Chapter 9 Supporting Works by Author

"Edgar Allan Poe once argued that a certain chess-playing 'machine' had to be fraudulent because it did not always win. If it were really a machine, he argued, it would be perfectly logical – and therefore could never make mistakes! What is the fallacy in this?"

Marvin Minksy (1986).

Chapter 9 Supporting Works by Author

This chapter reviews and discusses relevant material from published papers and parts of COVEN deliverables, from 1997 to 2001. These were written or co-written by the author of this thesis, and are part of the support for the wider interpretation of the results of the experiments presented in this thesis.

9.1 Introduction

The work presented in this chapter aims to support the choice of directions that were taken during the research; to capture the richness of the methodological considerations and research findings, and to present the personal conclusions the author derived from these experiences. The papers summarized here, are largely presented in chronological order, with the exception of documents that cover the same topic in iterative development over longer periods time, which are presented together. It is made clear where and which parts of the papers have been written solely by the author of this thesis, and which parts included major efforts and contributions from others.

Section 9.2 presents work that was conducted by the author prior to the COVEN project, on the JISC/BT funded “Inhabit the Web” project that was similar in its methodological set-up: a longitudinal network/usability test of a locally produced CVE called MASSIVE (discussed in Chapter 2). The local experiences on this project largely provided the professional expertise that she and her local colleagues at the University of Nottingham, brought to the COVEN project. Finally, section 9.3 describes the various COVEN papers and deliverables relevant to this thesis.

9.2 Pre-COVEN Papers

Three published papers are discussed here. The first paper (section 8.2.1) covers issues of methodology for the evaluation of virtual environments, which were raised from a review of available interface design and evaluation methods and a critical analysis of the type of evaluations needed for CVEs, and the limitations inherent in CVE technology, CVE usage, and the open issues in CVE design. The second paper (section 9.2.2) presents the evaluation results of the 3-year BT/JISC funded "Inhabit the Web" Project, which focused both on network requirements and on usability for CVEs, although its emphasis is more on the network requirements than the usability findings. The third paper (section 9.2.3) presents the usability findings of the "Inhabit the Web" project in greater detail, embedded in the available usability requirements at the time the paper was written, and proposes an automated virtual body language tool for CVE users.

9.2.1 Issues of Methodology

"Methodology for Distributed Usability Evaluation in Collaborative Virtual Environments" (Tromp, 1995), was presented as a poster during the 4th United Kingdom Virtual Reality Special Interest Group (UKVRSIG) conference, and as a paper in the proceedings of this conference. The paper argues that there is a need for the development of a methodological approach to the evaluation of VR, and CVEs. The author states that CVE technology is a relatively young multidisciplinary science, which aims to discover new concepts within the field of human communication technologies. The concepts introduced by CVE technology are not fully identified, understood, or placed within a model of usage. A model is needed to come to operational definitions used in empirical research. Operational definitions are essential for the measurement of human behaviour, because they specify how to measure a concept by turning it into a variable, which can be assigned values such as high, medium, and low. These variables should be part of a theory about CVE usage, which can be tested by performing experiments. The paper goes on to argue that our understanding of CVE concepts is so limited that we need to employ exploratory research approaches in order to identify those aspects of CVE technology and human behaviour which affect performance and satisfaction. Two exploratory research techniques are identified as most suitable: observation and open-ended questionnaires. The author points out that there are two ways in which CVE evaluation differs from traditional HCI evaluation. Firstly, to test for CVE usability does not just mean testing usability with the application interface, but also how well human needs are supported inside the application. Secondly, instead of one goal per application, CVEs seem to try to satisfy at least two goals: the general goal of creating a sense of presence, and the specific goal of allowing users to manage multiple tasks of a collaborative nature within the CVE.

The author lists four threats to the validity and the reliability of experiments with CVEs. These threats, summarized in table 9.1 need to be addressed carefully, and minimized, in order to be able to interpret the results from experiments with any value for future research.

Validity/Reliability Threats	**Description**	**Solution**
Selection threats	The application is often unstable so that real end-users cannot be used easily. Often the developers and their	Developers are experts in the technology and as such may be able to provide more insightful comments than end-users. During the developmental

	nearest colleagues are used when testing, instead of a random selection from the population of representative subjects.	stages of the software this type of subject should not pose a real threat as long as the evaluator is aware of the bias, the data gathered is interpreted with this knowledge in mind, and presentation of the results is includes mention of the type of selected subjects.
Ecological validity threats	CVEs are multi-user applications. Users are typically geographically distributed. If the application is not tested on this level, we are not measuring true usability.	CVE needs to be tested on at least two levels: single-user interface; multi-user interaction in geographically distributed setting.
Threats to internal validity due to distributed setting	Due to geographical distribution of subjects in the experiment the researcher cannot have full control over the experimental setting.	Recruit assistants at each site who have knowledge of usability evaluation issues, who can protect and control the subjects' direct environment factors from interference in the experiments. Prepare written instructions for assistants and all subjects.
Threats to hypothesis testing	The independent variable, cannot be manipulated due to time and money constraints on the project.	Some projects simply do not allow for testing by creating two or more different solutions to a design problem. Valuable usability results can still be gathered by documenting the design choices very carefully.

Table 9.1: Threats to validity and reliability of the measures in CVE experiments.

The validity and reliability threats are not insurmountable; even in the most restricted usability testing settings valuable data can be gathered by observing the designers and users very carefully, as shown by the "Inhabit the WEB" usability evaluations (section 9.2.3 and 9.2.4). The author continues to provide a discussion of the problems and solutions encountered during the usability evaluations. The issues raised in this discussion are summarized in table 9.2.

Evaluation Problems	Description	Solutions
Single researcher, many sites.	Assistant usability researchers have to be recruited at each site, and instructed.	Allow plenty of time to prepare assistants. Build extra time into the experiment schedule for each person to be in the right place.
Motivation of subjects to attend.	Only a small number of subjects can be used for each experiment (limited by the processing power of the CVE), and if one does not turn up the others are waiting in vain.	Use team-building techniques, and rewards. Plan arrival and departure of subjects carefully, and stick to timetable.
Motivation of subjects to return questionnaire.	Difficult to control and motivate subjects to answer questionnaire because subjects are distributed over many sites.	Make task very obvious. Make questionnaire as easy to answer as possible.
Explanation of task difficult to control.	Subjects are distracted by interface interaction during the first part of their arrival in the CVE. They tend to listen badly to instructions and do not remember their task clearly.	Provide the subjects with a written task description at each stage of the task explanation. Let local experimenter read task descriptions out loud.
Training and documentation for interface controls not available.	Training and manuals are often not available in the developmental stages of a software application.	Create your own manual by using web pages. Use short training before each experiment.

Table 9.2: Problems and solutions found whilst designing experiments within the constraints of CVE evaluation methodology.

The paper concludes by discussing the benefits and drawbacks of distributed usability testing. The direct benefit of the distributed setting for the evaluation results was that the observational data complemented the network measurements in that extreme numbers in the network measurements could be related back to the activities observed, and vice versa. The types of data found using observation and questionnaires were: user misunderstandings of interface, interface improvements suggested by users based on their experiences during the experiments, better understanding of human behaviour in CVEs, and last but not least a better understanding of the methodology of CVE evaluation. Amongst the drawbacks are: the fact that assistant usability evaluators are needed, which did not always prove easy, although in principle more data can be gathered and analysed if the assistants do their own analysis of the same experiment. Additionally, because only experts used the application this limited the generalisability of the findings; however the experts expressed highly informative and useful opinions. Finally, the author urges all CVE usability researchers to pool their data in order to get to testable hypotheses of human needs and behaviour in CVEs.

During the presentation of the poster at the UKVRSIG conference it became very obvious from the amount of attention and the type of questions asked, that the topic was deemed important and there was a need for more focus on these issues. The first contacts between the author and other researchers ultimately resulted in a VR usability workshop, described below (section 9.3.5.2).

