COllaborative Virtual ENvironments:
Experiments on Small Group Behaviour in the COVEN Project

Jolanda Tromp[1], Anthony Steed[2], Emmanuel Frécon[3],
Adrian Bullock[1], Amela Sadagic[2], Mel Slater[2]

Abstract

This paper describes work being undertaken by the multi-partner European Union funded project COVEN (COllaborative Virtual ENvironments). COVEN is developing collaborative virtual environment (CVE) systems and technologies in order to support a group of travel rehearsal applications. One of the major themes of the project is the evaluation of CVE applications through extensive network trials, usability evaluations and in depth case-controlled experiments. This paper concentrates upon an experimental study where groups of three strangers meet in a CVE to carry out a simple collaborative task. The experiment was designed to investigate issues about group behaviour such as the relationship between emergent leadership and computational resources, presence of being in a place, and the co-presence, the sense of togetherness, amongst the participants. This experiment and others undertaken by the project highlight a number of issues to do with participant representation, interaction style and system implementation that need to be addressed in future CVE development.

Keywords

Virtual reality, collaborative virtual environments, user evaluation, small group behaviour, presence

1. Introduction

The Collaborative Virtual Environments (COVEN) project, is a four year European project that was launched in October 1995, to design and explore collaborative virtual environment (CVE) technology. Its aim is to investigate the feasibility of scaleable CVE worlds through the development of CVE systems development and the demonstration of prototype applications in the area of virtual travel rehearsal.

We work from the premise that there will always be a limit to available computing and communications resources with resulting tradeoffs between realism and interactive performance. As a result, we prioritize specific user and application needs and then find ways of supporting them within a limited computing resource. Thus our approach to design has started with relatively simple worlds and embodiments, while incrementally

[1] University of Nottingham, UK
[2] University College London, UK
[3] Swedish Institute of Computer Science

introducing richer solutions, approaching anatomically correct embodiments, geometrically correct architecture, geographically correct locations, and real tourist information. Our suggestions for improvements are driven by requirements analysis and usability evaluation, and not by a blind desire for photo realism or a totally immersive platform.

In this paper we outline a recent set of experiments designed to explore what happens when people who have never met before, meet for the first time in a CVE. For instance: what social interactions develop, and how similar or different are these to what happens when they meet to carry out the same task in reality? These issues are of fundamental importance to the potential future use of this technology. CVEs will not be used, for example, if it is found to distort 'normal' social relationships, independently of the technological issues. Alongside this discussion of social behaviour we outline the technical background of COVEN - what makes these experiments possible, what services the system must provide and how it is possible for several people at remote sites to meet together in a shared VE.

2. The COVEN Platform

The technical basis of COVEN is formed by CVE systems developed in two of the partner institutions: DIVE from the Swedish Institute of Computer Science (SICS), and dVS from Division Limited. Both applications are mature toolkits, and both support some fundamental services in order to provide a shared environment for collaborative work and play. These include support for awareness between users and ability to communicate. DIVE is a general tool for the exploration of future CSCW applications and within COVEN we are using it to prototype CVE services. dVS is focused towards robust collaborative engineering applications and we are using it as an application delivery platform. Thus application content is developed for dVS, whilst the services required to supported large scale CVEs are in prototyped in DIVE with the requirement that the required functionality will filter through to dVS.

Both DIVE and dVS provide basic humanoid embodiments to represent participants in their systems. Replacement avatars that possess rich interactive capabilities and a greater degree of human likeness have been developed within the COVEN project [4]. These avatars can be used with both DIVE and dVS.

2.1 DIVE in COVEN

The Distributed Interactive Virtual Environment (DIVE) is the experimental platform used within the COVEN project. DIVE is especially tuned to support multi-participant virtual environments over the Internet.

At the networking level, DIVE is based on a peer-to-peer approach, where peers communicate by reliable and non-reliable multicast based on IP multicast. Conceptually, all peers share a common state that can be seen as a memory shared over a network. Processes interact by making concurrent accesses to that memory [3].

DIVE is fully integrated with the World Wide Web. Any file or document necessary to a DIVE session can be accessed using the http or ftp protocols. DIVE supports 3D formats such as VRML, and 2D image formats such as GIF, JPEG and PNG. In addition DIVE can visualise web documents using MIME compliant mechanisms. DIVE also supports live audio and video communication between participants. Sounds are spatialised and video streams can be texture mapped on to objects in the virtual scene.

In a typical DIVE world, a number of actors, i.e. the representations of human users leave and enter dynamically. Additionally, any number of applications exist within a world. Such applications typically build their user interfaces by creating and introducing necessary graphical objects. Thereafter, they “listen” to events in the world, so that when an event occurs, the application reacts according to some control logic. Events can be user interaction signals, timers, collisions, etc.

2.2 dVS/DVISE in COVEN

The dVS/dVISE system from Division Ltd is the delivery platform used within the COVEN project. Its architecture consists of a set of application processes, called “Actors” that are connected by a distributed database [10]. Services provided by the Actors include rendering, sound spatialisation, collision detection, Newtonian physics, 3D position tracking and input device handling. dVISE then provides an application specific environment authoring and interaction tool that runs within dVS.

In single-user mode, the Actors that make up the runtime environment are normally connected to each other via a shared memory database. For multi-participant systems, some of the Actors are replicated for each participant to provide the local services required for that person to visualise the virtual environment, while some other Actors are shared by all participants at a master site. Thus, the multi-user architecture is based on a peer-to-peer unicast approach.

A typical dVS multi-user session is controlled by the master session, which passes the state of the environment to each client as it starts up. While dVS supports audio from the application to the participants, there is as yet no direct support for audio communication between participants. Hence an auxiliary audio system, the Robust-Audio Tool (RAT) v.3.0.23 [5], is used alongside dVS.

