


# • D3.5, Usage Evaluation of the Online Applications

# Part A: Overview and Summary of Evaluation

| | |
|---|---|
| **Project Number:** | ACTS Project N. AC040 |
| **Project Title:** | COVEN - Collaborative virtual environments |
| **Deliverable Type:** | P* |

| | |
|---|---|
| **CEC Deliverable Number:** | A040-UCL-CS-DS-P-035a.b1 |
| **Contractual Date of Delivery to the CEC:** | 31st October 1998 |
| **Actual Date of Delivery to the CEC:** | 6th November 1998 |
| **Title of Deliverable:** | Usage Evaluation of the Online Applications |
| **Work package contributing to the Deliverable:** | WP3 |
| **Nature of the Deliverable:** | R** |
| **Author(s):** | A. Steed (UCL) and J. Tromp (Nottingham) |

**Abstract:**

This document summarizes the work undertaken in the usability evaluation of the COVEN online applications. This is the first part of a four part deliverable, and it provides an overview of all results and summaries of the other three parts.

Several threads of work were undertaken in this evaluation effort. These include network trials, development and evaluation of an inspection method for Collaborative Virtual Environments (CVE), development of a theoretical overview of CVE tasks, a series of small group experiments, and a series of experiments that investigate the sense of presence.

The results of this activity include a list of recommendations for future implementation and a set of evaluation methods and premises for future investigation

**Keyword list:**

Collaborative Virtual Environments, Usability Inspection, Presence, Cognitive Walkthrough, Heuristic Evaluation, Design Methods, Network Trials

***Type: P-public, R-restricted, L-limited, I-internal**

****Nature: P-Prototype, R-Report, S-Specification, T-Tool, O-Other**

## Executive Summary

The second phase of the COVEN usability evaluation activities addresses the evaluation of the online versions of the COVEN demonstrator applications. In this we attempt to further identify further refinements to the applications and investigate the effects of network and interface properties on collaboration in virtual environments.

This Deliverable consists of four parts in which we outline results about the applications and new methods we have developed for general application in the evaluation of collaborative virtual environments (CVEs). This activity of COVEN has been pioneering techniques for CVE evaluation since, as we noted in the first evaluation phase, standard human-computer interaction techniques are not easily applicable to CVE applications and systems.

Five threads of work are reported in this deliverable:

- Theoretical overview of Collaboration issues for CVEs.
- Network Trials of the applications over the experimental network with *Expert Users*.
- Usability Inspection of the on-line applications using method COVEN has developed.
- Auxiliary Trials to investigate more fundamental properties of CVE technology.
- Small Group Experiments using *Naïve Users.*

The main threads from the first usability effort are represented here in the inspection, network trials, and auxiliary trial threads. In addition we have developed a separate experimental thread that investigates the behaviour of small groups of participants under taking tasks in a CVE. Another new thread is a more formal formulation of the tasks and actions involved in operating within a CVE. The results of this thread (a CVE task analysis and a CVE observational scoring system) are used to under-pin the inspection and network trial.

The main results of this activity include an analysis of usability issues with the current applications to be used as input to the final implementation phase of the demonstrators. We also highlight a number of CVE technology issues that should form the basis for future experiments.

**1 INTRODUCTION ..... 4**

1.1 EVALUATION ACTIVITIES WITHIN COVEN ..... 4
1.2 PROGRAM OF WORK FOR SECOND EVALUATION PHASE ..... 4

**2 FRAMEWORK FOR EVALUATION OF THE ONLINE APPLICATIONS ..... 6**

2.1 THE PARAMETERS OF THE PROJECT ..... 6
2.2 THE GENERAL VALIDATION SCENARIO ..... 8
2.3 TASKS, USER GROUPS, AND CONTEXT OF USE ..... 10
2.4 SUCCESS FACTORS, QUALITY FACTORS, AND ASSESSMENT CRITERIA ..... 11
2.5 SELECTION OF METHODS - EVALUATION PLAN ..... 13

**3 SUMMARY OF USABILITY INSPECTION ..... 14**

3.1 INTERACTION CYCLES ..... 14
3.2 OVERVIEW OF INSPECTION METHOD ..... 14
3.3 RESULTS AND INPUT TO APPLICATION DESIGN ..... 15
3.4 METHOD DEVELOPMENT ..... 16

**4 SUMMARY OF COLLABORATIVE ACTIONS IN CVES ..... 17**

4.1 SPECIFIC GOAL OF CVES ..... 17
4.2 OUTLINE OF METHOD ..... 17

**5 SUMMARY OF THEORY AND METHOD DESCRIPTION FOR SPATIAL AND TEMPORAL ACTIVITY CHARTING OF CVE INTERACTIONS ..... 18**

5.1 OVERVIEW OF THE THEORY AND METHOD ..... 18
5.2 THE CATEGORIES ..... 18
5.3 APPLICATION AND RESULTS ..... 19

**6 OVERVIEW OF NETWORK TRIALS ..... 20**

6.1 ISDN/DVS TRIALS ..... 20
6.2 INTERNET/DIVE TRIALS ..... 22
6.3 DVS GENERAL COLLABORATION RESULTS ..... 25
6.4 DIVE GENERAL COLLABORATION RESULTS ..... 27
6.5 CONCLUSIONS ..... 29

**7 SUMMARY OF SMALL-GROUP EXPERIMENTS ..... 30**

7.1 INTRODUCTION ..... 30
7.2 EXPERIMENTAL FRAMEWORK ..... 30
7.3 EXPERIMENTAL DESIGN ..... 31
7.4 TECHNICAL DESCRIPTION ..... 33
7.5 RESULTS OF THE STATISTICAL ANALYSIS ..... 34
7.6 DISCUSSION ..... 35
7.7 CONCLUSIONS ..... 37

**8 OVERVIEW OF AUXILIARY TRIALS ..... 38**

8.1 INTRODUCTION ..... 38
8.2 BREAKS IN PRESENCE EXPERIMENTS ..... 38
8.3 VISUAL ATTENTION EXPERIMENTS ..... 40

**9 COMBINED RECOMMENDATIONS ..... 44**

9.1 COMBINED RECOMMENDATIONS ..... 44
9.2 METHOD DEVELOPMENT ..... 45

# 1 Introduction

## 1.1 Evaluation Activities Within COVEN

The COVEN project life is planned along three main design cycles, with the production of three versions of the Demonstrator application prototypes (initial, on-line and final). Usability evaluation activities in COVEN are broken down into three stages corresponding to these three design cycles. In addition to these three major steps in the application design, it is foreseen that small-size design adjustments may be performed on the prototypes of a given cycle depending on the results of the evaluations performed during this cycle. The COVEN usability evaluation process has been planned for each of the three stages so as to take into account the constraints (e.g. application status), the context (in particular, experience and results of the previous evaluation stages) and the specific objectives at each design cycle.

At end of the second implementation stage our primary goal is to assess the usability of the second implementation of the demonstrators stage of the applications (the "Online Applications") and to inform the design of the final application. Thus we are in the second part of the iterative engineering process, working with extended protoypes of the application. In many ways we are still exploring how to support collaboration in Collaborative Virtual Environments (CVEs) and thus we have formed a work program that is broad in scope. We try to demonstrate in this deliverable some of the added value of CVE technology.

## 1.2 Program of Work for Second Evaluation Phase

The work program adopted in Activity 3.2 was very successful in identifying different types of usability issue with CVE technology. We followed the same strategy in this iteration of the trials, following the revised Framework decsribed in Section 2. Six strands of work have been undertaken:

1. Theoretical overview of Collaboration issues for CVEs. (see Sections 4 and 5)
2. Network Trials of the applications over the experimental network with *Expert Users*. (See Section 6)
3. Usability Inspection of the on-line applications using methods COVEN has developed. (See Section 3)
4. Auxiliary Trials to investigate more fundamental properties of CVE technology. (See Section 8)
5. Small Group Experiments using *Naïve Users.* (See Section 7)
6. Instigation of *Workshop about Usability Evaluation of Virtual Environments* (Dec 17, 1998*).*

The first thread of work consists of a number of different elements, all addressing the central CVE issue of collaboration. A review of theoretical psychology on verbal and non-verbal human behavior has been made, to gain a detailed understanding of how people collaborate, (e.g. communicate and interact) in the real world and how this compares to the virtual world. Next, a hierarchical task analysis of collaboration in CVEs has been created (see summary in Section 4), describing the consequences for effective collaboration when certain elements of the task are unsupported by the application. This informed the development a collaboration cycle of interaction for the Inspection method (the third thread, described below). Finally, a new research method has been developed to compare *the actual collaborative behaviors*, as they can be observed from video-recordings of users performing tasks, with *the ideal collaborative behaviors* as predicted by the hierarchical task analysis. This method ("Spatial and Temporal Activity Charting for CVEs", see summary in Section 5) has been pilot tested on two COVEN network trials, and has been passed on to the Network Trial team for further testing against the network activity measurements of user activity.