9.2.2 Evaluation Results of Inhabit the Web Project

"Evaluating the network and usability characteristics of virtual reality conferencing" (Greenhalgh, Bullock, Tromp, and Benford, 1997), has been published in the BT Technology Journal. The paper presents the results of the BT/JISC (Joint Information Systems Committee of the Higher Education Funding Council for England), funded "Inhabiting the Web" project, which conducted a series of twenty virtual meetings using the MASSIVE virtual reality teleconferencing system between BT Laboratories and five universities (Nottingham, Lancaster, Manchester, Leeds, and UCL). The aims of the project were firstly, to assess the characteristics of the network traffic generated by MASSIVE, and secondly, to identify key human factors arising from the long-term and regular use of the system. The paper gives an overview of the MASSIVE system, the organisation of the virtual meetings, and the results of the network and usability data analysis. A description of the network data analysis goes beyond the scope of this chapter and this thesis (the interested reader is referred to Chris Greenhalgh's PhD thesis (1998), but, in summary, the key findings were that on average users move about 20% of the time and speak (i.e. send network audio data) about 25% of the time. When users are involved in a common task, a small, statistically significant correlation can be found between their activities. Group transitions from one MASSIVE space to another, was a common occurrence, suggesting that group transportation mechanisms can be used to reduce network traffic caused by this activity. Audio traffic could be reduced by using more aggressive silence detection mechanisms, although the authors note that by implementing this solution there is a risk of missing potentially significant non-speech and background noises. A replicated time-series design was followed for the network traffic data collection, as well as for the usability data collection. The repeated collection of usability data involved constructs such as: satisfaction, experience, ease of use, group involvement, and awareness of other users and objects. In addition each session was used to explore newly identified constructs such as dealing with multiple embodiments, distributed awareness and switching between the virtual environment and the real environment.

Answers to over one hundred questions were collected from six subjects during the 20 experimental sessions. The questionnaires were handled by web-forms and consisted of attitude statements with Likert-scales, and open-ended questions. For example, after an experiment aimed at learning a new interface feature called “Focus”, which allows the users to change the degree of their awareness of each other and their environment, typical questions would be: “How easy or difficult do you find it to use the focus command? Please score on a scale of 1 to 7, if 1 = very difficult, and 7 = very easy”, and “Please explain how you would change the commands for the focus feature to improve its ease of use.” Interviews were used as a follow-up for some of the more interesting observations. The paper continues by presenting three key issues that emerged from the usability evaluations: user embodiments, navigation, and the use of the spatial model of interaction around which MASSIVE was designed. The findings that relate to user embodiment issues largely indicate a lack in the degree of control of the virtual embodiments, a lack of expression of non-verbal interaction elements, and a lack of natural mappings of body movements and expressions to the virtual body controls. In terms of navigation the main problems found were: difficulties with fine-grained movements, such as aligning to other people and objects. Finally, problems with the spatial model of awareness, which can be both system controlled and user controlled, largely related to the invisibility of the effects of the mechanism, including the lack of feedback when system controlled changes in the settings occurred. Proposed solutions to these problems were to automate actions such as gestures, alignment to objects, tracking of moving objects, and making all system actions sufficiently visible.

The author compiled a list of topics based on a content analysis of all questionnaire results, which has not been published before. The result of this analysis is presented in table 9.3. The author selected what seemed to be the most pressing topic, “automation”, for a further paper, which is reviewed below (section 9.2.3).

General topics	Sub topics
Awareness	Spatial presence
	Spatial location
	Spatial navigation
	Self
	Other speaker
	Other group
	Objects
	Immersion/Presence
Automatic behaviors	Objects
	Recurrent activities
	Gestures and expressions
Spatial Lay-out	Adjacency and neighborhood
	Objects (exits)
	Navigation
	World structure
Community	Development
	Exclusion/Inclusion
	Social distance
	Group building
	Audio
	Views
Audio	
Methodology of evaluations	Coordinator of networked trials
	Usability evaluator
	Interviewer

Table 9.3: Usability topics derived from content analysis of all ITW questionnaire results.

9.2.3 Virtual Body Language

“Virtual body language: providing appropriate user interfaces in collaborative virtual environments” (Tromp and Snowdon, 1997), has been published in the proceedings of the Virtual Reality Systems and Technology (VRST) conference. The paper describes a set of requirements for user embodiments and user interfaces for CVEs. The proposed requirements are based on a review of VR literature and the longitudinal study of MASSIVE (the ITW study described in section 9.2.2). The authors propose an automated virtual body language tool based on a general software architecture that might be used to manage the selection of interface components guided by the users' current context. The user interface components are selected from a user modifiable database, allowing the user to create new user interface components as required. The authors argue that the design of CVEs, objects and embodiments should be considered in terms of how they afford social interaction. The embodiment should not simply be thought of as an interface device for navigation, but in terms of how it can support higher order movements of social significance, such as approaches, turnings, glances, etc. In the paper the claim is made that the central usability finding from the ITW studies was that the embodiment does not provide enough control. The six respondents to the questionnaires agreed that with the embodiment they had available, 70% or more of the interaction activities depend on voice only, and 30% or less on the embodiment. A quote from one of the respondents illustrates this claim:

> *"Most of the time I do not use my body to interact with other participants. [...] Usually it's all vocal interaction with others – we form a circle to hold a discussion, but we don't really use the fact that we can all see each other to govern what happens. Most of the cues are aural. Perhaps it's because the ways of making the body do things are not very natural???".*

The commands to control the virtual body are listed in Table 8.4 below. It may be obvious from this how few commands the users had available to express any semblance of body language.

Command	Explanation
0-9	Change of viewpoint
r	Reset orientation
R	Reset orientation and go to ground plane.
C	Go to centre of world
P	Change your focus setting by toggling it (default is wide).
<	Zoom focus in.
>	Zoom focus out.
B	Blush
S	Sleep
T	Point
H	Both arms in front of you.
K	Both arms in the air.
J	Left arm in the air.
L	Right arm in the air.
Space bar	Return body to normal.

Table 9.4: The key control commands for the virtual embodiment in MASSIVE-1.

Although the users tried to use the commands that they did have available in creative ways, this was not perceived as sufficient to support their social interaction. Another quote from a respondent may illustrate the scope of this usability issue:

> *"[To get attention in a meeting I use] animated arm waving. Attention may be attained through movement and as such I resort to shift-l / space in order to attract attention. It would be nice to automate the process of waving."*

This and other problems all pointed to the need to introduce automatic embodiment behaviours, in order to come to higher level interaction abilities. The suggested improvements in the paper are listed in table 9.5 below.

Types of automated behaviours	**Examples**
Lock onto an object in order to follow it automatically.	For instance, automated tracking of gaze direction to current speaker.
Behaviour macros.	For instance, gestures for waving, nodding, running, walking and facial expressions, and also more higher-order sets of actions, such as the actions a chairperson would want to use to open a meeting.
Objects should react to the user, instead of the user to the object in order to interact with it.	For instance, having the object's functions triggered and made available by approaching or selecting the object.
Buttons for often used behaviours.	For instance, having an 'applaud' button which produces the sound of hand clapping, the more users press the button at the same time, the louder the applause.
Transparency of control.	For instance, it should be obvious who is the current speaker, or who is currently manipulating an object, and a user should be able to go against the automatic behaviour of an object when possible.

Table 9.5: Suggested improvements to CVE interface derived from the ITW study.

It is beyond the scope of this thesis to discuss the proposed software architecture that the authors suggest could take care of automatic common behaviours, so the interested reader is referred to the original paper for more details. However, it would seem important to add that the authors envisioned this tool to be user controllable at any time, user extendable, and sharable, so that users can inherit behaviours from each other. This creative editing and inheriting of behaviour macros, could contribute to a naturally evolving CVE on a social and functional level. The authors state that recording the changes and additions made to the macros by the users, could provide an insight in the perceived needs of the users, which could ultimately contribute to evaluation of the CVEs.

9.3 COVEN Papers and Deliverables

This section reviews the work the author performed during the COVEN project. The first subsection (9.3.1) reviews COVEN project dissemination publications, the second subsection (9.3.2) reviews the Inspection Method developed for CVEs, which was gradually refined during the three iterations of the usability activities of the COVEN project. The third subsection (9.3.3) reviews the three papers that presented the results of the COVEN usability evaluation activities to the scientific research community. The fourth subsection (9.3.4) reviews the design documents that were produced during the different stages of the iterative design and summarise the experience gathered from this project on the issue of design of CVE. The fifth subsection (9.3.5) reviews the international discussion that was instigated to address open issues regarding CVE usability design and evaluation at the time, including a summary which has not been published before.

9.3.1 COVEN Project Dissemination

The members of the COVEN project were contractually committed to disseminate the existence and results to the academic and industrial communities by means of papers, presentations and demonstrations. The papers presented in this section are part of a large body of papers produced by the COVEN project members (see COVEN website for further details http://www.coven.lancs.ac.uk), in the production of which the author of this thesis was directly involved.

9.3.1.1 COVEN Project Dissemination I

"Collaborative Virtual Environments: the COVEN Project" (Normand and Tromp, 1996), was the first official presentation made to the academic and industrial research and development community. The presentation was made at the highly popular "Frameworks for Immersive Virtual Environments" (FIVE) conference in London by Veronique Normand, the manager of the project. The paper describes the project as a European project addressing technical and design-level requirements of VR-based multi-participant collaborative activities for professional and citizen oriented domains. It gives a short overview of the objectives and activities of the project, highlights the results at the time the paper was written and indicates the plans for future research. Normand wrote this paper and the present author made additions to it.