3. The COVEN Applications

COVEN’s application area is virtual travel rehearsal, and the project has developed two distinct applications that assume different classes of user and different requirements for collaboration.

The *business conferencing* application supports two main tasks, audio-visual presentations and a business game simulation. The application supports text message passing publicly or to named groups of receivers (users or message boards) to enable a basic level of communication, and participants are aware of each other through avatars which carry name tags. The presentation support is provided in a virtual conferencing suite shown in Figure 1. In this room, users can communicate by passing objects to each other, or to a large-screen display. The picture shown on the large display seen in Figure 1 is a view of the spreadsheet simulation.

Figure 1: A view of the business application

The *citizen application* consists of visiting a virtual travel agency in order to retrieve holiday information through 3D media (a visit to a virtual version of the holiday destination), and 2D media (such as web pages and audio-visual presentations). It consists of three zones, the virtual travel agency, which has doors leading to different zones for each possible destination. Only one of these destination specific zones exists at this time; this is the Rhodes Zone. In this room, groups of people can access information, and take advantage of the communications services to plan a group excursion (see Figure 2). The individual or group can then use the teleporter (seen on the left in the figure) in order to travel to a virtual version of Rhodes which includes visualisations of tourist information data, and re-constructions of ancient sites, see Figure 3.

Figure 2: A view of the citizen application

Figure 3: Two users in the virtual flight zone over Rhodes.

4. Usability Studies

Collaborative Virtual Environments technology that poses a problem when trying to apply existing HCI evaluation methods. Indeed there has been a general tendency to ignore or minimise VE evaluation as noted in [1]. In the absence of an existing dedicated CVE evaluation methodology, our first task was the design of a framework for the COVEN usability evaluation, addressing the methodological constraints specific to CVE evaluation [12]. The main constraints are caused by the novel interfaces that are presented to the user, the prototypical nature of CVE applications, and the breadth of social interaction that might take place.

4.1 Methodological Framework

A number of aspects of CVE technology direct and constrain our evaluation approach. Firstly, the COVEN project is developing applications that are intended to demonstrate the added value of the CVE concept from the end-user and the customer point of view. This requires an evaluation of the overall usability of the CVE applications, together with elements of a cost/benefit analysis. This places our evaluation activities in the general framework of standard usability engineering [7]. However, standard usability engineering methods are developed for 2D applications and cannot be expected to be directly applicable to 3D, collaborative applications.

Secondly, the CVE technology on which these demonstrators are built is in its early stages. In particular the human factors impact of VE technology is still poorly explored [1]. Investigating the human behavioural aspects, which influence performance and satisfaction in CVEs, is precisely one of the objectives of COVEN. This requires focused exploratory studies of specific phenomena, such as situational awareness, presence and co-presence.

Thirdly, due to the three dimensional, distributed and prototype nature of CVEs there are a number of factors that further characterise the specificity of CVE usability studies. CVEs attempt to create a 3-dimensional place in which people may interact, so we are not only interested in human behaviour and performance with the application, but also in human behaviour and performance inside this CVE. Thus, in general observations and experiments need to be performed both from outside the CVE and from inside the CVE. Furthermore, a CVE allows multiple geographically distributed users to interact simultaneously within the CVE in real-time, regardless of the physical location of these users. As a result, network traffic influences system performance, which forms a research topic in itself and is addressed in our network trials. Another implication due to the geographic distribution of the subjects is that it is rather complicated to conduct proper controlled experiments, such as the one reported in this paper. The prototype nature of the applications becomes another methodological issue in that it is often not feasible within the time and effort available to create different conditions for experiments. Therefore the process of scientific inquiry is constrained to the given state of the applications at the time of testing. Finally, remaining serious defects in the functioning of the application might only be found during the network trials, which means that it is often hazardous to conduct end-user experiments.

These methodological considerations form the basis of our framework for evaluation. From this framework three main threads of work were derived:

- Usability inspections of the applications using inspection methods adapted to 3D collaborative tasks.
- Observational evaluations of participants performing tasks in networked trials.
- Case-controlled experiments exploring CVE concepts such as *presence*, *co-presence,* and *collaboration*.

From the traditional usability engineering approach we identified the heuristic evaluation and cognitive walkthrough methods [8] as suitable techniques for evaluation of our prototypes. These techniques provide relatively quick feedback about the main design flaws, and no experimental subjects are needed

Network trials were found to be useful for collecting longitudinal network traffic data, exploring CVE technology concepts, exploring how people use the various communication media, and to what extent they can collaborate successfully.

Additionally, auxiliary case controlled experiments have been very useful to systematically explore CVE technology concepts and CVE user behaviour. The small group experiment reported in this paper, fits within this thread.

4.2 Network Trials

Network trials take place on a weekly basis. In total 60 trials have been held in two phases by the end of September 1998 with each trial lasting approximately two hours. Data is gathered by logging network traffic data from each site, video and audio from participants' screens, and questionnaires posed after the end of each trial. The four participants during the first phase of the trials were all experienced in the use of CVEs, though the particular platform was new to some. In the second phase the number of participants was increased to eight and trials were run with both novice and expert users.

The primary goal of the trials is to assess the system and network performance of the COVEN platforms and applications. In turn, this enables us to understand the more general network infrastructure and bandwidth requirements of CVEs.

4.3 Experimental Task Scenarios

The experimental tasks that subjects are asked to perform during the network trials are aimed to allow us to examine the main COVEN applications functionally, and to address the scalability problems associated with these services. During the first iteration of the COVEN usability studies two broad exploratory tasks (e.g. experiment 1 and 2 in Table 1) took place, which enabled us to identify and subsequently isolate concepts for further exploration. New experimental tasks (3-7 in the table) were then designed to allow focused explorations of the human factors issues associated with these concepts, see Table 1. These experiments are taking place during this 2nd phase of the COVEN usability and network trials, and they are aimed at providing us more in depth understanding of the CVE concepts. Experiment 3 in Table 1 is the main experiment reported in this paper.