The second thread of work used the results of the first iteration of the usability evaluations (Del 3.3) to clarify topics of importance for the scientific inquiry into usability issues for CVEs. A structured workshop-style approach with domain-experts ("Group Elicitation Method", (Helander et al, 1997) has been used to sort through the many concepts relevant to CVE research and development, refining, combining, and ranking the concepts into a task structure for use in the Network Trials. In total 7 feasible tasks were defined (see Table 5). The actual work carried out deviates somewhat from the plan developed in the Framework in Section 2. Due to problems in producing a working version of the online application, the number of tasks carried out in the dVS application had to be cut down. Some of the tasks were carried over into the experimental Dive trials, and in other cases the factors were covered by the Small Group Experiments (the fifth thread, described below). Additionally, a questionnaire, inquiring into the general collaboration issues (as identified by the activities in the first thread, reported above), has been administered to the users in the network trials when no planned tasks could be performed.

The third thread of work regards the development of the Inspection Method for CVEs, as initiated during the first iteration of the Usability Evaluations (Del 3.3). Based on the conclusions from the first Inspection, the method has been rigorously adapted to inquiry into the unique 3D and collaborative aspects of CVEs, see Section 3. The 3D Inspection issues have been adapted from previous work conducted by Dr. Kulwinder Kaur and Prof. A. Sutcliffe, from the Centre for HCI Design, City College, London (Kaur, Maiden, Sutcliffe, 1997). The collaborative issues have been derived from activities in the first thread. This adapted method has been tested on the COVEN Platform. Advice and cooperation to further refine this method has been sought from Kaur & Sutcliffe. Additionally, in order to gain an understanding of CVE designer problems and needs two small scale inquiries were made. One, a questionnaire, into the reception of the previous Inspection results, the format, and the effectiveness, and the second, a set of interviews with CVE designers, into their design practice.

The fourth thread of work, auxiliary trials, has proved to be effective to explore fundamental issues for the design and development of CVEs, in order to work around the constraints on usability evaluation methods available with a prototype CVE (see Section 8). This has resulted in two main experiments. One was conducted in order to demonstrate how presence might be objectively measured within a virtual environment using a method called the *presence counter*. The second explored perceptual issues associated with VE technology.

The fifth thread of work has been developed for a number of reasons (see Section 7). These experiments were designed to inquiry into the effect of latency on user performance, the effect of network performance on display, and the study of network performance on interaction and cooperation. We decided to make use of novice users in networked set-up, once this was flagged as feasible by the Network Trial team. Firstly, the use of novice users prepares us for the final evaluation in Activity 3.5. In the first iteration of the Usability experiments, working with novice users as subjects, was identified as a constraint on the scientific inquiry process. During this iteration the stability of the platform and our experience in running trials seemed sufficient to contemplate carrying out scientific trials over a distributed network, given short training on how to use the interface to the applications. Secondly, we have observed turn taking and body movement conventions with expert users and we wished to observe whether and how such conventions arise with novices, especially since our experts users are likely to have adapted to the media in order to maximize their productivity and collaboration in the CVE.

The sixth thread of work involves the instigation and co-organization of the first International Workshop on Usability Evaluation for Virtual Environments, to take place in the Montfort University, Leicester, United Kingdom on December 17, 1998. The workshop is sponsored by COVEN, the Montfort University, and the British HCI Group. Support in the form of advice has been sought from Prof. J. Wilson, VIRART, Dep. of Manufacturing Engineering, University of Nottingham; Prof. S. Benford, Communications Research Group, Dep. of Computer Science, University of Nottingham, Prof. A. Sutcliffe, the Centre for HCI Design, City College, London. The workshop is organized by Jolanda Tromp, University of Nottingham (COVEN), Chris Hand, Napier University, Edinburgh, Howell Istance, De Montfort University, Leicester, Kulwinder Kaur, City University, London, Anthony Steed, University College London (COVEN). Further details on the workshop can be found on http://www.crg.cs.nott.ac.uk/research/technologies/evaluation/workshop/workshop.html.

# 2 Framework for Evaluation of the Online Applications

This section describes our evaluation process for the second stage of the COVEN usability evaluation, to be performed on the online version of the applications. We step through each of the evaluation planning stages (steps 1 to 5 in Figure 1) so as to reach a full and well-argued definition of our usability evaluation plan.

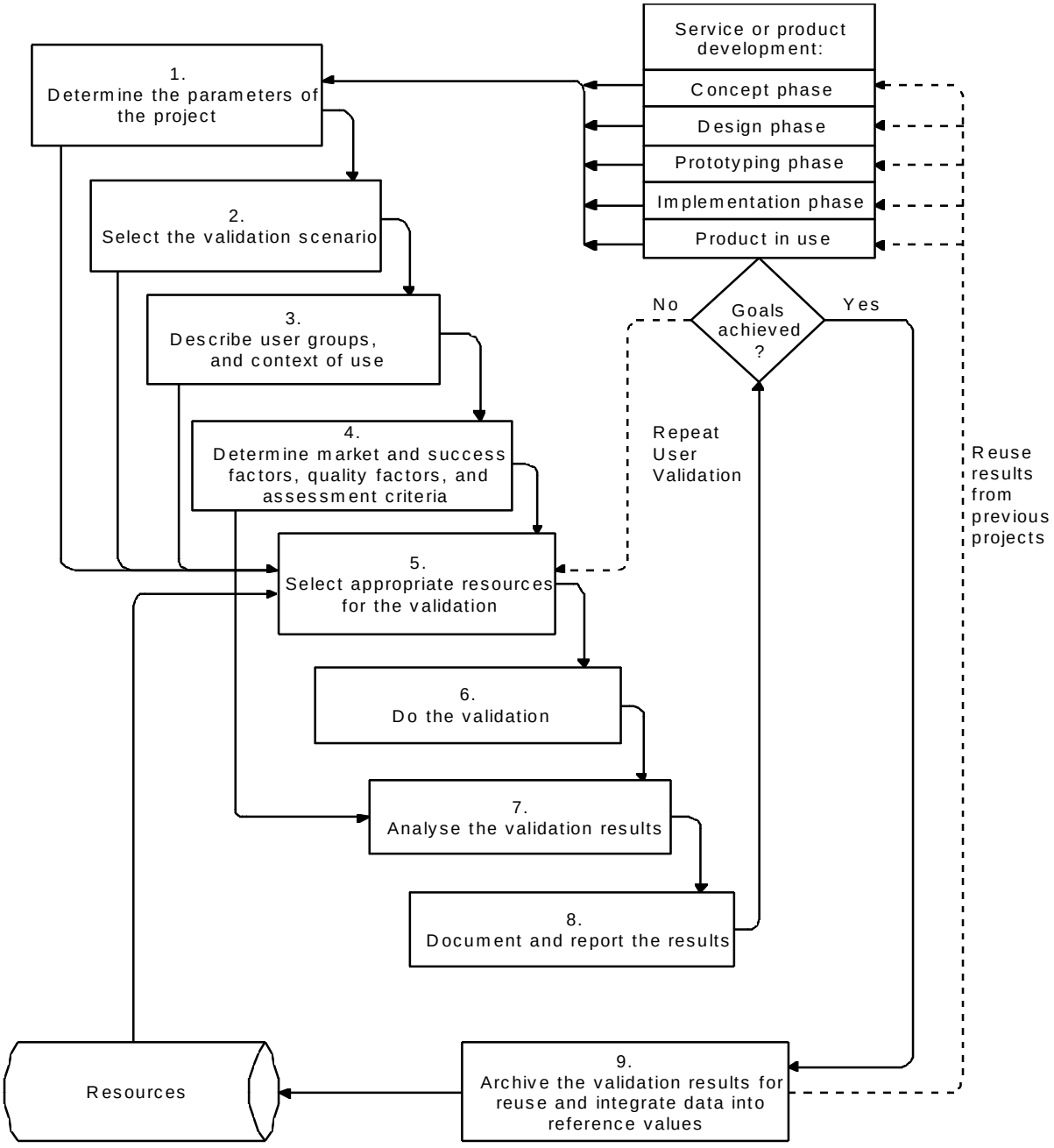



*Figure 1: Structure of a Standard Usability Evaluation Process*

## 2.1 The parameters of the project

*The constraints of different projects in terms of cost, time, and other factors are very heterogeneous. The*

*precise understanding of the parameters of the project is the prerequisite for planning the user validation process and for an agreement with users of the validation results[28].*

It has to be noted that the applications are still at prototype stage of development, and the networking activities are increased from 4 simultaneous users to a planned 8. While planning the Usability studies it is not clear how fast or effective the network trials will support a usable connection between all planned users. We have therefore planned a number of alternative avenues of inquiry into the general area of usability for CVEs, and the usability of the COVEN Platform in particular, where possible.