9.3.1.2 COVEN Project Dissemination II

"Collaborative Virtual Environments: the COVEN Project" (Tromp and Steed, 1998), was the first official publication made for the academic and industrial research community participating in a human-computer interface oriented conference; the British HCI conference. The paper describes the COVEN project, with the emphasis on the evaluation activities that took place throughout the project. It states that adaptation of standard HCI methods are needed in order to address the new interaction issues introduced by 3D interactive multi user VR environments. It presents the first results of the application of the standard HCI Inspection method and discusses how the usability issues uncovered by this method fall into three categories: system issues reflecting fundamental properties of the CVE system, interaction issues rising out of the complexity of interacting with 3D scenes, and application issues needed for the presentation of functionality inside the 3D world. It further describes how the results of explorations of human needs within CVEs support the findings from the Inspection, and mentions that further auxiliary case-controlled experiments took place as part of the project activities that focussed on the evaluation of central

CVE concepts such as the sense of presence and requirements for collaboration. Finally, future research is discussed including the development of the Inspection method to be more applicable to CVEs and the attempt to extract design guidelines for CVEs from the usability findings of the project. The paper has been largely written by the author of this thesis, with a review and additions by Steed, whose collaboration on the collection and analysis of the usability data was fundamental to the content of the paper.

9.3.1.3 COVEN Project Dissemination III

"The COVEN project: exploring applicative, technical and usage dimensions of collaborative virtual environments" (Normand, et.al., 1998), was published in the journal Presence, and presents the COVEN achievements after two years of work. In the paper a large collection of COVEN partners present the main features of the COVEN approach and results: the driving applications, the main components in the technical investigations, and the experimental activities. With different citizen and professional application scenarios as driving forces, COVEN presents the exploration of the requirements and supporting techniques for collaborative interaction in scalable CVEs. It explains how the technical results are being integrated in an enriched networked VR platform based on the dVS and DIVE systems. Furthermore, it is made clear how the project is taking advantage of a dedicated Europe-wide ISDN and ATM network infrastructure, explains that a large component of the project is a trial and experimentation activity that should allow the build up of a comprehensive understanding both of the technical network requirements of such systems and of their usability issues and human factors aspects.

The author contributed sections on experimental work in the project, assessing the usability and human factors aspects, and social interaction research. The main points that the author of this thesis made in these sections are summarized here:

1) The evaluation of the COVEN Platform is influenced by three considerations: usability engineering, scientific inquiry, and development of a methodology for CVE-specific usability studies.
2) Because of the methodological constraints involved in CVE prototype evaluation, identified above, a qualitative research approach was chosen. Because of the relatively small number of available subjects, repeated measures were made so that a large number of observations are obtained from each person.
3) Under focus were the COVEN services, which were identified at the beginning of the project, and were implemented in different degrees of detail in the initial application under evaluation.

These COVEN services were: support for mutual awareness, with possibly varying degrees; support for communication between participants; support for resource management, support for participants roles and rights, support for object manipulation, support for group navigation within the environment, support for human actors, support for a global map; and support for subjective views of the environment, possibly in relationship to the notion of participant roles. The other main sections of the document consist of expositions such as collaborative interaction from a computational point of view, scalability enabling techniques, and a summary of the achievements and perspectives on further work during the project.

9.3.1.4 COVEN Project Dissemination IV

Towards the end of the COVEN project all partners were invited to write an article on the results from the respective work packages they had delivered for the project. The present author wrote an article covering all usability activities that took place during the project, titled "Systematic Usability Evaluation and Design Issues for Collaborative Virtual Environments", Steed and Wilson reviewed it and provided additions (Tromp, Steed, Wilson, in press). The paper presents the results from the longitudinal usability studies within a framework of various complementary methods (i.e. usability engineering methods, observational methods, and empirical experiments). The need for such a framework is argued and the results from the evaluation activities are used to extrapolate further research directions for CVE development. These topics are summarized in table 9.6 below.

Research Method	**Research Findings**	**Solutions**
Inspection	Lag causes different delays for different users creating consistency problems.	Research area
Inspection	The 2D input device does not match the 3D task space.	Research area
Inspection	The perceptual affordances of objects are not obvious.	COVEN Design Method (Del2.9)
Inspection	The sequential affordances of object interactions are not obvious.	COVEN Design Method (Del2.9)
Inspection	The narrative affordances of the task space are not obvious.	COVEN Design Method (Del2.9)
Observation	Scanning occurs during the total meeting, and has the highest frequency during the middle part of the collaboration.	Automation
Observation	Communication acts during collaboration are supported by more than twice the amount of meta-collaboration acts.	Automation

Observation	Navigation often involves many fine-tuned positioning acts in order to encompass the most advantageous viewpoint for collaboration.	Automation
Consumer Evaluation	Commercialising CVEs for virtual travel information involves the need to develop new interaction paradigms both for the information providers and the customers.	Commercial development area

Tab. 9.6: COVEN Project usability research results summarized.

9.3.2 Inspection Method Developed for CVEs

Inspection methods are traditionally presented as particularly well suited to perform a quick and 'cheap' clean-up of a design, as described in (Nielsen and Mack, 1994). COVEN proposed to apply both heuristic evaluation (Nielsen, ibid) and cognitive walkthrough (Wharton et al., 1992) approaches, each approach focusing on different usability aspects of the design: heuristic evaluation focuses on the general usability of a user interface and uses design guidelines, while cognitive walkthroughs mainly address ease of learning and in a generally 'freer' format. Drawing on the HCI expertise of the consortium, usability inspections were thoroughly performed during the three evaluation phases of the project, involving the usability experts from the COVEN partner institutions. The method and a summary of its findings are described below.

The COVEN activity to test and develop the traditional HCI "Inspection Method" (Nielsen et al, ibid), was based on the premise that there were neither Cognitive Walkthroughs (CWs), nor Heuristic Evaluations (HEs) specifically for the Inspection of CVE interfaces. The COVEN Inspection method went through three iterations of development. The first iteration was a direct application of the traditional 2D

Inspection method, with additions to cover the collaborative aspects of CVEs. This was useful, but not sufficient to cover all aspects of 3D applications (COVEN Del3.3). The second iteration was a mix of an early version of the pioneering development of an Inspection method for single user VEs (Sutcliffe & Kaur, 1998), and the 2D method, again with additions to cover the collaborative aspects of CVEs, based on a model created by the author of this thesis to clarify the collaboration task cycle (see figure 9.1).

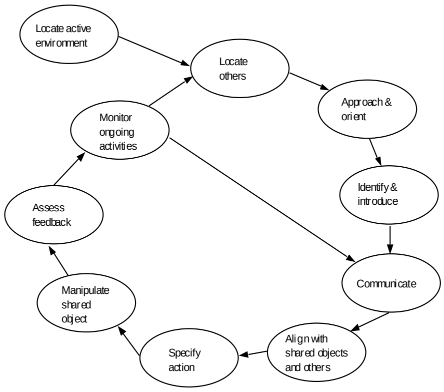

Figure 9.1: Model of CVE Collaboration.

This was also useful, but it was found that the total CVE experience needed to be inspected specifically for continuous task-flow. The total Inspection method was updated into a more coherent method description (COVEN Del3.5). The third iteration of the Inspection method was applied to three different CVEs, by six Inspectors (two single Inspectors looking at two distinctly different CVE applications, and two teams of two Inspectors looking at two distinctly different applications). Additionally, three CVE designers were asked to use the Inspection method and give feedback about the effectiveness of the method. In particular, we wanted to know whether it gave them more insight into usability design issues. Typical reactions we received were: “Particularly useful was the way the method emphasizes that users should get feedback on their actions and should always know what to do next." Traditionally, hierarchical task analyses are used to describe the tasks in a 2D interface, and one generic 2D task cycle for inspection guidance. For 3D interfaces we proposed using a floor plan to analyze task-flow. This should help to find the ideal positioning of interactive objects, the ideal path through the 3D space, and the ideal perceptual message we want to convey to the users. Additionally, we use multiple generic task cycles, which have been created specifically to guide the inspection of 2D, 3D, navigation, collaboration, and system controlled tasks. These task cycles have a strong emphasis on supporting continuation of the task flow for self governed, free moving users in an unconstrained 3D task space. We isolated three pertinent novel areas of VE design which needed special attention:

- Freedom of navigation in a pseudo 3D space.
- Interactive 3D objects.
- Human-human collaboration as mediated by the CVE.

The COVEN inspection method helps CVE developers accomplished the following main goals:

- Check the design of a CVE at the level of objects, interactions and task-flow.
- Establish how well task-design and task-flow meets representative user needs.
- Recommend redesign solutions for specific areas of potentially serious user interface struggles.