4.4 Usability Inspection and Network Trial Results

The usability inspection provided us with an extremely useful list of usability problems, although the method clearly needs to be developed further in order to be able to address usability problems associated with 3D, multi-user issues. To this end several key points which future inspections methodologies should try to address are briefly discussed next. Once the inspection issues were collated and it was found that they could be separated into three classes: system, interaction and application specific problems.

System problems include lack of functionality, performance and display quality issues. These issues pervade all applications built upon the system. A typical issue is that of the simulation slowing or stopping for a fraction of a second when new scene components are loaded during exploratory navigation of the environment. This is an issue for which many VE systems developers are trying to find solutions. The important point here is that interaction and navigation can become temporarily impossible, as a result of a seemingly innocuous action. Given the participant's expectation of free movement at all times, this suggests to the participant that an error has occurred, or that the operation failed, leading to a serious interruption of their sense of presence. This is also potentially serious for immersed people since the visual and proprioceptive cues will conflict, which could also lead to simulator sickness. This points at *latency, inconsistency,* and the associated confusions over *causality* as a key issue for future CVE research and development.

Table 1
Experimental Tasks

	Experiment	Concepts Explored
1.	Business Trading Game	Communication services
2.	Plan a holiday	Small group collaboration
3.	Word Hunt	Communication in small group, object centred interaction, representation of avatar, differences for desktop and HMD users
4.	Treasure Hunt	Non-verbal communication, level of detail on avatars, use of explicit focus, object manipulation
5.	Murder Mystery	Object manipulation, group navigation, shared and private communication
6.	Switching between Real and Virtual Worlds	Managing 2 embodiments, absence vs. presence, spatial behaviour of small group
7.	Travelling back	Mental model, group navigation, conceptual knowledge, information visualisation

Interaction problems concern the actions of navigation, and selection and manipulation of objects. The most fundamental issue with the interface of the COVEN demonstrators is the multiplicity of layers at which interaction occurs. There are aspects of all of the following in use in the non-immersed applications: keyboard input, 2D Widget interaction, continuous mouse driven 3D control, and discrete 3D Widget interaction. Deciding which “level” of input is required was considered a major burden. Many quite simple actions would involve several interaction stages, and much of the difficulty was due to the need to select suitable viewpoints from which to perform certain tasks. This is a fault found with most VE desktop interaction systems that support free navigation. Integrating all controls into one coherent structure is difficult on a desktop display because of the relative difficulty of providing an easy to use 3D control. It is also difficult in an immersive system because of the need to provide alternatives to text. This points at the broad issue of *3D interaction for design and usability* as a second key issue for future CVE research and development.

Application problems concern the actual actions and meaning of objects within the environment. The application issues are broad in nature, from problems with objects whose operation is not obvious such as the teleporter, to wider topics such as how best to represent group services to group members. The application issues can be classified into issues concerning the collaboration services and more application specific problems, which are too detailed to describe here in detail. The main issues that are raised about the collaboration services include problems with the communication channels. In particular, problems were found with the lack of error feedback when sending text messages to named recipients, inconsistency of object reaction to selection and manipulation, mutual awareness between participants, and the mechanisms of group formation. This points at the need for *guidelines on which object affordances to support* as a third key issue for future CVE research and development.

The results from the usability experiments during the network trials (tasks 1 and 2 in Table 1) put strong emphasis on a few inspection issues. Notable amongst these was the occasional lapse in synchronisation between events in different media. This vindicates the extra effort involved in studying user behaviour and opinions at an early stage in the project, even if those users are not typical end-users.

General feedback from the network trials yielded detailed information on audio communication and small group collaboration on two levels: network traffic and human factors. There were three communication channels in the trials: audio, text and visual information. All three proved valuable for different reasons, but actual usage was heavily influenced by the perception of reliability of the channel. The audio channel was regarded as useful but not totally reliable. In particular not knowing if the others had received one's audio segment was found to be a problem. The reliability of the text communication was found to compensate for lack of clarity in the audio, thus vindicating the provision of both channels. In addition it was found to be a problem that the avatars of the users did not animate when the person was speaking or typing. As it is, spatial proximity in the VE is neither a sufficient nor necessary condition for two participants to be communicating. Thus a final key issue is the need for the *occurrence of text and speech communication to be reflected in the representation* of the user within the CVE.

These key issues are some of the ones that COVEN is addressing in its ongoing technical program. However more work is required on isolating the key requirements that underlie the issues we have raised.

5. Experiments on Small Group Behaviour

The subjects in the COVEN network trials discussed above were usually members of the COVEN project itself. Of more interest in the longer term is the evaluation of the CVE systems with non-expert users. In this section we present one of a series of experiments to investigate the behaviour of subjects involved in a collaborative task.

The following questions were of interest for these experiments:

- Does computational advantage confer social power?
- Is the sense of presence of being in the virtual place associated with 'co-presence' - the sense of being and acting with others in a virtual place?
- How does the sense of enjoyment and feelings of group affection vary as between the virtual and the real experience?
- Can reactions such as embarrassment, shyness and conflict be generated in the virtual environment, and if so to what extent does this carry over to the real?

These questions are important for practical reasons. People will generally use heterogeneous networked machines to meet in shared VEs. If this is the case, and the one with the 'most powerful' facilities (however these might be defined) has a social advantage, then this should be understood by the participants - especially where it is thought important by the group members to maintain certain predefined social roles.