### 2.1.1 Type of services and products

The COVEN Demonstrators currently under evaluation are the Citizen-application (in the previous usability evaluation referred to as 'Rhodes traveller'), the Business-application (in the previous usability evaluation referred to as two parts 'Virtual conferencing' and '3-D spreadsheet') both built on top of dVS, and a separate research application built upon DIVE. Together these applications allow us to evaluate and explore collaboration in CVEs, relating to four categories of research:

- Group Navigation
- Group Interaction
- Group Communication
- Co-Presence

The Citizen application and the Business application contain implementations of the four categories of research (e.g. group navigation, group functionality, group communication, and co-presence), although each application makes use of these categories to different degrees. The DIVE application shares characteristics with the first two types of dVS systems; although it has different implementation strategies. For the Usability studies this means that instead of having only one kind of CVE to test, with the risk of generalizing findings based on one case only, we have the ideal situation of comparison between different design solutions. Additionally, some of the constraints on the scientific inquiry caused by the fact that we work with prototypes is alleviated because some things dVS can not do (for example: integrated audio) are possible in DIVE, and vise versa.

### 2.1.2 Objectives, requirements and constraints of the project

The main objectives, requirements and constraints of the project are described in Deliverable 3.3. To summarize: COVEN is developing Demonstrator applications intended to demonstrate the added value of the CVE concept from the end-user and the customer point of view. General constraints have to do with the 'scientific inquiry' dimension of the COVEN project. To be precise:

a) There are no evaluation methods which specifically inquiry into the usability of CVEs.

b) It is relatively unknown what the specific usability criteria are for CVEs.

c) Existing HCI methods need to be adapted. New methods may have to be developed.

d) In order to satisfy these needs (which are shared by all members of the scientific community involved in the design and development of VEs and CVEs) the COVEN Usability team has adopted a broad exploratory approach, assembling, integrating, and disseminating the results in a generally re-usable way.

More specific constraints on the usability activities stem from the immersive aspect of CVEs, as well as from the prototypic and distributed nature of the applications.

The objective of this second evaluation is to refine the evaluation techniques for CVE applications and technologies, and apply them to the existing online demonstrators in order to direct the final implementation effort. In particular we generate guidelines and report documents that comment upon the implementation of the 2$^{nd}$ versions of the demonstrators, which will be exploited in the context of A2.9. The output from this second evaluation is aiming to address the effect of latency on user performance, the effect of network performance on display, and the study of network performance on interaction and cooperation.

At this intermediary stage of the applications, the available prototypes still feature a limited set of functions; although redesign has taken place based on the usability reports from the first evaluation. As in the first evaluation, some bugs may remain, although efforts are made to remove high level bugs by allowing the designers to apply the Inspection method, by allowing ample network trial tests to take place before starting the experiments, and by slowly introducing more participants into the applications, thus building up a more secure level of understanding of the participants in the network trials. To make it more feasible to test the applications with representative end-users, efforts were made to develop a short training to be used when testing with novice users.

### 2.1.3 Users of the evaluation results

At this second stage, users of the evaluations results are the designers and developers of the applications, as well as the human factors researchers in the project. Additionally, results are used to create seminars for the SID-Chain of the ACTS Projects, a joint conference submission paper to VR'99, and a joint journal submission paper to IEEE Computer Graphics & Applications. Finally, two of the COVEN usability experts are involved in the organisation of the first International VE Evaluation Workshop, taking place in December 1998.

### 2.1.4 Decisions to be taken on the basis of the usability evaluation results

Based on the results from the Inspection of the Initial Applications and the results from the usability experiments (reported in Deliverable 3.3), the applications have undergone a major re-design during Activity 2.4 and 2.5. It is our expectation that the results from this second iteration of the usability evaluations will inform the designers again. A questionnaire has been administered to make sure that the usability advice comes in an effective format for the designers. Additionally, the results from the usability studies are expected to help define usability criteria for the final COVEN usability evaluation.

### 2.1.5 Resources constraints for the usability evaluation

Constraints have not changed from previous Usability effort.

## 2.2 The general validation scenario

### 2.2.1 Phase of the project life cycle

The currently planned evaluations are taking place at the end of the second design cycle iteration; one more iteration is planned.

At this intermediary stage of the applications, the available prototypes still feature a small set of functions; and it is certainly expected that some major bugs are found.

### 2.2.2 Focus of the usability evaluation

*(Merchior at al., 1995): Classically, user validation can focus on different related aspects of the application:*

*Functional content: adequateness to the target domain tasks*

*Information content: correctness of the information as well as adequateness of the media (audio, video, graphic, text) selected for the presentation of the information.1*

*Functionality of the user interface: the user validation of the user interface looks at features and properties of the dialogue and presentation components between the user and the user-interface of the application. Assessed is the quality of use of the functions provided for realising the target domain tasks.*

*Load imposed on the user: the assessment of load imposed on the user looks at the properties of the tasks which the user intends to perform with the service. This may be the properties of information content or the functionality of the application, and the nature of decisions the user performs on the basis of this information.*

*Stress perceived by the user when using the service: stress in terms of cognitive load and subjectively experienced discomfort is a function of the load on the individual user, and factors such as individual differences, learning state, and fatigue.*

*Cost/benefit aspects of the service.*

*...*

*Validation can focus on only one of these aspects or take more than one viewpoint into account.*

---

1 The information content aspect is particularly emphasised in 'classical' information engineering applications; this seems also relevant for the CVE applications due to the focus on multi-media presentation and 3-D rendering of information; moreover, as stated before, our Citizen Application can be looked at as an information engineering application.

Results from our previous evaluation work lead us to emphasize a specific conceptual evaluation focus. The exploratory research (the “Business Trading Game” and the “Plan a holiday” scenario) which took place during the first usability evaluations (A3.3) were reworked in a focused workshop-style meeting. This meeting provided the following decision structure:

1. Issue statement and formulation.
2. Recruit domain experts: KPN: Industrial designer, Presence research Psychologist, UCL: Presence research Computer Scientist, NOTT: Presence research & methodology Psychologist.
3. View-point generation: brainstorm
4. Reformulate into more elaborated concepts: reduce large list into smaller number of more central concepts (used: COVEN Services & Quality Criteria as guide to central concepts, see Del. 3.3 for more details).
5. Refine, reduce, rank, and establish relationship among the concepts.
6. Design experiments & tasks based on concepts in relationship; link to Network trials (see Table 1 below).

| Nr | Concept | Contains |
|---|---|---|
| 1. | Rebuild Rooms | • Object Manipulation<br>• Group Behaviour |
| 2. | Treasure Hunt | • Gestures<br>• Level of detail on Avatars |
| 3. | Cluedo | • Effectiveness of communication media<br>• Non verbal communication<br>• Object Manipulation<br>• Group navigation |
| 4. | “balletje balletje” | • Gestures<br>• Use of explicit focus<br>• Object manipulation |
| 5. | Cannibals (Teleport) | • Explicit group formation<br>• Implementation of metaphors<br>• spatial behaviour of small group |
| 6. | Plan a holiday | • Guided tour vs. non-guided tour<br>• Communication media<br>• Group formation |
| 7. | Travelling back | • Mental model<br>• Group navigation<br>• Conceptual knowledge<br>• Information presentation |

*Table 1 Usability experiments and related concepts*

The focus of this second stage of evaluations is on functional and information content, functionality of the user interface, and collaboration support, and latency effects. An Inspection method to clean up the application design has been adapted to inquire into CVE specific problems, such as: collaboration and 3D. A hierarchical task analysis (HTA) of collaboration has been created, as a guide to our explorations into the concept ‘collaboration’ in CVEs. Additionally,

three experiments were designed which all address the concepts raised in more or less depth, so that the combined results give us information on all concepts in relation to each other.