The representation of the environments has a strong bearing on how user tasks will be carried out. By going through the actions which a user will be expected to perform in the CVE, and noting the things that are not obvious or difficult, one is inspecting the design. Additionally, this method encourages positive feedback like the formulation of redesign suggestions, and should give a clear impression of recurring weak design areas.

The COVEN Inspection method has been developed by the present author, with help from Anthony Steed (UCL, UK), Veronique Normand (Thomson, Fr), Judith Dijkhuis (KPN, NL) Anne-Marie Sandos (Nottingham, UK), Stefan Thie (KPN, NL), Kulwinder Kaur, (CCL, UK), Alistair Sutcliffe (CCL, UK), Dave Lloyd (Nottingham, UK), Boriana Koleva (Nottingham, UK), Sophie Drijver (KPN, NL), Gidi van Liempd (KPN, NL) and Eugenio van Mierlo (TNO, NL).

Large parts of the COVEN Inspection document are based on other people's work. To be precise: The COVEN Inspection method has been built on top of the Inspection method for single-user VEs by Sutcliffe and Kaur (1998). The examples in this document are adapted from "The Inspection of the London Demonstrator", by Steed (COVEN Del3.7). The User Context Analysis is based on parts of the User Requirements Framework Handbook, Deliverable 5.1 of the RESPECT Project (RESPECT, 1997).

9.3.2.1 COVEN Inspection Method

The traditional HCI Inspection method consists of two parts: a Heuristic Evaluation technique, and a Cognitive Walkthrough technique. In first instance both were used to evaluate the COVEN platform, however in the second and third iteration of the COVEN usability evaluations only the Cognitive Walkthrough was used, albeit a version adapted for CVE interfaces, and this is referred to as “the COVEN Inspection method”.

9.3.2.1.1 Heuristic Evaluation

For the first Inspection Nielsen’s ten Usability Heuristics were used (from: http://www.useit.com/papers/heuristic/heuristic_list.html), these are listed in the table below (9.7).

Standard Heuristics	**From: http://www.useit.com/papers/heuristic/heuristic_list.html**
H1: Visibility of system status	The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.
H2: Match between system and the real world	The system should speak the users' language, with words, phrases and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order.
H3: User control and freedom	Users often choose system functions by mistake and will need a clearly marked "emergency exit" to leave the unwanted state without having to go through an extended dialogue. Support undo and redo.
H4: Consistency and standards	Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions.
H5: Error prevention	Even better than good error messages is a careful design which prevents a problem from occurring in the first place.
H6: Recognition rather than recall	Make objects, actions, and options visible. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate.
H7: Flexibility and efficiency of use	Accelerators -- unseen by the novice user -- may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions.
H8: Aesthetic and minimalist design	Dialogues should not contain information which is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.
H9: Help users recognize, diagnose, and recover from errors	Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.
H10: Help and	Even though it is better if the system can be used without documentation, it may

documentation	be necessary to provide help and documentation. Any such information should be easy to search, focused on the user's task, list concrete steps to be carried out, and not be too large.

Table 9.7: Nielsen's Ten Usability Heuristics.

The Heuristic Evaluation makes use of heuristics for evaluating 2D applications. Quoting Nielsen from his web-page:

> *"I originally developed the heuristics for heuristic evaluation in collaboration with Rolf Molich in 1990 [...]. I since refined the heuristics based on a factor analysis of 249 usability problems [...] to derive a set of heuristics with maximum explanatory power, resulting in this revised set of heuristics [...]."*

Adjustments were proposed to some of the definitions above, which were definitely specific to 2D graphical interfaces (especially 'Aesthetic and minimalist design') or to full-scale products ('Flexibility and efficiency of use' and 'Help and documentation'). The following adjustments were suggested:

H1: visibility of system status: In addition to visibility of system status**,** we may add '**awareness of other participants**' as a simple general CSCW application usability principle, allowing capture of the most obvious design deficiencies with regards to collaboration within the VE. Our assumption was at that stage of the evaluation, such a general-level principle may be enough.

H3: User control and freedom: Nielsen's concern is to avoid users being trapped in dialogue modes, as can happen in some 2D interfaces. An interpretation in our context relates to the freedom and flexibility in navigating the VE.

H7: Flexibility and efficiency of use: focus should be on efficiency of use rather than flexibility (no need for tailorable interaction, accelerators definition at this stage of the prototype).

H8: Aesthetic and minimalist design: aesthetic design certainly is relevant here; minimalist design makes sense for 2D dialogue sequences (especially dialogue panels) but may be irrelevant in the CVE context outside of such dialogue sequences, if any. More precisely, minimalist design may be relevant in some professional usage applications, but when strictly applied may contradict the premises of information retrieval CVEs (such as the Citizen application), which call for opportunistic exploration of the data base by the users. For these applications, the design should be clear and structured so that users can understand the different options offered to them and don't get overwhelmed by the amount of options, nor get lost in the VE.

H10: Help and documentation is not a relevant heuristic here, at this stage of the prototype.

These heuristics were used during the first inspection of the COVEN platform and uncovered some issues that were looked into in the further iterations of the development of the COVEN Inspection method. Based on the first inspection results, more amendments to the Heuristics were suggested, which take into account the 3D, and collaborative aspects of CVEs. Below is a listing of the new heuristics, followed by argumentation of what issues should be taken into account for CVE evaluation.. Rather than introducing more heuristics, an effort was been made to include general 3D and collaboration design principles in the existing heuristics.

H1 (new)	***Visibility of system status*** The system should always keep users informed about i) what is going on, ii) where it is going on, iii) what interaction mode they are in, iv) other users and their status, v) delay is scene transition, through appropriate feedback within reasonable time.

Since viewpoint in a CVE is unconstrained, this heuristic has to be updated. The environment is made of objects, with or without associated functionality. A user may or may not be paying attention to an object for which the system status is changing. Additionally a user may have shifted his/her attention away from an object for which a system change is occurring.

Users can select different modes of interaction within the CVE, such as navigation modes: walking, flying, and teleportation, object interactions: selection and manipulation. The user needs to be made aware of the different states of interaction, of the current state, and of changes in the current state. Users may want to be informed by the system about the number of other users present in the CVE, who is leaving or arriving, and possibly where they are and who they are. Lastly, scene transitions are accompanied by a time delay. In order to inform the user that there is nothing to worry about, and possibly to give an indication of the expected delay, a convention has to be developed, which can be used every time a delay is expected.

H2 (new)	***Match between system and the real world*** The system should speak the users' language, with words, phrases and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions where possible, making information appear in a natural and logical order.

It is not always possible to find real world concepts to design CVE issues, an example is the ability to teleport from one location to another location in a CVE. In this case metaphors must be found to convey the function of an object or action by using an analogy from the real world. Where possible literal metaphors should be used, but when this is not possible "magical" facilities can be used, such as a flying carpet to signify a bird's-eye view. Care must be taken that the information still appears in a natural and logical order. It is also important to make sure that this metaphor does not call forth any extra conceptual knowledge that does not fit the CVE phenomenon for which the metaphor is used, because this will confuse the user. The old heuristic is qualified with 'where possible' when addressing the issue of following real world conventions.

H2 - part B (new)	***Match between metaphor and the real world counterpart*** The metaphor should be based on real-world counterparts where possible, so that users can employ their conceptual knowledge of the real world to predict what will happen in the CVE, metaphors should not overlap. Magical facilities should appear natural and logical within the frame of reference of the CVE.

The heuristic is supplemented with H2 - part B, to address the issue of metaphor to convey affordances for the CVE that are similar to the real world, but not the same. This heuristic is supplemented with the note on the introduction of magical features for something in the CVE that simply does not exist in the real world.

H3 (new)	***User control and freedom*** Support undo and redo. In particular, assess freedom and flexibility in navigation. Group navigation, subjective views, remote object manipulation need to be solved in a way that the other participants are not unnecessarily disturbed in their actions.

Users can navigate freely through the 3D space alone or in groups. Group navigation means that users belong to a group and follow the movements of the group leader. When moving alone the user may need to reorient back to their point of departure or a few ‘steps’ back. Similarly, the system may take control from the user in order to ease navigation. The user should be able to undo this control over their embodiment when desired. A solution for such undo/redo actions needs to be found. When moving in a group a user should be allowed, when possible, to leave the group at any point. The different levels at which undo/redo must be used have to be explored further.

Users are theoretically in a multi-user space. This has implications for undo/redo functions that affect the CVE. The issue of subjective views needs to be explored further in order to come to recommendations for implementations of this heuristic in CVEs. Similarly, because of the distributed nature of CVEs it may not be possible to have a global undo/redo function.

In addition, issues of manipulation on remote or distant objects needs to be explored. Users should not be restricted in terms of manipulation of objects that are not in their direct vicinity; one of the advantages of CVEs is that users can act on remote objects.