For many applications it is essential that participants develop a sense of presence in the place depicted by the VE, this in order that behaviour appropriate to the situation be supported (for example, in a virtual therapy session concerned with 'fear of flying'). At the same time for effective joint working a sense of co-presence - the sense of 'togetherness' amongst the participants - is also essential [2]. For practical purposes in the design of such CVE systems it would be useful to know whether these two types of presence are associated, and if so whether they are mutually supportive or perhaps generated by the same underlying factors.

Positive affective relationships between group members (a high *'group accord'*) may be an important factor in overall group task performance. It is important to know what factors hinder or strengthen the development of such affect, and how similar or different this development would be in comparison to real meetings.

Finally, an important indicator of the 'reality' of virtual meetings is the extent to which everyday social reactions might be generated in the virtual - such as shyness, embarrassment and other social phobic responses. The greater the extent to which these can occur, the more 'similar' virtual meetings might thought to be to real ones. Again the factors that might inhibit or strengthen these everyday reactions are important to understand.

Two experiments have been carried out regarding these issues, both employing the same scenario but under different conditions. The first study is summarised below and fully described in [11]. The second study builds on top of the first study, and further exploits the capabilities of the COVEN platform.

5.1 The Experimental Scenario

Groups of three strangers meet for the first time in a VE. The subjects have avatars labelled as Red, Green and Blue, and they call each other by these names throughout the experiment. They meet in a small room, and have to find their way together to another room that has a series of puzzles written on pieces of paper stuck around the walls. On each piece of paper is a set of words or phrases each prefixed by a number. The subjects have to rearrange all the words corresponding to the same number in order to form a well-known saying. This task was chosen since is requires a good deal of collaboration in order to solve. One person cannot easily remember all the words, and the task can be partitioned by each subject taking a different area of the room.

5.2 Summary of the Initial Experiment

The first study was carried out at UCL under using a CVE supported over a local network. All subjects were physically in the same large laboratory, but care was taken that they did not meet one another or even know that they were in the same location. There were 10 groups involved. In each group the Red person was immersed using a Silicon Graphics Onyx with twin 196 MHz R10000, Infinite Reality Graphics and 64M main memory, running Irix 6.2. The tracking system has two Polhemus Fastraks, one for the HMD and another for a 5 button 3D mouse. The helmet was a Virtual Research VR4 which has a resolution of 742×230 pixels for each eye, 170,660 colour elements and a field-of-view 67 degrees diagonal at 85% overlap.

The Green subject used a SGI High Impact system with 200Mhz R4400 and 64MB main memory. The scene was shown on the full 21 inch screen display. Navigation was accomplished by using the keyboard arrow keys, with up and down arrows giving forward and back movement, and left and right keys providing rotation. All movement was on the horizontal plane of the floor.

The Blue subject used an SGI O2 running at 180Mz on Irix 6.3, with an R5000 processor, and 32MB main memory. The scene was shown on a full 17 inch screen display. Navigation was the same as for the SGI Impact.

The virtual reality software used throughout was DIVE 3.2. The sound system used was the Dive-Audio Tool. A basic DIVE avatar was used for each of the participants. An image of such an avatar is shown in the left of Figure 4.

Figure 4: A selection of the avatars used in the two experiments.

The avatars were simple, with limited movement and no capability for any kind of emotional expression. The embodiments looked identical, apart from colour.

The total scene consisted of about 3500 polygons, which ran at a frame rate of no less than 20 Hz in stereo. The latency was approximately 120 ms.

The Green subject was given an additional task, not revealed to the others. Green was asked to monitor Red as closely as possible, always trying to be in Red's line of vision, although taking part in the puzzle-solving task as much as possible. If Red objected Green was to comply temporarily with Red's wishes, but then continue anyway with this monitoring task. The purpose of this was specifically to examine the question related to the engendering of embarrassment referred to above.

After about 15 minutes the virtual session was terminated, and the subjects completed a questionnaire, which took about 10 minutes. Then each subject was required to put on a waistcoat of their colour and then they all met together in real life for the first time just outside the real room which had the real pieces of paper placed on the walls.

They were then invited to continue the task in the physical location, which lasted for about another 15 minutes. At the end of that time they completed another questionnaire, and then met with the supervisor for a face-to-face debriefing.

An analysis of the questionnaire results, including a comparison of the virtual and real meetings showed the following statistically significant results:

- Immersion enhances leadership capability: the immersed person was overwhelmingly recognised as leader in the virtual session, but this disappeared in the real session. This was confirmed by a separate question on which person did the most talking - invariably the Red (immersed) person.
- Presence (being in a place) and co-presence (being with other people) were positively correlated.
- Reported presence was not significantly different between the immersed and non-immersed people.
- Group accord increased in the real session compared to the virtual (though it is not possible in this study to rule out the effect of time).
- Higher individual accord was associated with higher overall, place- and co-presence.
- Individual accord tended to be higher for females than for males, and was positively associated with more successful performance of the task.
- There was no reported effect of the attempt to deliberately introduce some ‘embarrassment’ into the virtual session by having one subject monitor another - no differences between the three role-colours were reported on any component of accord.

The debriefing sessions were carried out with the three subjects (Red, Green and Blue) simultaneously - i.e., they were group discussions rather than individual de-briefings. These revealed additional, more dynamic information, hidden behind the inevitably static questionnaire responses. It turned out that there was a significant impact of the monitoring task of Green on Red in three of the ten groups. In one group Green and Red were mutually hostile - each thinking that the other was deliberately spoiling things. In another Blue, seeing Green and Red often ‘together’ felt excluded from the social interaction. In another Green was concerned about what the other two might think of her strange behaviour of always ‘staring’ at Red.

One thing reported by almost all Green subjects was the difficulty of carrying out the monitoring task at all. Red moved faster than the other two subjects (on the more powerful machine and immersed). Also it was difficult for Green to know Red’s field of view. There being no virtual equivalent of ‘eye contact’ in any meaningful sense, Green could never know whether or not Red was aware of Green’s activities - there could be no ‘exchange of glances’. More generally several people mentioned this lack of feedback about body movements and body language from the avatars.