### 2.2.3 Selection of the user validation scenario

*(Merchior at al., 1995): A suggestion of the most common user validation scenarios is:*

*innovation: this is the case where an innovative application is developed. Competitive applications do not exist. The emphasis of user validation is on identifying critical deficiencies early in the development process.*

*development of an electronic information product or service: here it is assumed that traditional products, e.g. paper-based documentation, newspaper, journals, exist which are in competition with the electronic information product or service under development.*

*benchmarking: competitive applications exit and will be evaluated to provide a baseline for the development of the new application.*

*product development process: here, prototyping is supported by feedback from user validation in several iteration cycles.*

*conformance test: the objective is to check whether minimum requirements set by standards, legal requirements, design guidelines, etc. are fulfilled.*

*user acceptance test: the objective is to test user acceptance of an application including aspects as attractiveness and enjoyability of the application.*

*etc.*

In this second stage of evaluations, our main scenario is a product development scenario or, more accurately, a prototype development scenario as our objective with the COVEN Demonstrators cannot be to reach a 'product' level of achievement. This is not a shortcoming of COVEN, it is simply the place the COVEN research task has in the product development cycle. Based on the place in this cycle, certain (usability) activities are needed, wanted or impossible.

Note that we are here in a situation that potentially mixes several scenarios: in some regards, the services we are developing are fully innovative, with no competitive applications; however competitive applications do exist for some functions of our services taken separately: video-conferencing, 2-D shared editors, multimedia catalogues, web services, paper-based documentation, 3-D interactive VEs, etc. In has to be notes here also that for some of these services, and also for the CVE service as a whole, there are no specific usability guidelines. We see it as part of our task to develop and adapt methods of evaluation and to generate guidelines for design. It is for this reason that we are working on an adaptation of the tradition Inspection method, which is now in it's second iteration of development. Additionally, we are developing the Spatial and Temporal Activity Scoring method to establish a quick and systematic measurement technique for user action in CVEs.

## 2.3 Tasks, user groups, and context of use

### 2.3.1 User groups

Concerns about user groups are similar to those during the first usability evaluations. This results mostly in a concern about the stability of the application. The current, unpredictable instable state of the applications does not allow easy testing with end-users. More positively, from a Usability point of view is that the number of available subjects has increased from 5 to 7. The subjects in the trials are 7 programmers and developers of CVE technology and computer science students. Additionally, efforts have been made to test the application with novice users, who have undergone a short training of the interface to a subsection of the application (which has been chosen for its stability).

### 2.3.2 Tasks users intend to perform

*(Merchior at al., 1995): Regarding only the functionality of the service, you can neither tell whether the functionality is sufficient or insufficient, nor can you detect that there is too much functionality which the user does not need. Based on task analysis results, the adequateness of the application for the user tasks can be estimated. Some users may use the application on a regular basis, others only occasionally. A best guess about the frequency of use and the execution of tasks with the application should be made for a*

*defined time interval (total life time of the application, per day, per week, or other interval).*

The general task of collaboration users perform is fully analysed in a hierarchical task analyses. It provides a breakdown of user actions for collaboration in CVEs (see Del3.5b: Usage Evaluation of the Online Applications, Part B: Collaborative Actions in CVEs).

### 2.3.3 Context of use

*(Merchior at al., 1995): Finally, user validation must be performed with reference to the context in which the service will be used. The context of use is defined by:*

*the environment in which a service is used, at home, in public places, at the workplace. Noisy environments and lighting conditions may impair the perceived quality of the application.*

*the properties of the technology into which the application will be integrated.*

The foreseen context of use for the Citizen application is the home; the foreseen context of use for the two business applications is the workplace, although some target Virtual conferencing usages may be the home also, in a 'working at home' or teleworking perspective.

The COVEN developments are currently performed on high-end equipment; however, target technology (especially for the citizen-oriented Citizen application) is 'low-end' PCs - we are relying on technological advances to take place in the coming years.

## 2.4 Success factors, quality factors, and assessment criteria

### 2.4.1 Critical success factors

*(Merchior at al., 1995): To arrive at a precise description of assessment criteria for the user validation, it is necessary to start with initially identifying the critical success factors. Critical success factors determine the success or failure of the product or service for the organisation which is developing the application.*

As stated in previous Deliverables, in the specific COVEN organisation context, the most important success factor for our applications is to produce services with added value to the user. Factors like aesthetics and enjoyability are also important, especially with regards to citizen-oriented services. Additionally, we have declared 'support for collaboration' as the critical success factor for our evaluations.

### 2.4.2 What needs to be validated: quality factors

*(Merchior at al., 1995): Quality factors are variables which reflect different independent quality aspects of the product or service. Most of these quality factors refer to the user interface and have been developed over a long time. They are included in standards and norms (ISO 9241). These quality factors have also been a basis for developing measurement procedures and measures which can be used to measure the quality factors. The quality factors identified so far are2:*

*Adaptedness of the application to the intended task domain*

*This factor describes whether the application provides sufficient functionality to perform the intended tasks. Superfluous functionality may impair the use of the application because of increasing learning effort.*

*2. Efficiency of task performance*

*If an application cannot be used efficiently, users may not want to use it. Fast and efficient execution of tasks must be provided, and procedures for the execution of tasks should be homogeneous regardless whether the state of the application has changed.*

*3. Information content*

*This quality factor indicates the correctness and other aspects of the information content.*

---

2 N.B.: dependencies exist between these quality factors (e.g. Display deficiencies and User problems or Information content).

*4. User problems and design deficiencies*

*This quality factor indicates the shortcomings of the functionality of the application.*

*5. Display deficiencies*

*This quality factor indicates the shortcomings of the display of the application.*

*6. Robustness*

*This factor describes how well the application can cope with user errors and indicates the effort for error correction.*

*7. Error frequency*

*This quality factors indicates the frequency of error occurrence.*

*8. Learning cost*

*Especially users from the general public may not be motivated to learn how to use the application. This factor describes the learning and training effort required by an application.*

*9. Workload*

*This quality factor describes load imposed on the user by the application and stress perceived by the user when using the application.*

*10. Subjective ratings of users*

*Motivational factors and aspects such as attractiveness, innovativeness, enjoyability are described by this factor.*

*11. Conformance with standards and minimum requirements*

*This quality factor indicates whether standards and other minimum requirements defined by legal requirements, guidelines, etc. are fulfilled.*

The quality factors under consideration in our working context have not changed from the first Usability evaluation efforts..

In this second stage of evaluations, we focus on the quality factors directly related to function and information content, functionality of the user interface, and collaboration support aspects: 1, 1bis, 2, 3, 4 and 5. Quality factors 7 (error frequency), 8 (learning cost) and 10 (subjective ratings) are also in our focus as these factors precisely correspond to intrinsic claims of the CVE technology - see below.

#### 2.4.2.1 CVE technology claims - hypothesis under focus

An overall hypothesis can be deduced from VR development effort. Implicit in the documentation of VR projects is claimed that the added value of VEs is that they are intuitive to use. From these implicit statements the following deduction can be formulated:

A. VE are realistic by making use of 3 dimensional representations

B. A realistic interface is intuitive

So from A) and B) follows:

C. Virtual environments are intuitive

The question we investigate is how we can make this wish come true. We use this as an overall hypothesis to our experiments.

### 2.4.3 Assessment criteria for quality factors

*(Merchior at al., 1995): Finally, for each of the quality factors, measures, by which the quality factor will be specified, must be determined and assessment criteria must be defined. Criteria are values for the measures which are the basis for the assessment of the application. A criterion can be defined as*

*a critical value which should under no circumstances be exceeded*

*an acceptable value*

*an optimal value which is not necessarily required from the user validation results, but which would be advantageous for the success of the application on the market.*

*It must be noted here that quality criteria can sometimes conflict. Normally, not all quality factors can be given equal weight in a user validation project. The priorities must be made clear on the basis of the information about the user groups, tasks, and context of use. E.g. for users who will use the application only casually, learning cost will be a critical quality factor.*

The assessment criteria remain the same, see section X for details.

## 2.5 Selection of methods - evaluation plan

In this second evaluation stage, our evaluation plan consists in a number of evaluation actions making use of both inspection and empirical methods, including case-controlled experiments and general observations.

### 2.5.1 Choosing evaluation methods

Several methods have been chosen to address the usability of the applications:

- Inspection updated cognitive walkthrough. We have expanded and updated our Inspection method so as to directly address CVE specific issues.
- Empirical measurements. We have defined a number of empirical actions addressing different usability concepts.
- Model of Collaboration. We have created a hierarchical task analysis of collaboration in a CVE, based on scientific literature.
- Technique to measure user actions. We have created a method for scoring the behaviours of small groups, the Temporal and Spatial Interaction Scoring method.

# 3 Summary of Usability Inspection

The Inspection method described in Part C of this deliverable uses a new method developed by COVEN to address issues unique to usability testing of Collaborative Virtual Environments. This method is an adaptation of the traditional HCI-based Inspection[31] and is based upon an early version of the pioneering development of an Inspection method for single user VEs [16]. The need for adapted techniques for the evaluation of CVE applications was identified during the first evaluation phase of the COVEN project[3].

