 1	9
Problems with the sound	-	5	1 – 7	3 – 5	4
Colours	-	6	4 – 14	9 – 0	2
Hard to see all sides	4	2	1 – 5	3 – 7	5
Physical devices, e.g. glasses, cables, phone line, mice, physical surrounding space, objects	7	32	9 – 13	28 – 16	5
Difficult task, idea that it is impossible to solve	13	10	2 – 1	11 – 11	12
Control of the social situation	9	9	4 – 5	5 – 1	19
Cooperation with a stranger, e.g. communicating, misunderstanding	9	1	9 – 4	10 – 6	12
Time pressure	5	0	2 – 1	0 – 2	5
Illness, e.g. nausea, minor headache, eyes	0	7	2 – 6	5 – 1	1
Explicitly wrote: “I don’t have any” problems	19	5	3 – 0	2 – 0	2

In the FtF condition, participants complained about the “impossible logic”, “feeling stupid”, “wearing the microphone” or “being watched”, or they simply did not like “the surrounding physical space”. For the HMD, the data obtained may point to the old and rather heavy HMD that was used. For the settings with less good performance (I – HMD, I – D, and D – D), additional technical hindrances are often mentioned. Manipulating objects, marking them and controlling the devices was found hard for the HMD and the Desktop system. Interestingly, those people who worked the Desktop in the I – D setting found more hindrances to manipulating objects and controlling devices (36 complaints from 11 subjects) than people who worked in D – D setting (26 complaints from 22 subjects) - perhaps the benefits of their immersed partners made non-immersed participants more aware of their limitations. In the asymmetrical settings people commented more on difficulties than in the symmetrical settings, perhaps because lacked knowledge about their partner’s technology. In relation to social interaction, the main problems related to communication, e.g. a person was not sure if he or she saw the same thing as their partner, or misunderstandings in relation to references for the asymmetrical settings. The amount of complaints about social interaction point to the need for further examination of these aspects in relation to effective collaboration, while the positive comments after the trials demonstrate that people like to solve problems and to achieve goals. A large number of participants appreciated in their own words that it was fun to collaborate, and many used words such as “pleasant” to describe “collaboration with a stranger”, both in the networked and in the FtF scenarios.

3.2.3 The Use of Observations made by Outside Observers

There are obvious problems with self-reports about working together. It is difficult to experience the task, solve the problem, and identify problems at the same time. One example is that participants using the HMD frequently bumped into physical obstacles but often they did not report this. Another type of example concerns the fact that participants used the different systems in particular ways without being aware of this. For example, for the I – D settings, when the subjects sitting with the desktop system observed that it was easier to ask their partners about the colors of the hidden side of the cubes rather than to manipulate the cubes themselves in order to see the colors, they did this. Thus a pair would move their avatars in such a way as to face each other and have the model between them, making it easier to collaborate this way (Heldal 2003). Social coordination thus "saved" unnecessary manipulation activities without participants being aware of this and reporting it as an important feature of their collaboration.

Another set of examples of a similar type can be observed for the I – HMD setting. People did not necessarily report differences during their collaboration. One pair for example estimated their effort at 50%-50%. However, the observer noticed: "The subject with the HMD talked much more". In response to an open-ended question the person who wore the HMD commented: "He [the partner in the IPT system] did most of the checking [of the colors on the cubes]. He found the right cubes and I placed them." In other words, dividing the work in this way may not be something that participants are aware of as constituting an uneven collaboration, but they may rather think of this as the use of different tools, and it therefore requires an observer to correlate questionnaire responses and observations. Or again: the HMD person often needed to ask the person in the IPT system to help her or him. For example, one person in the IPT system in the I – HMD setting at first answered that there was "no difference in effort" in collaboration, but noticed later: "It seemed easier for me to grip and move the cubes. Hence, I did this almost exclusively."

In general, the observers noticed if that people talked a lot, they performed better. For example, in the case of one problematic collaboration, the observer noted: "They came close to the solution – then they got worse again. They talked a lot in the beginning." Unlike the participants, the observers were able to gauge that more intensive communication generally supported problem solving. On the other hand, there were instances when the observer could not observe differences in collaboration even though the subjects did. Here is one example from a D – D pair: "It was cool to be able to give orders. I also had better motor skills." Interestingly, this person ranked his contribution as being greater but did not mention any other differences in collaboration.