A major issue explored in the de-briefings was the relationship of the people to their avatars. The most interesting way in which this was realised was through projection - that is, individuals were respectful of the avatars of the other people, and tried to avoid carrying out actions that would cause distress or be impossible in real life.

This process of being mindful of the avatars of others was surprising, they were taken seriously in spite of their simplicity. This relationship to the avatars was noticed in another way - the surprise that some people experienced on meeting the real person.

This analysis of the post-experimental group discussion revealed a surprising degree of attachment and relationship towards the virtual bodies (avatars). Although, except by inference, the individuals were not aware of the appearance of their own body, they seemed to generally respect the avatars of others, trying to avoid passing through them, and sometimes apologising when they did so. If even these simplistic avatars can evoke such responses It is interesting to wonder what responses more powerful avatar representations might evoke, and we tried to take this into account in the set-up of our second experimental.

5.3 The Second Small Group Experiment

The second small group experiment focused more directly on social interaction that would occur 'naturally' in a more realistic use of the CVE. In this case the three subjects were separated by considerable distances - the Red person was at the University of Nottingham (UK), the Green person was in SICS (Sweden) and the Blue person was in UCL (London). This had as a consequence that only the virtual part of the experiment was carried out.

Heterogeneous IP networks were used to connect all the sites together and the world and bodies which were used were accessed via HTTP from SICS throughout. On top of these networks, we used the DIVEBONE to connect the workstations at both UCL and SICS to Nottingham. For sessions involving peers located in different local networks, DIVE has long relied on the existence of the MBone, the IP Multicast backbone, a structure for interactive multimedia communication over the Internet. The DIVEBONE is an application-level backbone, which can connect sub-islands of the MBone and/or single local networks. The use of the DIVEBONE has been dictated by our previous trials, which has shown that it is more reliable than the existing MBone infrastructure in-between the three sites.

There was no special advantage between the UK sites in terms of network connectivity. Theoretically there is a faster connection, in terms of routing between SICS and UCL than between UCL and Nottingham. However, variability in network traffic cannot result in a reliable prediction as to what the situation was at any particular moment, and for this trial no network analysis has yet been carried out. Moreover, the packets from SICS to UCL were transiting through the proxy server running at Nottingham - so the situation is far from clear. The important point is that it is unlikely that there was a general tendency for two of the sites to be connected more strongly than any other pairing.

The scenario was the same as in the first experiment, though there were several important differences in the set-up:

There were four groups of three subjects each, none knowing each other, and not knowing the purposes of the experiment.

No person was immersed - though there were differences between the machines used - this would help to find if immersion accounted for the leadership effect in the earlier experiment or whether it was more to do with speed of machine. At SICS an SGI O2, R5000 180Mhz with 96MB main memory was used. At Nottingham a High Impact R4400 200Mhz with 192MB main memory was used, and at UCL a High Impact R4400 CUP 200Mhz with 64MB main memory was used. All monitors were 21inch, and the full monitor screens were used.

As in the first experiment, none of the avatars had any expressiveness or limb movements. In this second experiment, in order to examine the potential impact of avatar appearance, one avatar, the one at SICS, was different to the other two (in the first experiment the three avatars used were identical apart from colour). The Red and Blue subjects had basic male caricature avatars, similar to the one in the middle of Figure 4. The third (Green, at SICS) was represented in a very realistic manner, as depicted in the right of Figure 4. It was also smaller than the other two. Note that the SICS (Green) person would see two other avatars of the same cartoon-like appearance apart from colour; whereas each of the other two participants would see one avatar looking very realistic and the other cartoon-like. None of the subjects had any idea as to how they themselves were portrayed.

When the Red or Blue people spoke a wave-like emanation would be emitted from their avatars. This was not the case for Green - another difference between the Green person and the other two.

6. Results

We present results of this study in relation to the general questions posed in Section 5.1. This second study had a limited number of groups compared to the first one. For this second study we were more interested in the qualitative results than quantitative. Sections 6.1 to 6.3 present results from the questionnaires. Section 6.4 reports on the interviews and de-briefing sessions. Section 6.5 reports and discusses the results from the observational analysis of the video-data. Finally, section 6.6 summarises some common findings and highlights the problems found.

6.1 Computational Advantage and Social Power

In their questionnaires subjects were asked to rank the degree of leadership, and the degree of talkativeness, portrayed by each of the three subjects including themselves. In the first study Red (immersed) had emerged as the clear leader in both senses. In this study there were no statistical differences between leadership rankings or degree of talkativeness. This supports the hypothesis that leadership was conferred by immersion, rather than by processing power of the machine, or the impact of the colours themselves (it had been suggested by some commentators that Red emerged as leader because ‘red’ is a leadership colour). However, a salient result here was that Green was not rated as leader by any of the 12 participants in the four groups, and was rated as 'most talkative' by only 2 of the 12 participants.

6.2 Place Presence and Co-Presence

While the first study found a significant positive correlation between reported place-presence and co-presence, this was not repeated in the current study, although with only 12 subjects instead of 30. The overall average place-presence was 6 on a 7-point scale (high score meaning higher reported presence) and the overall average co-presence was 4.

6.3 Group Accord

A group accord score was constructed from a number of questionnaire responses: the degree of enjoyment, the desire for the group to form again, the degree to which there was a desire to meet any of the individual members, the degree of comfort with individual members, and the extent of perceived cooperation of the other members. In the first study this score was positively associated with an overall combined presence and co-presence score, as well as with each individually. In this study there is a significant positive correlation ($R^2 = 0.39$) with the co-presence score only.