## 3.1 Interaction Cycles

In order to perform the inspection, six task interaction cycles have been identified[16]. These action cycles are used to identify user needs, task scenarios and cognitive walkthrough items for the Inspection. The cycles appear throughout the description of the Inspection method and help guide the usability engineer through the Inspection task. It is very well possible that for one task several task action cycles are possible. This is not disturbing but very helpful in clarifying what can be done with a CVE desktop interface and how.

- System Initiative Cycle, where the user has to deal with the system temporarily taking control over the cause of events in the CVE, either because the user has caused this to happen or because the system has instructions to do so.
- Normal task action Cycle 2D, where a user is interacting with 2D information in the environment in order to achieve a certain goal, such as text menu's or 2D pop-up displays.
- Normal task action Cycle 3D,w here a user is interacting with a 3D object in the environment in order to achieve a certain goal.
- Goal Directed Exploration Cycle, where a user is searching for something known to be in the environment in order to achieve a certain goal.
- Exploratory Browsing Cycle, where a user is navigating through the CVE in order to achieve an understanding of the world layout, or world order.
- Collaboration Cycle, where the user is interacting with other users in the CVE, either to collaborate on a certain task, or to socialize.

## 3.2 Overview of Inspection Method

The inspection method follows the following stages.

- **The User Context Analysis**

An analysis of the user context is required in order to get a good understanding of the specific user needs, which one needs to keep in mind whilst doing the inspection.

- **The Task Scenarios**

The major tasks in the application should be described as they are expected to occur during a typical task performed by representative users, resulting in a scenario description.

- **The Task Analysis**

Tasks, system functions, and 3D objects should be identified from the scenario description, they may be listed as general activities, and more specific activities subsumed under general ones.

- **The Interaction Cycle Analysis**

The tasks, system functions and 3D objects should be linked to the one or more of the 6 interaction cycles.

- **The Cognitive Walkthrough**

For each task, system function and 3D object a number of Inspection questions exists, which are used in the CW.

- **The Report**

For each problem found with a task, system function or 3D object, a note is made and a reference number is assigned (see Part C of this deliverable for more details).

## 3.3 Results and Input to Application Design

The method was carried out on the COVEN On-Line Citizen Application[4] by two inspectors. It provided a more robust and complete classification of the usability problems in the application than the methods used for the first versions of the applications. However, a direct comparison between the two applications is not possible because of the methodological developments. In this inspection we have attempted to rate the problems based on a more complete analysis of the possible actions and tasks of a CVE and to rate the problems in such a way that a comparison can be made between the on-line and final versions of the COVEN demonstrators.

Overall, 52 problems were identified with the application which were classified by their interaction cycle (see Section 3.1). To illustrate the results we example we consider the use of the presentation player in the Rhodes Meeting Room (See Deliverable Part C, Section 7).

| **Function** | **Generic Task Scenario for a Normal Task Action** | **Potential Problems and Design Solutions** | **Severity Rating** | **Ref. Numb** |
|---|---|---|---|---|
| CD Player | | | | |
| | i: Can the user form or remember the task goal? | Not totally obvious that there are presentations, though the controls don invite exploration. Titles could be “Show General Info”. Map on screen could be replaced by an introductory slide. E.g. “Press slide show controls to see presentations”. | 3 | 23 |
| | ii: Can the user specify an intention of what to do? | Once they have discovered the interface then yes. | 1 | 24 |
| | Iii: Are the objects or part of the environment necessary to carry out the task-action (users new intentions) visible? | Yes though the buttons could be larger and labelled. | 1 | 25 |
| | iv: Can the objects necessary for the task action be located? | Yes since they are in front of the presentation screen. | 0 | |
| | v: Can the users approach and orient themselves to the objects so the necessary action can be carried out? | Yes | 0 | |
| | vi: Can the user decide what action to take and how? | Yes | 0 | |
| | vii: Can the user carry out the manipulation or action easily? | Objects are tricky to click since they are quite small. | 1 | 26 |
| | Viii: Is the consequence of the users action visible? | There is a considerable time lag. Immediate audio or text feedback should be given. | 2 | 27 |
| | ix: Can the user interpret the change? | Yes | 0 | |
| | x: Is it made clear to the user what the next correct/needed action could be? | Yes | 0 | |

There are five reported problems here (numbers 23-27). They are not considered too serious and the main issue is the first "Can the user remember or form the task goal?". The perceived problem here is that although the

presentation controls are quite visible it is not obvious what they control. The interface does not invite exploration since the display screen shows a map of Rhodes and thus looks like a poster. The user might even expect the presentation to appear on the control box. This is quite a severe problem, though a possible solution is simply to make the default slide on the presentation screen to be a "Help" or "Instructions Screen" to prompt the user to press the controls. Maybe even a short animation depicting a user pressing the buttons could be produced.

The other issues are that the buttons on the controls are quite small and could be made larger and easier to press, and that the presentation does take some time to start so immediate feedback on the display is required.

Note that the current design of the presentation system is radically different from the previous one. In our opinion the current one is much superior and the changes to be made are simple cosmetic changes rather than re-designs.

## 3.4 Method Development

The method is now undergoing revisions and testing on other applications. One way in which the development is being informed is by structured interviews with CVE application designers so that the results of the inspection can be tailored to their needs. Currently there are few CVE design methods and most development follows the two rules of making the environment simple to keep the frame-rate up but to make it realistic so that real-world metaphors can be applied. As we have seen from our inspections, the latter rule is far too broad. The presentation player metaphor is now less "realistic" but still comprehensible and easier to use. A next step is to take the results of the inspection and develop design guidelines.

A critical area that needs to be addressed is to investigate the design of "flow of interaction" in the virtual environment (see Part C, Section 9.3). Although the Inspection method does address the design of the flow of interaction for *separate* actions, and objects in its current state it does not emphasize the flow of action *between* actions. The perceptual affordances of 3D objects in VEs need to be improved by choosing simplification of the objects so that the available functions on the objects are amplified as much as possible. The sequential affordances of a task that involves interaction with multiple objects in a certain, specific order needs 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 multi-user VEs. However, the use of spatial and temporal cues and audio warnings may be much more effective. Designing the spatial lay-out of rooms and objects more carefully, and providing more carefully designed object affordances could improve the usability of VEs. Simplifying the VE by deliberately designing caricatures of objects and situations may be a more effective way of keeping machine load down, without loosing usability points.

# 4 Summary of Collaborative Actions in CVEs

This section summarizes the results of Part B of this Deliverable. Part B outlines the ideal behaviours of users in a CVE, when collaborating with other users inside that CVE. This theoretical breakdown of the tasks undertaken by the users is used to under-pin the usability inspection (see Section 4and Part C of this Deliverable) and the observational review of the activity in a CVE session (see Section 5 and Part D of this Deliverable).

In order to improve our understanding of the atomic tasks involved in collaboration between multiple users in CVEs, a hierarchical task analysis[6] has been created. This detailed analysis of actions and sub-actions until the level of atomic tasks has been reached, gives us criteria which can drive usability evaluations, it defines a generic cognitive walkthrough task-tree for the Inspection of any CVE. Finally, it helps focus the observational analysis of video-recordings of user actions. When certain user behaviours can not be observed on video-recordings, it could either be because these behaviours are not supported by the application, or it could be because the users do not wish to display such behaviours. It can be difficult to measure the absence of behaviours.

## 4.1 Specific Goal of CVEs

The specific needs of users in a CVE are managing multiple tasks. Depending on the goal of the application the tasks may take a different form, such as teleconferencing, group-wayfinding in a tourist application, information visualisation and searches, etc. However, certain tasks are common to any CVE. These tasks can be described in terms of high level tasks and decomposed into their constituent sub tasks and actions. This will show the overall structure of the main CVE user tasks and a detailed overview of the sub tasks and actions which are specific to collaboration in CVEs. Below the reader will find a break down of generic CVE specific tasks, into the constituent sub tasks until the level of a single action has been reached. These single actions, once identified can be incorporated as user requirements at system level of CVE application design. In order to remain focused on the collaborative aspect of CVE user needs, only the tasks involved in CVE specific collaboration have been broken down to the bottom level of single actions and these tasks and sub tasks are discussed in more detail than other tasks.

## 4.2 Outline of Method

The task decomposition starts with those tasks outlined in Figure 2, and recursively breaks them down until atomic actions are reached. For example the first unfolding of the "Find Objects" task is shown in Figure 3.