H4 (new)	***Consistency, standards and appropriate affordances*** Users should not have to wonder whether different words, VE objects, situations, or actions mean the same thing. Objects should suggest appropriate affordances to users, and affordances should be applied throughout the interface in a consistent way.

The topic of consistency and standards has been addressed in CVE design discussions as the 'affordances' of objects. Affordances are functional meanings conveyed from objects to the observer. Objects suggest possible actions to the observer, based on the conceptual knowledge of the observer. When designing new objects and functionality in an existing CVE the affordances should not only be natural, compared to the real world, but also consistent with similar actions in the existing CVE. Solutions have to be explored to make this easier.

H5 (new)	***Error prevention and User Guidance*** Even better than good error messages is a careful design which prevents a problem from occurring in the first place. The user needs to be guided from one expected action to the next, by drawing her attention to it; either by putting the next available action in her field of view or through the use of some other cue.

In principle, objects can be completely or partially obscured by other objects in a scene. It may not be apparent to a novice user, or any user who has never seen the scene before, which object the user is expected to interact with. Often it is also not clear, which objects a user can interact with and which ones the user cannot interact with, because they are mere decoration to make the scene more believable. This issue could be addressed in H5: Error prevention. This heuristic should be extended to include 'user guidance'. Because the user is free in navigating the environment, where the designer intends the user to take certain actions, the user should be guided in their choices. This can be achieved by putting the next expected action in the user's view or making it otherwise noticeable by the user.

H6 (new)	***Recognition rather than recall*** Make objects, actions, and options visible. The user should not have to remember information from one part of the dialogue to another. The user should at any time be able to request information about objects and actions.

Many actions which can be performed on objects are not obvious until the object gets closely inspected. This issue is really addressed in CW question B: Are the actions which are available to the user clear? However, this issue could also be included in a heuristic because it is such a typical problem for objects in CVEs. This issue could be added to H6: Recognition rather than recall. Functions should be made obvious to the user by making use of the correct choice of metaphor or realistic representation, calling forth the appropriate affordances. This heuristic complements H4: Consistency, standards and appropriate affordances.

H7 (new)	***Flexibility and Efficiency of use*** General: Users should not have to perform cumbersome navigations in order to operate on objects or functions of the interface. Prototype: Focus on efficiency of use rather than flexibility. No need for tailorable interaction, accelerators definition at this stage of the prototype. End-User Product: Accelerators -- unseen by the novice user -- may often speed up the interaction for the expert user such that the system can cater for both inexperienced and experienced users. Allow users to tailor frequent actions.

Many problems have been found with navigation and selection. In order to manipulate an object, or perform an action on an object, the user has to be able to navigate the embodiment and view quite accurately. This is certainly a problem for the novice user who has no knowledge of the correct moves yet, and it is simply cumbersome for the experienced user. This issue could be addressed in H7: Flexibility and efficiency of use. The ultimate solution to problems of this kind is to introduce a certain amount of automatic behaviour or accelerators: once a user reaches a certain proximity with an object the system automatically performs the actions for the user, unless the user interrupts the system. Users should be able to override automatic system activity at any time. This makes H7 the complement of H3: User control and freedom.

H8 (new)	***Aesthetic and clear/structured design*** Spaces should contain atmospheric elements to create places where users feel 'at home'. Information which is irrelevant or rarely needed should be avoided. Every extra unit of information in the space competes with the relevant units of information and diminishes their relative visibility. The design should be clear and structured so that users can understand the different options offered to them and not be overwhelmed by the amount of options. Landmarks should be provided in order to recover from getting lost in the VE.

In CVEs a minimalist design of the environment is not necessarily a better design. Users have been shown to use extra features in the environment as landmarks to help them orient themselves when navigating through the CVE. Also, users have been shown to feel more present in the CVE when it is rich in features, thus creating a realistic populated space. These factors are relevant for creating a sense of presence, one of the main goals of CVEs.

H9 (new)	***Online user help.*** *Help users recognize, diagnose, and recover from errors, by providing high quality error feedback, help and documentation inside the CVE.*

Documentation about functions of objects could be made available to the user inside the VE. The user could request information or help on the functions of objects by request, or objects could automatically produce information on their function when the user approaches them.

H10 (new)	***Awareness of own embodiment and that of other participants.*** General qualities of the embodiments in terms of uniqueness, visibility of the presence and activity of the other participants.

Users should be able to distinguish their own embodiments and those of other participants. They should be able to identify and recognise other participants easily. Users should also be able to identify who is performing actions on objects that affect their own embodiment or environment.

All evaluators agree that the Cognitive Walkthrough is the easier to perform, because it is more a step-by-step process, which follows the users' thoughts closely. For this reason the CW was adopted as the primary inspection method.

9.3.2.1.2 Cognitive Walkthrough

The questions used for the cognitive walkthrough were not changed prior to the first inspection, except in terms of wording. A copy of the original questions follows below (See http://www.cc.gatech.edu/computing/classes/cs3302/documents/cog.walk.html for the original source):

- *Will the users be trying to produce whatever effect the action has?*

Are the assumptions about what task the action is supporting correct given the user's experience and knowledge up to this point in the interaction?

- *Will users be able to notice that the correct action is available?*

Will users see the button or menu item, for example, that is how the next action is actually achieved by the system? This is not asking whether they will know that the button is the one they want. This is merely asking whether it is visible to them at the time when they will need to invoke it. An example of when this question gets a negative supporting story might be if a VCR remote control has a hidden panel of buttons that are not obvious to a new user.

- *Once users find the correct action at the interface, will they know that it is the right one for the effect they are trying to produce?*

This complements the previous question. It is one thing for a button or menu item to be visible, but will the user's know that it is the one they are looking for to complete their task?

- *After the action is taken, will users understand the feedback they get?*

Assuming the users did the correct action, will they know that. This is the completion of the execution/evaluation interaction cycle. In order to determine if they have accomplished their goal, the user needs appropriate feedback.

Severity ratings

The severity of a usability problem is a combination of three factors (from http://www.useit.com/papers/heuristic/):

- The frequency with which the problem occurs: Is it common or rare?
- The impact of the problem, if it occurs: Will it be easy or difficult for the users to overcome?
- The persistence of the problem: Is it a one-time problem that users can overcome once they know about it, or will users repeatedly be bothered by the problem?

The following 0 to 4 rating scale is one option for rating the severity of usability problems:

0. I don't agree that this is a usability problem at all.
1. Cosmetic problem only: need not be fixed unless extra time is available on project.
2. Minor usability problem: fixing this should be given low priority.
3. Major usability problem: important to fix, so should be given high priority.
4. Usability catastrophe: imperative to fix this before the system is tested by users.

The inspection during the Cognitive Walkthrough was based on the following four questions (A to D), see figure 9.2.

A. Will the users be trying to produce whatever effect the action has?

B. Will users be able to notice that the correct action is available?

C. Once users find the correct action at the interface, will they know that it is the right one for the effect they are trying to produce?

D. After the action is taken, will users understand the feedback they get?

Figure 9.2: The original four Inspection questions (A to D).

Based on the experiences with the first Inspection, it was proposed to slightly change each of them and also to insert an extra question (see figure 9.3 below). The justification for this is that these questions elicit a clearer description of the usability problems, because they are easier to apply. The advantages are that they are:

1. application non-specific (see problem classification)
2. unique to VEs because of control mapping/3D nature of display/nature of tasks
3. better categories for the problems found.

A: Is the user goal clear for the user?

B: Are the actions which are available to the user clear?

C: Is it clear which action is the correct/needed one?

D: Will the users manage to take the action, is it easy/immediate to perform the action?

E: Is there clear feedback after choosing the action?

Figure 9.3: The new five Inspection questions (A to E).

Because of the freedom of movement through the 3D environment users need to be guided in special ways through the 3D space, to go from one expected action to the next, if that is what the designers want. It may therefore be beneficial to add a sixth question to the list of CW questions: Is it made clear to the user what the next correct/needed action could be? (See figure 9.4.)

A: Is the user goal clear for the user?

B: Are the actions that are available to the user, clear?

C: Is it clear which action is the correct/needed one?

D: Will the users manage to take the action, is it easy/immediate to perform the action?

E: Is there clear feedback after choosing the action?

F: Is it made clear to the user what the next correct/needed action could be?

Fig. 9.4: The final six Inspection questions (A to F).

9.3.2.1.3 Feedback from Inspectors

The Inspectors were asked about their experiences with using the Inspection method, after each stage of its development. In general the Inspectors were quite impressed by the number of usability problems that were uncovered by using this method. Although the activity is time-consuming, it allows for usability testing without the use of end-users. Problems noted by the Inspectors were:

- **Performing the CW and HE at the same time creates a cognitive overload for the Inspector.**

It was suggested by the inspectors to use the CW as the first and main inspection method, and to use the HE as an additional method for refinement of the issues found by applying the CW.