6.4 Group Interaction

At each site immediately after the questionnaire had been completed, the local experimenter held an unstructured interview with the participant. A number of issues were explored - their overall comments, their reactions to the avatars, their responses to the other people, and any other issues that arose during the conversation.

We consider each group in turn. In the Table 2, we show a summary of the salient questionnaire responses, where each subject showed any preference in reactions to the other two subjects. A blank entry means 'no preference'. Average enjoyment is measure on a 1-7 (low-high) scale.

Table 2 - Group Response

Group	**1**			**2**			**3**			**4**		
Riddles solved	0			1-2			0-2			0		
Average enjoyment	3			4			5			2		
Responses	**R**	**G**	**B**	**R**	**G**	**B**	**R**	**G**	**B**	**R**	**G**	**B**
Gender	M	F	F	M	F	M	M	M	F	M	F	M
Leader	R		R	R	B	R	R		B	B		
Most talkative	R		R	R		R	R		B	G	G	
Would like to meet	B		R				B				B	
Most comfortable with	B		R		B	G	B				B	G
Most isolated	G		R	G	G			G			R	R
Most cooperative	B		R		B		B	R		B	B	G
Most embarrassed by	G				B							

Group 1. From the table we can see that there was a clear leader (Red), with seemingly a good rapport between Blue and Red.

The Red subject found that the Green avatar was ‘scary’, ‘like a zombie’ because there was a conflict between its greater visual realism but lack of bodily movement. Blue on the other hand was more ‘cartoonlike’ and therefore it was easier to understand the fact that it was not ‘functional’. The female voices in the male embodiments were thought of as ‘weird’.

Green reported problems in making audio contact with Red and Blue, though considered the situation as similar to a three-party phone conversation. Green found contact difficult also in the sense that she did not know the other two people, and generally finds communication easier after she knows them. Green was concerned about the lack of facial expressions and gestures, which would be important in a real meeting.

Blue was concerned about no one speaking at the start of the meeting, although since Red eventually spoke more than anyone else, she felt that Red emerged as the leader. Green seemed more real, but didn’t say much.

Group 2. Again Red appears to emerge as leader, with Green and Blue most comfortable with each other. Note that although Green is most comfortable with Blue, she is also most embarrassed by him. She also puts herself as most isolated, concurring with Red.

Green in this group reported feeling sometimes lost when the others disappeared and she was left ‘alone’. She also reported a language problem (recall that she was in Sweden, the other two in the UK). She was also never quite sure whether the other two people were real or ‘robots’ - but finally, after the experiment, decided that they were real because they had been laughing - ‘something that a robot cannot do’. However, the earlier belief that they were robots meant that she ‘did not find it natural to talk with them’. The appearance of the other two was ‘nice’ - they looked friendly. (Recall that Green would see the other two as the same cartoon-like characters, apart from the colour).

The Blue person claimed to hear the Green one more clearly. When other avatars came very close to him, he found it to be ‘funny’ and ‘as if there were no personal boundaries’. He found this to be ‘nicer - they didn’t mind you being so close - didn’t seem to be offended even if talking right close - just seemed strange.’ He found it different from reality especially because expressions and body language cannot be seen, although a lot could be gathered from the tone of voice. He believed that the Red person was a ‘plant’, and was a natural leader - he seemed to know more about what to do. (In fact in this group Red was a member of the COVEN team and so would be more familiar with the DIVE controls, though he was not aware of the goals of this experiment, and had no prior experience with the environment or scenario).

Group 3. There are no clear patterns that emerge from in this group. Red reported that Green was ‘better because more realistic - seemed to be more important, more relevant somehow than the others' (note that this is not reflected in the questionnaire responses). He reported that 'it seemed to be a virtual body catering for a managing director - a senior person - more privileged than the others' with gender also implied by that body. This even carried over to a belief that Green seemed to know more about the system, because the body was more sophisticated. He reported that 'even though the people were anonymous I still felt the need to be polite'.

Green was hampered by a poor sound quality, and the fact that there were no gestures, only voice. He believed that 'if the sound had been better, communicating through that medium would be a bit better than the telephone'. He said that the three of them could not all talk with one another, but had to follow a 'radio discipline' each waiting for the others to be ready before being able to start talking.

Blue reported as a problem not knowing which voice was which person (there being no spatialised sound). Identifying who was talking was important, but there were no cues to this. She knew that there were two people 'somewhere' but had no identification of these people with their avatars. She found silence to be 'strange' - no chatter, no 'white noise', as would be the case in normal meetings. When there was silence she was wondering why this was the case.

Group 4. Again there is no clear pattern in questionnaire responses. Red reported that 'Green seemed to speak a strange language', and that it was not clear who was whom, and who was involved. He felt that Green didn't seem to understand what to do, got lost, and Red was never sure if he had been heard after speaking. He thought that the audio from Blue was the worst. He thought that standing in the 'radiation' which seemed to come out of Blue when Blue was talking would make the sound easier to hear, but that this meant that 'I had to stand so close to blue's virtual body that it was uncomfortable'. The female voice in Green's male embodiment was strange 'because Green's embodiment is so obviously made realistically male, and the voice so obviously female.'

Green found the experience frustrating because it was so difficult to talk with the other two, and she never understood how to start talking with them. She paid attention to the fact that the visual sound waves from the others indicated that they were talking. She was unsure as to whether the other two were real persons or 'robots'. Their sound was strange and mechanical - and she did not try hard to get in contact with them because she thought that they were machines. She reported talking mainly to Blue, with Red being 'out of contact'.