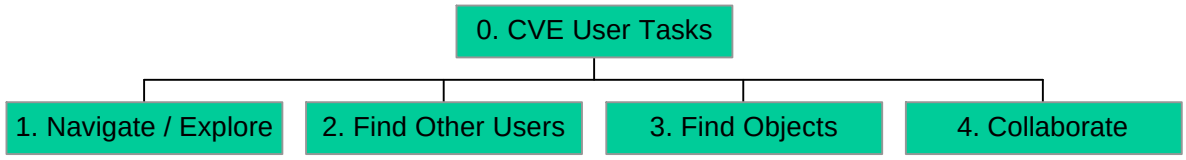



*Figure 2: Root of the task breakdown tree*

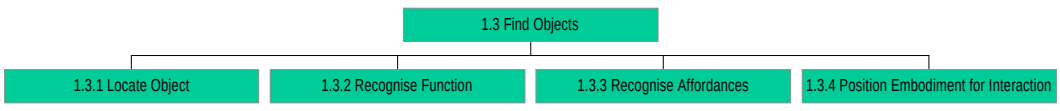



*Figure 3: "Find objects" sub-tree*

# 5 Summary of Theory and Method Description for Spatial and Temporal Activity Charting of CVE Interactions

This section summarizes Part D of this Deliverable. Part D describes the theoretical considerations and the method to score the temporal and spatial activities of user of Collaborative Virtual Environments (CVE). The method explains how to score the interaction of the participants which are recorded over time, using an integration of several theories of collaboration. This method is different from other observation methods in that strenuous efforts are made to clarify the bases upon these inferences are made, to cancel out the effects of value judgments from the observer's own particular point of view, to standardize the process of inference, and to determine whether the operation is reliable, by testing and refining the categories of behavior which are inquiry.

Within COVEN there are two primary uses of this technique:

- To verify that the usability problems identified in the inspection are real issues that need attention.
- To verify that hierarchical task analysis of CVE tasks is complete.

Results from the observations will thus drive the development of the usability evaluation framework.

## 5.1 Overview of the Theory and Method

Interaction process analysis is a method described by Bales (1951) for use in the original observation of small group interaction as it occurs. The heart of the method is a way of classifying direct face-to-face interaction as it takes place, act by act, and a series of ways of summarizing, analyzing the resulting data so that they yield useful information [1]. An observer scores the occurrence of small group behaviors as they occur, based on a system of pre-defined categories. The number of times the categories are scored and the order in which this occurs gives us information about the behavioral patterns of the member of the small group under observation. This method provides a quick way of collecting quantitative data from observations. For COVEN these observations are made from video-recordings of CVE interaction.

The categories to score CVE specific user behavior are being developed based on a review of three sources: use of the method[1] several instances of application of the method to study real-life groups [1][11], users interacting with single user VEs [30] and tele-presence devices [20]. The categories have been derived from actual video-footage of two users collaborating in a CVE (MASSIVE-2), by building up a descriptive account of observable behaviors.

By scoring behaviors as they occur from second to second, and minute to minute we will get an overview of regularities in collaborative behaviors, e.g. sequences of behaviors, transition probabilities of behaviors (i.e. the probability that a particular kind of act will be followed by another, either by the same actor, or by another one). Bales (1951) analyzed several observed sequences of interactional behaviors, including the frequency of various category-to-category sequences, indices of observed to expected sequences between groups of behaviors, and percentages deviation of category/categories from total mean, by sub-periods, etc.

In order to keep the amount of data, and the scoring work to a reasonable size, certain moments of collaboration will be made the main focus of the analysis, initially something such as two minutes of the beginning and end of a collaboration between 2 users, and 2 minutes exactly midway through the collaborative session. The next section briefly describes the categories.

## 5.2 The Categories

The categories which have been defined so far are Communicate, Manipulate, Navigate, Position, and Scan. Each of these categories can be subdivided into a category for the goal, or intention, or result of each action, see Table 2. These actions are to be scored over time.

| **The Categories** | **The Sub-Categories** |
|---|---|
| *Communicate (C)* | Communicate-Explore (C-E) |
| | Communicate-Manipulate (C-M) |
| | Communicate-Position (C-P) |

| | |
|---|---|
| | Communicate-Scan (C-S) |
| | Communicate-Test (C-T) |
| | Communicate-Verify (C-V) |
| *Manipulate (M)* | Manipulate-Communicate (M-C) |
| | Manipulate-Navigate (M-N) |
| | Manipulate-Operate (M-O) |
| | Manipulate-Position (M-P) |
| | Manipulate-Scan (M-S) |
| | Manipulate-Test (M-T) |
| *Navigate (N)* | Navigate-Communicate (N-C) |
| | Navigate-Explore (N-E) |
| | Navigate-Find (N-F) |
| | Navigate-Manipulate (N-M) |
| | Navigate-Position (N-P) |
| | Navigate-Scan (N-S) |
| *Position (P)* | Position-Communicate (P-C) |
| | Position-Explore (P-E) |
| | Position-Manipulate (P-M) |
| | Position-Scan (P-S) |
| | Position-Verify (P-V) |
| *Scan (S)* | Scan-Communicate (S-C) |
| | Scan-Find (S-F) |
| | Scan-Manipulate (S-M) |
| | Scan-Navigate (S-N) |
| | Scan-Position (S-P) |
| | Scan-Sweep A-B, B-A, (S-S) |

*Table 2: The categories of observable CVE behaviours.*

## 5.3 Application and Results

Unfortunately results of the application of this method to video taken from the COVEN network trials were not compiled in time for this Deliverable. Some results will be given in Deliverable 3.6 "Network Assessment of the On-line Applications". The results of the observations will be correlated against the inspection problems (see Section 3.3) and used to test and extend the hierarchical task-analysis for CVE applications (see Section 4).

# 6 Overview of Network Trials

The usability effort takes advantage of the network platform being provided by Activity 3.5 "Network Evaluation of the Online Applications". In this phase the network trials have expanded and diversified, with there being more sites involved in the trials, and both Dive and dVS being used as the CVE system. In addition Activity 3.5 has investigated the use of ATM as well as ISDN and Internet networks, though those trials have been technical in nature and no usability experiments were done.

The network trials phase was March-September 1998, with, roughly speaking, weekly trials alternating between dVS and Dive platforms. In the first few months the trials were going with through a start-up phase with new partners being added, and new software being tested. In the end the set-up phase took rather longer than expected due to software problems and the logistics of trying to co-ordinate trials between so many partners.

Table 1 in Section 2.2 had outlined a trial usability plan. This was designed so as to cover many interesting aspects of the CVE system for evaluation. The trials that were carried out are described in Table 3 and Table 4. In addition the network platform was used for the second and third parts of the Small Group Experiments, see Section 7.

## 6.1 ISDN/dVS Trials

Seven partners took part in the ISDN/dVS trials in this phase. Their locations are shown in Figure 4. The software used was dVISE5.0, and because this does not support audio communication between participants, a separate audio tool (RAT from UCL [15]) was used alongside the main application. In addition a simple text communication client was used for communication between users before the audio and VE were set up.

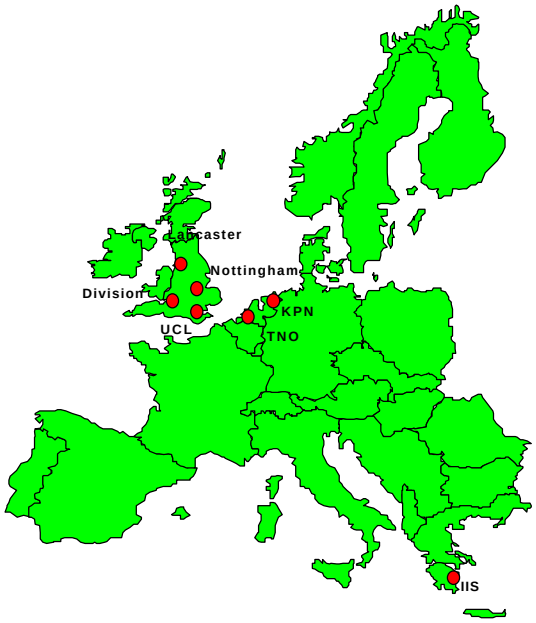



*Figure 4 Location of sites in ISDN/dVS trials*

The activities undertaken in this trials category are given in Table 3.

| Date | Activity | Participants |
|---|---|---|
| 10th June | Bronco | 5 |
| 24th June | Bronco | 7 |
| 8th July | Citizen Familiarisation | 5 |
| 22nd July | Cannibals (Teleport) | 5 |

| | | |
|---|---|---|
| 5th August | Cannibals (Teleport) | 6 |
| 9th September | Treasure Hunt | 5 |
| 23rd September | Treasure Hunt | 7 |

*Table 3 Usability tasks undertaken for the ISDN/dVS trials*

Overall the trials were quite complex to support since there were several problems with dVS's stability with so many participants, and the application's moving to a new platform for these trials. Most of these problems were resolved by the end of the trials and several stable sessions were held as can be seen from the table. Each participant completed a questionnaire similar to that in Figure 5 after each trial. Section 6.3 below reports on general usability results generated from these questions.