- **Writing down the findings , first on paper by hand and then typing them into the forms is too time consuming and rather clumsy.**

It was suggested by the inspectors to use a Web based form system in order to reduce time and effort of filling out the firms.

- **Low-level bugs make complete branches in the task tree unavailable for Inspection.**

It was suggested by the inspectors to apply three levels of inspection. First an inspection by the designers themselves, next an inspection by the usability inspectors (the developers of the Inspection method), and finally (when the application has reached a proper finegrained level of usability problems), an inspection by actual end-users.

- **Collaboration aspects of the CVE need to be tested with 1 inspector and 2 helpers, which is more time-consuming, complex to set up, and will require not just 1 trained inspector but trained teams.**

It is suggested to isolate collaboration scenarios from the final CW scenarios, so that the two helpers need not be present throughout the total CW/HE session.

- **Different interfaces will be apparent on each machine (at the desktop level and scene level). E.g. group leaders and group followers have different capabilities.**

By introducing helpers who are also inspectors for the collaborative activities within the application, differences in interface should become clear, especially when those helpers also do a CW/HE of the collaborative action that is being inspected.

- **It is not always clear how many actions one can perform on an object, so that there is a large number of branches on the task-tree, which will only be found during the inspection.**

It was suggested by the inspectors to let one inspector inspect the application first, based on the first scenario description. After that the scenario steps can be extended and consolidated for the other inspectors. A mechanism should be provided for the other inspectors to add scenario steps that the first inspector has not introduced. A number of necessary refinements were specified:

- Flow of actions part of Inspection.
- Steps of cycles synchronised as much as possible.
- Formulation of steps of cycles more precise.
- Allocation of cycles to functions, tasks and objects made easier.

These refinements were added to the final version of the COVEN Inspection method. This final version of the COVEN Inspection method is presented in Appendix G of this thesis.

A final note that the developers of the COVEN Inspection method added to their observations of the usefulness of the Inspection method for CVEs is that the interaction cycles provide excellent guidance for the design of actions and objects, while they are being designed. It is for this reason that designers are encouraged to use the interaction cycles as a first form of guidance, when they are (re-)designing the application. The use of these cycles during the design task is further explained in Chapter 10, section 10.4.

9.3.3 Evaluation Results

During the longitudinal network trials of the COVEN platform, experiments took place on a fortnightly and sometimes on a weekly basis. Although it could never be predicted with a hundred percent certainty that the application and the network would be stable enough on that day to go through with the experiment planned, we still managed to run quite a number of interesting tasks. The author of this thesis encouraged other partners in the project who were taking part in the network trials to employ their own local usability researchers who could be assistants in running the experiments. This was important for two main reasons, one being that experience with running network trials had shown that assistants at each site involved in a networked experiment were essential, the other being that the European Commission insists on collaboration between project partners with the intention of spreading the know-how involved in the work across Europe. The author also stimulated the formation of a tight group of usability researchers within the project, calling for usability planning meetings annexed to COVEN plenary meetings, and ultimately organizing stand-alone usability coordination meetings between the partners involved in the usability work package. This resulted in a strong triangular cooperation between three partner institutions that lasted the duration of the project. She also called for researchers to use the open slots in the network trial planning, to be used specifically for focussed experiments that could be turned into conference and journal papers in relatively short amounts of time; targeting specific conferences. This resulted in a number of neat experiments, some of which the present author was directly involved in, and some of which became successful publications.

	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.	Orienteering	Mental model, group navigation, conceptual knowledge, information visualisation

Table 9.8: List of experiments prepared for the COVEN network trials.

During the first iteration of the COVEN usability studies two broad exploratory tasks (experiment 1 and 2 in table 9.8) 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 exploration of the human factors issues associated with these concepts. These experiments took place during this second phase of the COVEN usability and network trials, and were aimed at providing more in depth understanding of the CVE concepts. Experiment 3 in table 9.8 is the experiment reported in sections 8.3.3.2 and 8.3.3.3. Experiment 5 in the table is the main experiment reported in Chapter 8 of this thesis.

9.3.3.1 Evaluation Results I

"Experiences with the evaluation of CVE applications" (Steed & Tromp, 1998), has been published in the proceedings of the second Collaborative Virtual Environments (CVE) conference, and was jointly presented at the conference by both authors. Steed presented the technical part of the presentation, and the author of this thesis presented the usability part of the presentation, for which she prepared a six-minute video and handout that attempted to illustrate the type of usability breakdown found during the network trials. See transcript below for the transcript that accompanied the video demonstration.

1: Illustration of Getting Connections into Place

Usability issues: In the top right hand corner of the screen we can see Adrian behind his desk, interacting with the VE and the other subjects inside the VE, thus providing us with a record of user action with and within the VE. A lot of time and effort from a lot of people is involved in getting everything to work, people need to be there, but may have nothing to do for unpredictable periods of time. No collision detection in the floor. Poor audio quality at times, makes it difficult to follow what is being said. All embodiments look the same, makes it difficult to know who is saying what and who is where.

Subjects: Subject 1 (Nottingham, UK), Subject 2 (Bristol, UK) , Subject 3 (Groningen, NL), Subject 4 (London, UK).

Setting: We are looking from Subject 1's point of view into the VE. Subject 3 is having problems with his connection, the other subjects are waiting around, occupying themselves with the embodiment controls; sinking through the floor and making themselves invisible.

Subject 3: [Subject 1] I suggest I try a reboot, because I tried to remove things from /tmp and it doesn't work.

Subject 1: Okay, that sounds sensible [...] that will hopefully get everything on a clean slate […] if that doesn't affect you too much.

Subject 4: We [...] we're restarting the server whilst we're waiting, just in case it […].

Subject 1: Yeah, probably is a good idea.

2: Brussels Demo of COVEN Platform at the EU Headquarters

Usability issues: Navigation through doors is troublesome. Each user has to individually go through the door and deselect the doorknob afterwards, before the next user is allowed to go through the door. Also, problems with latency, selection and manipulation can be observed, culminating in cognitive overload where is becomes hard to speak and manipulate at the same time.

Subjects: Subject 1 (Nottingham, UK), Subject 2 (Bristol, UK), Subject 3 (Brussels, Belgium), Subject 4 (London, UK)

Setting: The subjects are working from a previously agreed script, in order to show the Commission the state of the art of the COVEN Platform during an actual network trial. All subjects are assembled in the first room of the business application. A collaborative stock market game can be seen in this room. Kurt is presenting the show to the public in Brussels. The second room is a presentation room for teleconferencing, with an overhead screen, a red bin which is the slide projector, and the slides, tables and chairs, etc.

Subject 3: Uhm, […] you're all very welcome to be here, thanks for making it on such short notice. If you could follow me through to the other room please.

Subject 4: So, ehm, thanks for inviting us [Subject 3].

Subject 1: Cheers!

Subject 3: Ah, as you all know, we're giving a few presentations here, so I would like to give the word to [Subject 1] who is giving the first presentation.

Subject 1: Okay, thanks very much [Subject 3]. I just have to let me come out to the front here. Eh, right let me just select this <slide> here. [...] So, here we just to first of all introduce ourselves and where we are we have a map showing the location of the various partners. Uhm.. So, I'm [Subject 1], and I am currently in Nottingham. Do you guys like to introduce yourselves?

Subject 4: [.......................] this yeah! I am [Subject [Fo]ur], and I am at the University College London. [...]

Subject 2: Hi, my name is [Subject 2], and I am here at Division Limited, Bristol.

Subject 3: And of course I am [Subject 3] here in Brussel.

Subject 1: Okay, now let me prepare the first talk.

Subject 3: Okay [Subject 1] please begin.

Subject 1: Okay, I'd like to talk to you today to give you a brief overview of the COVEN Project, which is concerned with collaborative virtual environments, and this is a European platform for cooperative teleworking. (Subject 1 is trying to talk, position his embodiment, change slides and switch views between the audience and the overhead screen at the same time!).

3: Citizen Application: the world disappears!

Usability issues: The subjects had to develop a "door-policy" where only one could go through at the same time, but not everybody sticks to this rule. Web-based instructions are looked up to learn about the experimental task. Causes of problems are deduced 'on the fly', by trying different combinations and actions of users per site. An apparently completely unrelated textual conversation between subjects is going on near the bottom on the VE screen. Walls can be selected, although no actions can be performed on them.

Subjects: Subject 1 (Nottingham, UK), Subject 2 (Bristol, UK), Subject 3 (Groningen, NL), Subject 4 (London, UK).

Subject 1: Okay, do we have to go through the door one at a time?

Subject 4 (goes through door) (Door closes).

Subject 2 and Subject 3 (go through door at the same time) (Door closes).