4 Discussion

Clearly we can obtain different understandings about collaboration from quantitative data, from qualitative responses, and from observations. Quantitative data are perhaps most useful in evaluating those cases when the main factors influencing collaboration are known. Statistically significant results can pinpoint these relevant factors accurately. For example, quantitative data provided important insights about the correlation between presence and performance in different settings (Heldal et al. 2005). As we have seen, these quantitative methods can be usefully complemented by means of qualitative analyses. For example, the answers to open-ended questions combined with observers' observations presented here not only give a more nuanced account of collaboration, but also mean that certain factors contributing directly to more effective collaboration can be analyzed in a more complex way. Examples include the synchronization between speaking and doing from different partners, the ability to identify common points of reference, and the differences in how social interaction is better or poorly supported during certain phases of the collaboration – or interaction with the devices and the objects in other phases.

One direct implication of the evaluations of collaboration presented here is that it is critical to examine how different activities are *jointly* coordinated; that is, to look at how verbal communication is synchronized with collaborating via the different devices and the virtual objects – for example how people follow each others' verbal instructions and how this is indicated in words and movements. This is a more complex activity than that described by Hindmarsh et al. (2001) and Heldal (2004) since it refers to the process of coordinating speaking and doing, which means looking at the two activities of two people in concert. Since qualitative data provides useful information about this coordination, one question then becomes whether this data is sufficient, or if it needs to be used in conjunction with repeated viewings of video- and audio recordings.

5 Conclusions

We have provided a number of illustrations of how qualitative data can complement and add to quantitative data in evaluating different distributed VE systems for collaboration. These data will provide a better understanding of the advantages and disadvantages of the different systems. In particular, qualitative data can provide more insight into how people experience working together. It is clear that there are quite different disadvantages and advantages of collaboration in each type of system. Comparing the comments about the hindrances and benefits and about collaboration in tables 3– 5 with the summary in table 6, for example, it is easy to see that users picked up on quite different sets of issues in the different systems. Importantly, these concern not only the technology, but also how people work in relation to each other. We have also seen that it is useful to obtain several complementary sets of qualitative data – the responses of participants, aggregated sums of these responses, and observations made by observers. These three combined can address the gaps in each: the subjective point of the participant which is difficult for the observer to get a sense of (for example, frustration with one's partner), an aggregate picture of issues participants identified which does not emerge from their individual comments (how much do these individual comments weigh in the overall balance?), and things that participants have difficulties noticing or being aware of but that observers notice (for example, moving around and handling objects in a suboptimal way).

It seems that users are quite articulate about certain quite nuanced aspects of working together. They voice their frustration about how they are coordinating their activity with their partner, and yet they are unaware of certain obvious others, such as how they coordinate speaking and doing things. The latter, as we have seen, are more easily detected by observers. The implication is that a multi-facetted analysis of people's experiences is needed in order to elicit a well-rounded understanding.

For collaboration, the findings that can be obtained by different methods give quite different pointers to how well or poorly the collaboration worked: performance can discriminate between different systems but cannot, for example, capture the fact that people experience asymmetrical settings in a problematic way because they misunderstand what their partners can do. Further, how partners contribute cannot be captured by crude measures such as how much, in percentage terms, each contributed. And finally, we have seen that what the experimenters observed was sometimes at odds with what participants reported about their collaboration.

From Table 6, there were many problems that were technical in nature, such as control using the devices and update rate and quality of the display. However some of these technical problems are quite hard to alleviate and it may be that some assistance can be given to the participants so that they can coordinate themselves better using the existing technology. For example, certain problematic issues arose because users were not aware of the capacities that their partners had with their systems. One easy improvement for systems may therefore be to raise awareness of aspects of the system that are otherwise 'hidden' to users because, for example, they are not familiar with the features of the system that their partner is using.