1. Did you have any trouble locating the other user(s)? Please explain.
2. Did you have any trouble recognizing the identity of the other user(s), or tell the other users apart?
3. Were the communication channels between you and the other users effective? Please explain.
4. Were the actions of the other user(s) visible and recognizable to you? Please explain.
5. Could you act on a shared object while keeping the other user(s) in view? Please explain.
6. Could you easily switch views between a shared object, other locations/objects of interest and the other user(s) (sweep from one to the other)? Please explain.
7. Could you get a good overview of the total shared space and all other users in it? Please explain.
8. Could you tell when there were interruptions in the attention of the other user(s) to the CVE? Please explain.
9. Did you have any trouble performing the task today? Please explain.

*Figure 5 General usability questions*

### 6.1.1 Bronco Task

This task was performed in the standard Division Bronco demo. The task involved the participants reviewing the design of the car and using built in application functionality to decide upon a colour for the car. In this phase of the usability testing we were interested in establishing that collaboration was possible on this platform.

### 6.1.2 Cannibals (Teleport ) Task

This task was performed primarily in the Rhodes Meeting and Flight zones of the Online Citizen Application. This task exercised the teleportation functionality which had had previously been found to be a non-obvious facility in the environment. It requires a considerable degree of verbal negotiation and synchronized navigation through the environment. The task was described thus:

> *Today's task is Cannibals & Indians. It can be performed by each participant in turn, as the leader, and it needs assistance from 3 other participants as actors in the task, and one camera-person to record the actions.*
>
> *There are 4 characters in this task. 1) the ferryperson, 2) a wolf, 3) a goat, 4) a cabbage.*
>
> *The ferryperson has to take all 3 characters from the VTA to the island Rhodes. But! The ferryperson can not take more than one character at one time.*
>
> *When the wolf is left alone with the goat, the goat will be eaten by the wolf and the game is lost.*
>
> *When the goat is left alone with the cabbage, the goat will eat the cabbage, again the game is lost.*
>
> *Go to the teleporter. Use the group-functionality as often as possible. Do not tell each other the solution.*

### 6.1.3 Treasure Hunt Task

This task was performed primarily in the Lindos zone of the Online Citizen Application. The task was designed to exercise the ability of participants to communicate between themselves with reference to objects in the environment. Thus they had to explain spatial relationships between themselves and the environment and others and the environment. The task was described thus:

> You are Indiana Jones. You have just heard that very valuable archeological treasure can be found on Lindos in Greece. Together with 4 other Indiana Types you have to find these treasures. These are: a shield, a spear, a sword and a vase
>
> When all objects have been found unite the objects every participant has found. When these treasures are brought together every participant gains immortal fame.

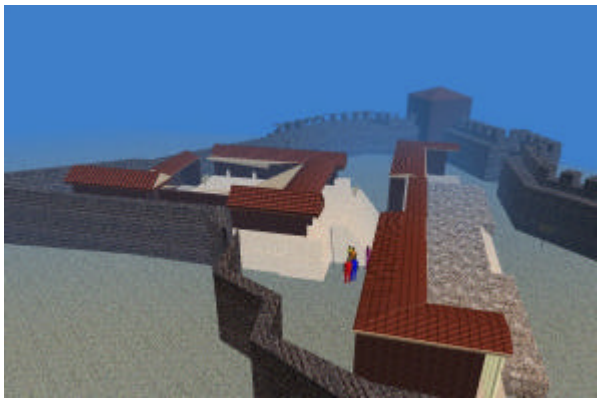


*Figure 6 Overview of the Lindos Zone showing the participants searching for an object*

Figure 7 shows some additional usability questions asked after the treasure hunt task.

1. Did you find an archeological object? If you did, what was it?
2. Could you find the place were every player moved its object to?
3. How did you find out where this place was? (By audio/text/graphical)
4. Could you move the object to this place?
5. What did you think of the overview you had when everybody and all the objects were gathered in one place?
6. What do you think of this experiment in general? What did you like, what did you dislike/miss

*Figure 7 Treasure hunt task questions*

## 6.2 Internet/Dive Trials

The four sites that took part in the Internet/Dive trials are shown in Figure 8. Initially one person per site was engaged in the trials, rising to four per site by the final trial. Standard Internet networking was used, with the Dive software being supported by the used of an application level multicast network, the DiveBone [12].

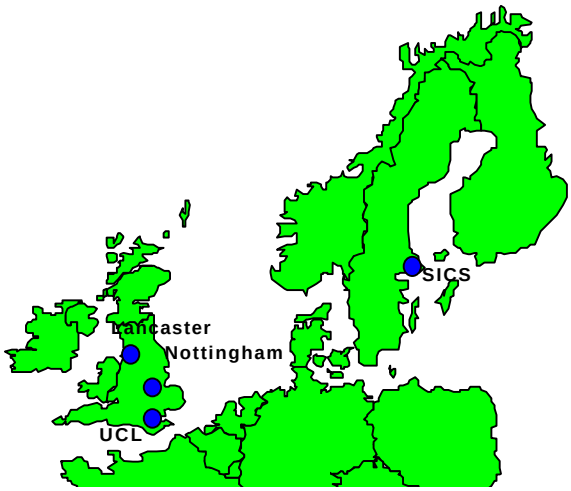



*Figure 8 Location of sites in Internet/Dive trials*

The usability trials that took part are shown in Table 4.

| Date | Activity | Participants |
|---|---|---|
| 01st July | Bowling Game | 4 |
| 15th July | Dive Town | 8 |
| 29th July | Curios World, Office World | 12 |
| 19th August | UCL Model | 12 |
| 26th August | UCL Model / Dive Town treasure hunt | 12 |
| 16th September | WhoDo game | 15 |
| 30th September | WhoDo game | 16 |

*Table 4 Usability tasks undertaken on Internet/Dive platform*

After each trial each participant completed the same usability questionnaire as used for the ISDN/dVS trials (see Figure 5). Dive was designed from the outset to support Internet networking and the setup phase and trials proceeded smoothly. The fact that in the final trial, 16 participants were engrossed in an hour long trial without major incident is a major achievement.

### 6.2.1 Curios World Task

The first focused task was the Curios World task. This was fairly open-ended in that the participants were allowed to collaborate however they wished to solve the task. There were two puzzles in a simple world:

> *Sliding block puzzle: Press "Scramble" button on reverse side of puzzle to randomise puzzle. Selecting one of the pieces slides it into the adjacent vacant spot. Press "Reset" to return the puzzle to it's solved state.*
>
> *Patience: The objects of the game is to remove all but one of the marbles from the board. Pieces move by jumping over one other piece into a vacant space on the other side. Press "Reset" on the back of the puzzle to return to the starting position. To move a piece, first select the piece (the piece will be highlighted red", and then select the vacant square.*

In this task it is important that the users can see what the other participants are doing. This includes being able to see what a participant is looking at, and what they are selecting or interacting with. In Dive these can be told by the position of the avatar, and through the "selection ray" which appears when the participant selects an object.

### 6.2.2 UCL Model World Task

This task was the first large scale testing of the platform. It took place in the UCL model, part of the Online Business Application[5]. The task was simply to navigate around the (large) environment, to test aspects of visibility and interaction in large spaces, see Figure 9. This is a graphically complex world, so it was necessary to test whether interaction was possible with the lower than usual frame-rate.



*Figure 9 The participants rendezvous in the UCL front quadrangle*

### 6.2.3 WhoDo Game Task

This application was designed after the original trials task design (see Section 2.2) in order to provide an interesting task that could involve 16 people for an hour. It exercises many interaction and collaboration styles, since there are aspects of team play, and individual collaboration. The task was explained so:

> *The aim of the game is to guess who killed the "Dead Dude" who is lying the in the central room. Participants play in teams and must collaborate with their team mates and interact with the other teams to be successful.*
>
> *You must guess who did the killing, what item they used to do the murder and where the dasterdly deed took place. There are six suspects in the mansion, six possible weapons and nine rooms.*
>
> *You progress through the game, by moving around the mansion asking questions to players from other teams. They are not obliged to answer you (or even acknowledge), but of course you can set up "deals" where you will only answer a question if they will. Of course if everyone is being obstinate, the game might go on for a while.*
>
> *You may submit a form with your suspicions about the crime at any point for checking, though you don't get a second chance if you get it wrong. That is, if you guess incorrectly, you have immediately lost but you may still answer questions for other people. Of course, if there is more than one in your team, you can try more than one solution.*
>
> *The winners are the team to first guess the correct solution. You may play on to determine rankings or decide to re-start straight away.*

The main tasks in the game were to collaborate with the other teams by asking questions, and to communicate privately with the team members to form a strategy to solve the murder mystery. Figure 10 and Figure 11 give two snapshots of the game in progress.