Subject 1 (opens door) (door opens)

Subject 3: [...] when I question who is in this zone here, I miss [..] eh [..] me!

Subject 1: Yeah I think you sneaked through the door when it was still open for [Subject 2].

Subject 3: Yeah, but me and [Subject 2] are in the same zone.

Subject 2 (leaves room).

Somebody (selects wall by accident) (wall high-lighted).

The paper presents the experiences of the authors of the usability evaluation of the COVEN platform. At the time this concerned two separate applications: one a teleconferencing system aimed at future business users, and the other a holiday planning environment for the use of the general public. The successful functionality of both applications was to be integrated into the final COVEN platform at a later stage of the project. The authors state that their concern in the evaluation was three fold: to refine the design of the applications, to gain further understanding of the component technologies of CVEs, and to reflect upon the methodologies for evaluation. The paper reports on two main threads of investigation: the usability inspection of the applications, and that part of the network trials that had taken place at the time. The paper presents the technical and usability basis of the project and project members involved, the framework used to plan the usability evaluations, an overview of the applications under evaluation, it also presents outstanding issues for the system designer, some guidelines for the application builder, and discusses the effectiveness of the methodologies chosen for the benefit of the usability engineers, including a focus for future evaluations for CVEs. The outstanding issues for system design are presented in table 9.9.

System design issues for CVEs	**Description**
System Issues	Latency in the propagation of updates of changes in the CVE to all users is a problem inherent in all VR toolkits. It makes interaction and navigation difficult at times, and causes asynchronous visual audio and collision behavior. Lack of general Undo for 3D environments. Lack of rules to allow users to lock objects during interaction. Lack of spatialisation of audio.
Interaction Issues	Multiplicity of layers at which interaction occurs: keyboard input, 2D widget interaction, continuous mouse driven 3D control, and discreet 3D

	widget interaction. It creates a considerable burden on the user to decide which level of input is required, a fault found with most VE desktop interaction systems that support free navigation.
Application Issues	Inconsistency of object reactions. Affordances of objects not obvious. Metaphors suggesting wrong or broader affordances than available. Lack of help with the VE itself.

Table 9.9: Outstanding issues for CVE system design.

The paper was jointly written by Steed and the author of this thesis. The video was created by the author with some technical help. The video footage was recorded during several COVEN network trials, and the demonstration sections were selected by the present author.

9.3.3.2 Evaluation Results II

"Small Group Behavior Experiments in the Coven Project" (Tromp, Bullock, Steed, Sadagic, Slater, Frecon, 1998), has been published in the IEEE Computer Graphics and Applications journal. The experiment was the third incremental follow-up of two previous experiments (Slater, et.al., 1998), this time using the COVEN platform as test bed. The experiments were designed to explore what happens when people who have never met before, meet for the first time in a CVE. The authors explain that this is an important topic for the future use of the technology because, or instance, if they distort "normal" social relationships, independently of the technological issues, CVE technology will not be able to become widely accepted as a serious communication means. The article describes the COVEN project, the COVEN platform, the COVEN network and usability trials, and the services the system provides in order to allow several people at remote sites to meet together in the CVE. The analysis of the experimental data focused on group behaviour issues, 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. There were 20 groups of 3 participants in the study. The task was a highly collaborative puzzle-solving task, and each participant in the group of three had a different style of embodiment, ranging from very realistic, to a very simplistic cartoon-style. The results of the experiment highlight issues to do with participant representation, interaction style and system implementation that need to be addressed in future CVE development. For instance, analysis of the data suggested that leadership within the group was conferred by the use of an immersive (HMD) interface to the CVE. This might mean that the person in the group with the most sophisticated equipment has an immediate advantage over others with less advanced equipment at their disposal. Another finding was that presence was experienced to a greater degree than co-presence, meaning that subjects had more difficulty feeling that they shared the CVE with real other people than they had feeling present in the CVE themselves. Finally, in terms of group interaction, indications were found that different levels of sophistication in the design of the virtual embodiments created different expectations about the capabilities of the people who were using those embodiments. The observational analysis revealed trouble with precise navigation, due to the small field of view on the desktop interface, interfered with their involvement in the task. Being able to point at objects that were essential to the task was important for the collaboration process to succeed. Being able to find other participants quickly, in order to interpret the meaning of their speech in relation to the shared objects, was difficult but essential to the collaboration. Finally, breakdown in the audio signal, or bad reception of the signal of one subject by the others, caused the others to ignore or disdain the subject with bad audio. Consequently, the subjects who experienced bad reception ended up feeling very frustrated and sometimes even angry, leaving them with a bad impression of the other subjects. The interviews confirmed these impressions.

The author of this thesis initiated the experiment, involved the third partner (IIS, Greece) needed to perform the experiment in a distributed setting, ran the subject drafting for the Nottingham site, did the subject guidance through the different stages of the experiment, conducted the post experiment interviews and questionnaire administration at the Nottingham site, analysed the video footage based on her observational method, wrote the section on "Observational analysis", and performed the final edit of the document in close collaboration with the designated editor of the journal.

9.3.3.3 Evaluation Results III

"Leadership and Collaboration in Shared Virtual Environments" (Steed, Slater, Sadagic, Bullock, Tromp, 1999) was published in the proceedings of the ACM Virtual Reality (VR) conference, and based on the same experiments as described in the section above (9.3.3.2). The paper is a shorter presentation of the same material, however the results of the analysis are presented in a more concise manner than the previous one. The present author of this thesis provided a transcript derived from the observational analysis, which illustrates the breakdown of trust between participants, see figure 9.6. The subjects in the experiments are named after the color of their embodiments: Mr. Blue, Mr. Green, and Mr. Red. This transcript was not printed in the final paper.

R: <is observing posters on the wall>.
G: Eh, I can't hear you, I am having problems, can you hear me Mister Red?
B: I really [..............] time [.........] words seem [........].
G: [........] find the [.......].
B: I will be incre[..]dibly h.h.h.h.ard to do communicating with Mister Green to do the task.
B: How [.]bout a twosome Mister Red?
R: Eh.m, yah, that's fine with me.
B: Mister Green? [....]
R: <turns around and positions to encompass both Blue and Green in his view>
B + G: <standing next to each other, facing Red>
B: So, we're discontinuing communication. We're not talking to you anymore [..] are you okay with that?
G: I ah [...] think [...] words [...] to understand s.s.s.s.s.s.omething.
B: Mister Red?
R: Sorry?
B: However, I don't know but [....] that [......] read the cards eh without eh talking to Mister Green.
R: I can't hear you either I'm afraid! Hahahah!!!
B: Okay, [........] just the task then and forget about Mister Green.
R: Eh, we're gonna do it without Mister Green than, yeah?
G: Can you hear me now? [......] NOW?
B: Okay [....]
G: Mister Blue, can you hear me now?
B: ah[..........] h.h.h.h.
G: Can you [.........] m.m.m.m.m.m.e?

Figure 9.6: Breakdown in the collaboration process. [25-06-98: Subjects B: 001, R:101, G: 201]

The author of this thesis initiated the writing of this paper, and provided a section on the observational analysis derived from the work presented in this thesis, which was dropped from the final paper due to space constraints.

9.3.4 CVE Design Documents

The two COVEN deliverables that cover CVE design guideline issues were not originally planned when the COVEN project proposal was written, however, the consortium felt that it had sufficient expertise to initiate the writing of such guidelines and an effort was made to create a new work package for the writing of these documents. The author of this thesis was the principal initiator of this shift in effort, backed up by the other partners present at the plenary meetings of the project where such issues were taken care of.

9.3.4.1 First CVE Design Document

The first guidelines deliverable, “Guidelines for Building CVE applications”, (eds. Dijkhuis, Liempd, Oudshof, 1997) was written by a large number of COVEN partners and consists of four chapters. Chapter 1 defines and clarifies the process of building a CVE and the associated tasks and corresponding roles involved in such a project: the decision maker, the designer, and the developer. The chapters correspond to these respective role definitions and further specify the type of knowledge and considerations such a role demands. Oddly, the role of the usability evaluator is missing from this overview, but this omission is rectified in the second guidelines deliverable. Guidelines for the decision maker included typifying the task for which the CVE is intended, cost/benefit decisions, and success and fail factors of CVEs. Guidelines for the designer covered what a user sees, how a user learns, and how to make a CVE usable. Guidelines for the developer involved computational issues involved in developing human actors, subjective views, mutual awareness, communication, resource management, participants roles and rights, object manipulation, global maps, group functionality, simplifying the architecture of a CVE, simplifying the scene, dividing the scene, and increasing the performance of a CVE. The author of this thesis wrote Chapter 2 “Guidelines for the CVE designer” pp.26-37.