This group was the most interesting with contradictory impressions amongst the group members. In this group Blue believed that there was no line of communication between himself and Red. He also believed that Red and Green were 'talking in German' to one another, and leaving him excluded. He did, however, manage to talk with Green. He said that the main difference to meeting in reality was the lack of body movement and gesture, although he did have a feeling of being together with the others, similar to the feeling in real life. His reactions in this regard were interesting - he found it especially significant when he ‘saw someone walk by’ - this kind of thing was most reminiscent of real life. However, he returned to the fact that he believed that the other two were talking together with one another in German, and that his reaction was also like that of ‘real life’ - he was somewhat ‘intimidated’, especially when they were laughing, and he wondered if they were laughing at him. ‘You become a minority’ he concluded.

The most important findings of this experiment were as follows: First, no leadership pattern emerged. This is significant in relation to the earlier study, where leadership was clearly associated with the immersed subject. There being no immersed subject here adds weight to the result of the earlier study that leadership capability with respect to this scenario and task is related to immersion.

The statistical relationship between place presence and co-presence was not reproduced in this study. The previous study had found a statistically significant relationship between both types of presence and group accord, but in this study this was found only for co-presence.

There is some evidence that the differing avatar of the Green subject did impact the relationships between the people involved - this was explicitly stated by two of the subjects. However, the effect is not consistent - for one, the greater realism of the Green body was not appropriate to the lack of body movement, and this was found to be disconcerting. For another, the greater realism was associated with prestige. The fact that the clearly male embodiment was associated with a female voice was also remarked upon.

Both cases where a subject thought that the other two were not ‘real people’ occurred for Green subjects - who would see the other two avatars as cartoon-like rather than realistic. The fact that they were not believed to be real people actually reduced the degree of communication offered by the Green person involved.

This study is realistic in the sense that the subjects were at genuinely remote locations, have different nationalities, were (for all but one) not involved in the COVEN project themselves, were all naive as to the purposes of the experiment and the scenario. It was as if three strangers had wandered into a place, with poor communication channels between them, wearing strange garb, with only the ability to move around but no ability to move their bodily limbs, and given an unfamiliar task to do together. They found this to be difficult. Admittedly they had limited time (15 minutes is not long enough to adapt and build strategies for coping). But on the other hand, 15 minutes is enough time for someone, for example, a business manager, considering using this technology to reach a conclusion that such systems are far from capable of achieving collaborative work on a business level.

6.5 Observational Analysis

The video-data have been used to analyse the collaborative activities during the experiment by comparing them with a hierarchical task analysis of collaborative behaviour. A number of generalisations could be made, which are presented below.

According to observations made in [6] and [9] collaboration between people sharing the same workspace - be it virtual or physical - involves the ongoing and seamless transition between individual and collaborative tasks. Thus collaboration can be broken down into unfocused collaboration, where the individual monitors the other participants' activities without getting involved, and focused collaboration, where individuals are closely working together. Both focused and unfocused collaboration are largely accomplished through alignment towards the focal area of activity, such as a document, where individuals coordinate their actions with others through peripheral monitoring of the others involvement in the activity 'at hand' [6].

During the second experiment the subjects were using desktop VR, which has a small field of view of the VE. Problems keeping the shared object and the other participant in the same view are thus largely due to the small field of view (55 degrees), but it also points at a real user need which is currently not met by CVE systems. In order for participants to collaborate smoothly they need to be able to keep both the shared object(s) and the relevant other(s) in the same view. Subjects in the experiments can be seen repeatedly making a sweeping move from object to speaker, and back to object, and so on, during collaboration. They can also be observed having trouble making this sweeping movement smoothly. Often they overshoot their goal, and pay for this mistake by losing track of the conversation. Also, subjects can be seen trying to 'back up' in order to increase their field of view, and therefore overview; encompassing as many objects and subjects in their view at the same time. Usually, this results in them 'falling out of the room', where they back up to the point of going through the wall. They pay for this mistake by inadvertently completely interrupting the conversation, because - even though it makes no difference for audibility - their conversation partner observes and comments on their visual disappearance. Interestingly, one subject in the post- test interview complained that the room was 'too small'. This points to a need for collision detection, larger field of view by using out of body views, automatic tracking of initially selected objects and subjects relevant to a conversation.

There was no pointing device available for the subjects so that they had trouble identifying objects in the room to each other, and hence could not really trust that specific objects had become shared objects. An exacerbating factor to this problem was that the objects under inspection all had the same colour and general appearance.

In order to switch between personal activities and group activities, users need to be able to monitor the ongoing activities in the shared space they inhabit while going about their own business [9]. This peripheral awareness allows them to make smooth transitions between focused and unfocused collaboration.

Due to the lack of ambient sound (sounds generated by objects being moved or manipulated by other participants), once a participant loses another participant from view, there is no way of telling what they are doing, unless the other participants accompany their activities with a running commentary, or until they cross or enter the field of view of the first participant again.

6.6 Discussion

There were several remarks common to more than one subject that were made on the questionnaire, during the de-briefings, or observed on the video-recordings. The most often found were the audio break-ups, and not knowing for sure if the others could hear what was being said, or missing what others had said.

The lack of facial expression and bodily gestures was also a common theme. This was not only concerned with the bodily movements of the others, but also an inability to point at the papers on the walls to indicate which one was being referred to during conversation. Subjects were frequently seen *physically pointing at the real computer screen, smiling, and nodding* while talking about a particular sheet, simultaneously exclaiming ‘this one!’ - as if the other participants could actually see what was being referred to.

Another common theme was concerned with navigation and positioning- this was too coarse, leading to frequent collisions with walls and with one another, sometimes leading to a situation where the participant could become temporarily lost. Especially for collaboration, the need to fine-tune navigation and position precisely in relation to other participants, and shared objects is extremely important. As in the first experiment, the lack of collision detection was an important issue for some of the subjects.

In terms of support for collaboration the common themes observed are that subjects had problems keeping a referenced shared object and other participants in the same view. Subjects had problems identifying the referenced object and also the referee (other participant currently speaking). Subjects had problems monitoring the activities of the other participants while acting or navigating themselves (breakdown of peripheral awareness).