Analyzing qualitative data can thus give a rich picture of what users find most and least problematic about working together; whether it is different features of the system, of the task, or of working with their partners. It is clear that collaborating in different symmetrical and asymmetrical distributed VE settings has different hindrances and benefits which quantitative data have so far been unable to pinpoint. Further work is needed to synthesize the findings from the different methods in order to draw out the implications for usability and system design. What has been shown, however, is how, for collaboration, a combination of methods can be used to arrive at a well-rounded understanding of how people work together, and how different aspects of collaboration can be elicited that quantitative measures, responses by subjects, and experimenter observation on their own cannot achieve. Future research which systematizes qualitative data will be able to address this shortcoming and provide useful guidelines for improving distributed VE technologies and their uses.

Acknowledgements

Thanks to Lars Bråthe, from Volvo Powertrain for his advices, and to Ann-Sofie Axelsson, Maria Spante, and Josef Wideström for much help during the trials.

References

Arthur, K. W. & Booth, K. S. (1993). Evaluating 3d task performance for fish tank virtual worlds. *ACM Transactions on Information Systems,* 11, 3, 239-65.

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. & Raja, D. (2004). A method for quantifying the benefits of immersion using the cave. *Presence-Connect,* 4, 4.

Churchill, E. F., Snowdon, D. N. & Munro, A. J. (eds.) (2001). *Collaborative Virtual Environments Digital Places and Spaces for Interaction.*, London: Springer.

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.

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

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

Heldal, I. (2004). The usability of collaborative virtual environments: Towards an evaluation framework. Ph.D. thesis, Chalmers University of Technology, Gothenburg.

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

Heldal, I., Schroeder, R., Steed, A., Axelsson, A.S., Spante, M. & Wideström, J. (2005). Immersiveness and symmetry in copresent situations. Accepted at *IEEE VR 2005*, March 2005, Bonn, Germany.

Heldal, I., Steed, A., Spante, M., Schroeder, R., Bengtsson, S. & Partanan, M. (2005). Successes and failures in copresent situations. *Presence: Teleoperators and Virtual Environments*, 14, 5.

Hindmarsh, J., Fraser, M., Heath, C, & Benford, S. (2001). Virtually Missing the Point: Configuring CVEs for Object-Focused interaction. In Churchill, E. F., Snowdon, D. N. & Munro, A. J. (eds.). *Collaborative Virtual Environments Digital Places and Spaces for Interaction.*, London: Springer, 115-139.

Meehan, M., Razzaque, S., Whitton, M. C. & Brooks, F. P. (2003). Effect of latency on presence in stressful virtual environments. *IEEE Virtual Reality Conference 2003*, Los Angeles, USA, 141-48.

Neubauer, B. J. & Harris, J. D. (2003). Immersive visual modeling: Potential use of virtual reality in teaching software design. *The Journal of Computing in Small Colleges*, 18, 6, 142-50.

Sadowski, W. & Stanney, K. (2002). Presence in virtual environments. In Stanney, K. M. (ed.) *Handbook of virtual environments. Design, implementation, and applications.* (791-806) London LEA.

Schroeder, R. (ed.) (2002). *The Social Life of Avatars: Presence and Interaction in Shared Virtual Environments.* London: Springer.

Schroeder, R. & Axelsson, A.S., (Eds.). (2005). *Avatars at Work and Play: Collaboration and Interaction in Shared Virtual Environments*. London: Springer.

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. (2003). How colourful was your day? Why questionnaires cannot assess presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, 13, 4, 484-496.

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*, 163-72.

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.* (153-64) Wien: Springer-Verlag.

Steed, A. & Parker, C. (2005). Evaluating effectiveness of interaction techniques across immersive virtual environment systems. Forthcoming in *Presence: Teleoperators and Virtual Environments*, 14, 5.

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

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.

Wideström, J., Axelsson, A., Schroeder, R., Nilsson, A., Heldal, I. & Abelin, Å. (2000). The collaborative cube puzzle: A comparison of virtual and real environments. *Proceedings of CVE 2000*, San Francisco.

Youngblut, C. (1998). Educational uses of virtual reality technology, Institute for Defense Analyses 1801 N. Beauregard St. Alexandria, VA 22311-1772.

Zimmons, P. & Penter, A. (2003). The influence of rendering quality on presence and task performance in a virtual environment. *IEEE Virtual Reality Conference 2003 (VR 2003)*, Los Angeles, USA.