9.3.4.2 Second CVE Design Document

The second design guidelines deliverable “Usability Design for CVEs” (Tromp, 1999) was aimed to directly cover the total CVE design task from beginning to end. It was written to guide all parties involved in the design of a CVE, to define and prioritise specific user and application needs and to find ways of supporting them within the limitations of the computing resources available. In other words, the method presented in the document enables all members of the design team to make informed assessments of the minimum usability requirements necessary to create a cognitively immersive CVE experience, based upon performance constraints imposed by the distributed, large-scale, multi-user nature of CVEs. In principle the method is an adaptation for CVEs of the traditional requirements analysis advocated in HCI-based design, informed by the COVEN usability experiences of the present author.

This deliverable was written by the author of this thesis. It is not yet published, and a summary version is presented in Appendix F of this thesis. These design guidelines were considered to be an important contribution to the commitments of the COVEN project to the EC to produce commercially applicable achievements from the project, and were received favourably during the presentation of COVEN achievements at the final audit of the COVEN project in Helsinki, 1999.

9.3.5 International Usability Discussion

Through the COVEN work and many conversations with other researchers working in the VR research and development field it became apparent that there was an increasing need for more usability knowledge about VE design and evaluation, a better overview of human needs and behaviours in and with VEs, and more integrated overview of the problems of VE technology. It was for this reason that the author instigated a workshop on this topic together with Dr. Chris Hand (de Montfort University, Leicester), and invited Dr. Kulwinder Kaur (City University, London), Dr. Anthony Steed (University College London), and Dr. Howell Istance (de Montfort University, Leicester) to join in. It was decided to first organise an international discussion about this topic in order to publicize this initiative, and to find out what type of questions people were trying to answer.

9.3.5.1 Usability Discussion

"Usability for VR Interfaces" (Tromp, Kaur, Hand, Istance, Steed, 1998a), was written to start off the discussion about usability and design specifically for VR technology. The topic was intentionally kept wide enough to include all types of VR technology and not just CVEs, as all of these areas were equally under-developed from a usability point of view. The discussion took place during the CVE98 conference, lasted for one hour and was well attended by approximately 40 people. There was a small group of researchers (two spoke out) who felt that existing HCI techniques ought be sufficient even for the evaluation of VR technology. However, the other researchers present argued that it might well be possible that the inherent 3D nature of VR interaction called for new evaluation and design techniques, or at least an adaptation of the traditional HCI methods. The workshop was announced at the conference and during the discussion, and via various mailing lists, internet news groups, and flyers at respective institutions.

9.3.5.2 Usability Workshop

"The First International Workshop on Usability Evaluation for Virtual Environments" (eds. Tromp, Istance, Hand, Steed, Kaur, 1998b) presented the combined papers (35) of all participants of the workshop, plus an introductory document written by the organizers to provide a background for the discussion; to situate the problem of usability for virtual environments in the context of what was already known about general usability evaluation and what might be novel about the VE interface type. The workshop participation was by invitation only, these were based on position papers submitted by the applicant participants. In total, 42 participants were chosen from a wide range of academic, industrial backgrounds (both theoretical and design), ensuring a broad range of expertise in different VE platforms and application areas. All papers were reviewed by two members of the organising committee, and judged on indications of sufficient background knowledge of the authors about general usability to ensure a high level starting point for the workshop discussions. All accepted papers were made available to the workshop participants before the workshop via web pages. The workshop was a one-day event consisting of interactive and free-flowing group discussions. The workshop was organized into two separate sections: CVE Usability Issues and CVE Usability Methods. These were derived from an analysis of the type of papers we received, and refined based on our own expertise. The usability issues we identified were: VE interaction, navigation, social interaction, presence, and utility. The usability methods we identified were: controlled experiments, user observation, user reports, interface inspection, and design guidance. All participants were assigned to a group by the organizers (although they were free to change group), based on the contents of their position papers, their background, and their personal interests. On average there were eight participants in each group. The groups, the expected contribution of the topic to the goal of the workshop, and the main question for discussion are summarized in Table 9.10.

Type	Group	Contribution	Main Question
Usability issue	VE Interaction	Understanding how users interact with VE	What are the criteria for evaluation?
Usability issue	Navigation	Key interaction activities	What are the criteria for evaluation?
Usability issue	Social Interaction	Important activity in CVEs/multi-user VEs	What are the criteria for evaluation
Usability issue	Presence	Affects the interaction experience	How does presence impact on usability and usability evaluation?
Usability issue	Utility	Important contextual issue	What constraints does the general utility of VE place on usability evaluation?
Usability method	Controlled experiments	Evaluating the effect of specific features on usability	How to perform controlled experiments as part of a usability evaluation?
Usability method	User Observation	Primary method for usability evaluation of an application	How to perform observation of users as part of a usability evaluation?
Usability method	User reports	Efficient for large-scale testing of general usability	How to use user reporting as part of the usability evaluation?
Usability method	Interface Inspection	Efficient for general usability data in the absence of end-users	How to perform inspection as part of the usability evaluation?
Usability method	Design Guidance	Important in solving problems found in usability evaluation	What design guidance is needed to support evaluation and development of usable VEs?

Table 9.10: The topics, contributions and questions for each discussion group of the UEVE98 workshop.

The organizers presented an introductory session in which the format of the workshop was explained. The group sessions took place in parallel and were chaired by one of the organizers, who took it upon themselves to moderate and facilitate the discussions. Large sheets of paper, plenty of colour pens and yellow tag note blocks of different sizes were made available for each group, to assist the participants in their discussion. Each group discussion was started by asking the participants to introduce themselves, their interests, and their affinity with the group topic. The chairs summarized the objectives of the group and their questions for discussion. Members of each group were selected to report the results of the group discussion back to all workshop participants in two plenary sessions. The final session of the day was dedicated to summarizing and bringing together any conclusions that could be made with all workshop participants together. During each group discussion common themes were identified, however no definite answers to each group's questions were found, instead a more clarified picture was derived from each discussion group about the type and direction of research needed to answer the questions; each researcher taking back their own trace of the workshop.

9.3.5.3 Early PhD Results: Usability Workshop Paper

"Designing Flow of Interaction for Virtual Environments" (Tromp and Fraser, 1998), has been published in the proceedings of the first international workshop on Usability Evaluation for Virtual Environments (UEVE98). The paper presents the first results of the research performed for this thesis. It argues that the design of spaces and objects for VE needs special attention, because the freedom of navigation and interaction in VEs make the flow of interaction less predictable than for 2D systems. The authors explain that VE users need additional guidance in terms of flow of actions. Real- time image generation puts a high load on machine processing time and, in the case of networked VEs (CVEs), on the network traffic generated by the real-time image update. To reduce this load the functionality and appearance of the environment and objects are reduced to the bare minimum, so it is often difficult for users to predict which operations are available and which are not. Because of the freedom of navigation and interaction in VEs, it is difficult to predict what actions users will take, and in what order they will perform their actions. Users have been shown to struggle with finding the right order in which to perform actions, with finding their way through the environment, and navigating into precise positions. The correct performance of actions and sequences of actions in the VE greatly relies on the design of the interface of the VE and its objects. User actions in VEs seem to oscillate between user created 'story-lines' (i.e. the successful performance of action sequences) and user interface struggles (i.e. the inability to perceive or perform the correct action, the absence of feedback, or both). The authors argue that VE users need more obvious structure in the environments and object interactions. The designer will have to help the user identify the actions and objects necessary to perform their tasks, especially the order in which they are to be used. Some chunks of information have to be interpolated by the user; some have to be attached to the objects, before the user can make sense of the environment. The sequence of appearance is extremely important. These items of information function as elements of a story, and while the arrangement may be flexible and open, the elements have to be assembled in a particular order to make sense of the story. This order should be designed to guide the users through their tasks. The perceptual affordances of 3D objects in VEs need to be improved by simplifying the objects so that their functions are amplified as much as possible. The sequential affordances of a task, which involve interaction with multiple objects in a certain, specific order, need to be designed with more care by directing the users' attention from one object to the next, as desired. Generally, partial tasks can be automated, and guidance for sequential affordances can be provided by structuring the lay-out of the rooms and position of the objects more deliberately. Instead of arbitrarily positioning objects and rooms, they can be grouped and ordered into meaningful parts which intentionally draw the user from action to action. Standard HCI alerting techniques for guiding user attention to the next action, such as the use of colour, flashing, and reverse video, are not very elegant solutions in a VE, especially when it concerns the design of CVEs. However, the use of spatial and temporal cues and audio warnings may be much more effective. Designing the spatial layout of rooms and objects more carefully, and providing more carefully designed object affordances could improve the usability of VEs. The paper concludes by stating that simplifying the VE by deliberately designing caricatures of objects and situations may be a more effective way of keeping machine load down, without losing usability points. Tromp wrote the paper, and Fraser provided additions.