7. Conclusions and Future Work

This paper has provided an overview of the work of the COVEN project. We have given an account of the evaluation work being undertaken by the project with an in-depth look at one experiment on group behaviour and have briefly introduced both the systems and applications development threads. It has been our intention to demonstrate how COVEN is identifying and exploring solutions to key usability issues that have been found in our trials.

The small group experiment described in this paper is interesting for several reasons. Firstly we back up previous results. Secondly we note that the experiment was managed by the experimenters in real-time between three sites in two countries using the DIVE COVEN platform over heterogeneous IP networks - itself a positive comment on the usability of the system to carry out a complex task.

From this and other experiments we have identified several key points for future CVE research and development:

In particular, COVEN is tackling the problem of implementing realistic virtual humans. In CVEs, their visual complexity and naturalness have an effect on the graphics rendering of each participant involved and on the network traffic generated by their animation.

The traditional rendering optimisation techniques fail when it comes to virtual humans. Indeed, these techniques usually simplify graphical representations with the distance. However, from a short distance high levels of details are still necessary to represent virtual humans, and, thus, many vertices and polygons will put a load on the rendering pipeline. This remains a serious problem for multi-participant applications that require realistic avatars, with no clear solution in sight.

At the network level, the COVEN platform will include the notion of virtual humans more deeply inside the application by adapting data treatment to the flow that concerns virtual bodies' animation. That adaptation will mainly consist of specific algorithms such as network message size optimisation and aggregation and adapted dead-reckoning techniques.

Additionally, COVEN intends to achieve a CVE platform that scales, i.e. that will contain many users, in spatially extended, when not infinite, environments with many detailed objects. The COVEN approach to this issue involves abstracting the key properties of scoping in the physical world (e.g., general principles of occlusion and attenuation with distance). These properties are then embodied in a set of lightweight mechanisms based on the idea of spatial world structuring, especially the use of regions. The implementation of these mechanisms requires relatively little computation and is flexible enough to support a wide variety of existing and new communication practices. This approach pioneered by NPSNET and the Spline system has led to the improved COVEN experimental platform, as described in [3].

Acknowledgements

The COVEN consortium consists of twelve partners: Arax Ltd (UK), Division Ltd (UK), EPFL (Switzerland), IIS Ltd (Greece), KPN Research (The Netherlands), Lancaster University (UK), SICS (Sweden), Thomson-CSF LCR (France; coordinator), TNO FEL (The Netherlands), University College London (UK), University of Geneva (Switzerland), and the University of Nottingham (UK).

The COVEN project is funded by the European ACTS Framework. We thank the anonymous subjects who took part in experiments described in this paper, and to all those at Nottingham, SICS and UCL who helped with the experimental arrangements.

References

[1] N.I. Durlach and A.S. Mavor (eds.), “Virtual Reality: Scientific and Technological Challenges”, National Research Council, National Academy Press, USA, 1995.

[2] N. Durlach and M. Slater, “Presence in Shared Virtual Environments and Virtual Togetherness”, BT Workshop on Presence in Shared Virtual Environments, Martlesham, UK, June 1998, http://www.cs.ucl.ac.uk/staff/m.slater/BTWorkshop.

[3] E. Frécon and M. Stenius, “DIVE: A scaleable network architecture for distributed virtual environments”, to be published in the Distributed Systems Engineering Journal - Special Issue on Distributed Virtual Environments.

[4] A. Guye-Vuilleme, T. Capin, I. Pandzic, N. Thalmann and D. Thalmann, “Nonverbal Communication Interface for Collaborative Virtual Environments”, in Proc. of Collaborative Virtual Environments 98 (CVE’98), University of Manchester, 17-19 June, 1998, D. Snowdon and E. Churchill editors, pp. 105-112.

[5] V. Hardman, M-A. Sasse and I. Kouvelas, “Successful Multiparty Audio Communication over the Internet”, Communications of the ACM, Vol. 41, No. 5, May 1998, , pp. 74-80, 1998.

[6] C. Heath, P. Luff and A. Sellen, “Reconsidering the Virtual Workplace: Flexible Support for Collaborative Activity”, Proc. of the Fourth European Conference on Computer Supported Cooperative Work (ECSCW'95), Kluwer Academic Publishers, Dordrecht, The Netherlands, 1995, pp. 83-99.

[7] E-M. Melchior, T. Bösser, S. Meder, A. Koch and F. Schnitzler “Handbook for practical usability engineering in IE projects”, Project ELPUB 105 10107, Telematics Application programme, Information Engineering section, 1995.

[8] J. Nielsen and R.L. Mack, “Usability inspection methods”, John Wiley and Sons, New York, NY, 1994.

[9] T. Robertson, “Cooperative Work and Lived Cognition: A Taxonomy of Embodied Actions”, Proc. of the fifth European Conference on Computer Supported Cooperative Work (ECSCW’97), Kluwer Academic Publishers, Dordrecht, The Netherlands, 1997, pp.205-220.

[10] M. Rygol, S. Ghee, J. Naughton-Green and J. Harvey, “Technology for Collaborative Virtual Environments”, Proc. of Collaborative Virtual Environments 96 (CVE’96), University of Nottingham, September 19-20, 1996.

[11] M. Slater, A. Sadagic, M. Usoh and R. Schroeder, “Small Group Behaviour in Virtual and Real Environments: A Comparative Study”, submitted to Presence, summarised in http://www.cs.ucl.ac.uk/staff/m.slater/BTWorkshop, 1998.

[12] A. Steed and J.G. Tromp, “Experiences with the Evaluation of CVE Applications”, Proc. of Collaborative Virtual Environments 98 (CVE’98), University of Manchester, 17-19 June, 1998, D. Snowdon and E. Churchill editors, pp. 123-130.