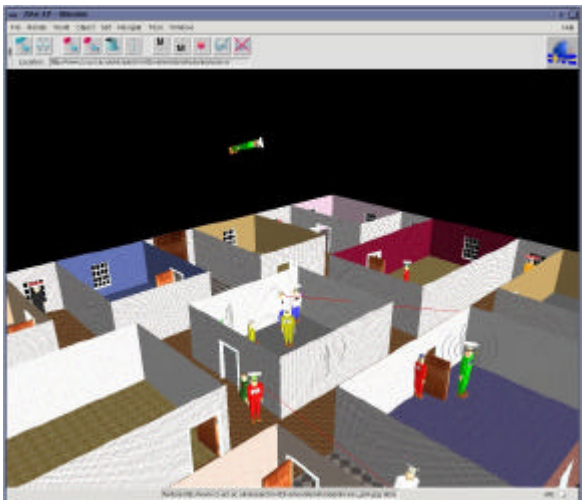


*Figure 10 Overview of the WhoDo Mansion*

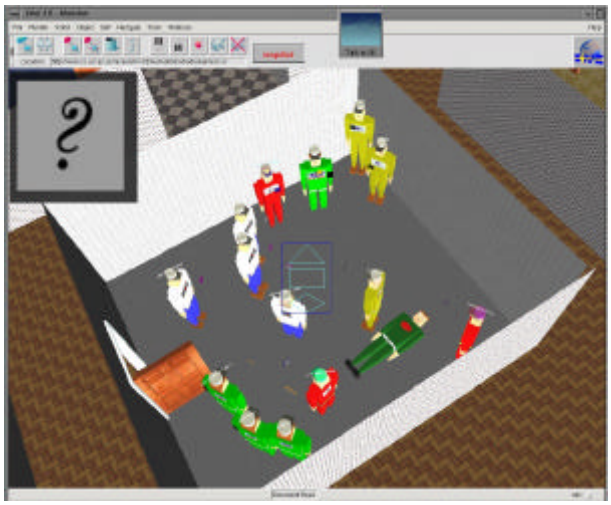


*Figure 11 Participants gathering at the start of a game*

## 6.3 dVS General Collaboration Results

### 6.3.1 Audio and Text Communication

The general impression was that the audio was more reliable in this trials phase, though the perceived quality varied greatly from week to week. Given that most of the sites had experience from the first set of trials, setting up the audio levels and so on took less time. The text tool was changed this time, and although an improvement, the text screen became very cluttered with several participants.

> *Yes, the audio is much better than before although I prefer to use the text chat channel because the message stays on the screen and you can quarantee that everyone will get it.*

One perceived problem that occur with such an audio tool with large numbers of participants is that only one conversation can go on at a time.

> *Not quite - at the start of the session I asked some questions using the chat app [text] without any response. I did not want to use the rat session at the time, because there two people were having another discussion.*

This highlights both the audio congestion and text chat complexity problems. It seems that audio is used to bring notice to text communication that is happening since this is not very noticeable on the screen.

The converse is also true one participant suspects that another has stopped listening to the audio either by forgetting to turn it on or because they have been interrupted by a real world task:

> *Sometimes if you cannot get someone's attention in audio you type a text message, and then they wake up in the audio client.*

One considerable problem that was reported many times was that the audio samples in the application seem to have been produced with a very loud volume so that an application audio event would deafen the participants. The interim solution to this was simply to turn the application audio off since there did not seem to control volume from the application.

### 6.3.2 Interaction and Navigation

A couple of the tasks required the users both to interact with the environment objects and navigate about to get a good overview. This was seen to be quite easy since most users were used to the navigation. If the environment was densely occluded, then it was not as easy to get an overview.

> *The bronco demo has a large enough virtual area for you to be able to pull backwards to be able to see each of the other participants in the virtual environment.*
>
> *Yes: small space, easy turning around, only one object (the car) blocking your view.*
>
> *By panning from side to side I was generally able to see the other users in the world. However there is no feature at the moment to fly to a particular user which might be useful should they become obscured by geometry.*

One re-occurring problem with the interface was that interaction and manipulation events are "detached" from the instigator. That is, it is difficult to associate an object selection to a user since it could be any user who was viewing the object. This user's comment indicates the problem and provides a solution:

> *I could see the results of other user's actions but was usually unable to associate them specifically with another user unless the user was in my field of view. It is not possible to tell who is selecting what item. Perhaps the selection colors could be changed from white to match that of the body doing the selection or picking.*

### 6.3.3 Collaboration

In general it was deemed necessary to have the users identified by colour and name tag. In this series of trials, the name tags were not added until the second half, and we had just reached the stage where there were usually 5 or 6 participants - the point at which remembering all the colours seemed to become too much of a burden:

> *I can keep three or four users apart from color, but when we are in with even I keep forgetting who is red, puprle or green.*
>
> *I had some trouble with this. It is sometimes difficult to remember which colour goes with which person.*

In collaboration situations it important to know which of the participants are currently interacting and which have been interupted. This is difficult since an interupted avatar is "static". And this is independent of whether they are using audio and text. Non-response in audio and text is usually used as the criteria for interruption, not just whether they are moving in the VE or not.

> *No, not immediately, although usually it became noticeable if they stopped responding to questions over the rat system and/or text client questions.*

A possible exception to this rule is suggested by this next quote. If the person is collaborating, you expect them to have the same focus of attention in the VE as yourself:

> *Difficult to tell. I could tell when they were not facing the object I was manipulating, so I assumed they were doing something else. Also, when a person is immobile for a while you begin to think he is doing something else.*

In general communicating spatial locations is a tricky business. Given the lack of a map overview, or audio spatialisation direction must be given explicitly with reference to well-known objects in the environment.

> *Also, when we were searching for objects, e.g. Anthony would say "Hey I found one, I am here!", but since there is no directional sound this did not help at all. What I did then was turn around and hope to see him, or*

*ask where he was. I noticed that people had to explain in more descriptive terms where they were ("I am in front of the temple", etc.)*

*Someone found an object then said (audio) "over here" and after finding out the context for here I was able to join them - often just by running around and looking for a group of users.*

## 6.4 Dive General Collaboration Results

### 6.4.1 Audio and Text Communication

The Dive platform provided two main facilities that were useful in the trials: built in spatialised audio and built-in text communication. The text communication provides for global, private or arbitrary groups of people to be addressed with a single message. The spatialised audio was perceived to be very useful for the task:

*Yes - spatialised audio is great! When someone went far away it sounded as if they were far away. And you could locate people from the sounds they made - i.e. hear in left speaker so look left for them.*

There are problems with the reliability of the audio inherent in the trials' using Internet networks. The first is the audio delay:

*… except that it is tiresome to have to wait one's turn before talking (network delay). It introduce situations when two different speaker think that it is their turn and start talking at the same time. In such situations, there is always a silence, each of them waiting for the other one to restart talking. However, a bit of discipline usually solves the problem.*

The second is audio loss. This is significant since although you could just repeat yourself, one doesn't know if the audio has been lost or if the person in engrossed in another task:

*Nobody seemed to hear me last time. I thought they were either ignoring me completely, or they couldn't hear me. I wasn't sure which one it was, because sometimes they *did* seem to react to me. I gave me a very bad mood!*

Some of these problems can be alleviated simply by making one's utterances longer, by including the name of the person. Indeed this behaviour was quite noticeable in the later trials:

*Even though sound was chopped up at times, in general I was able to detect who was speaking to me (even if you, in general, ended up starting every sentence with the name of the person you were addressing, like "Adrian, do you know if ...", to avoid ambiguities).*

Some users though were confused by the lack of audio "occluding surfaces", a problem that would appear very difficult to solve in the current platform:

*The sound would come out of one room and through the walls, so I was able to listen to what was said by people that I could not see (typically behind a wall).*

This is re-iterated in the next quote. This quote also suggests that private point to point communication would be desirable simply because it screens out other conversations:

*You also want to be able to hold scoped audio conversations, like the text hat supports. It is difficult to do a trade while you hear someone else's trade in the background.*

The text tool was used quite extensively, since although the audio was attenuated over distance, it was possible to eavesdrop on a conversation. However there seems to a problem with cross-media conversations.

*The mix of public audio and private group chats was good.*

*Yes, though it was sometimes confusing as I would ask a question in audio and get a reply in the text chat.*

*The group talk [text] tools was very useful for communication when the team was dispersed.*

The main problem with the text chat session is that one window is opened for every group or private channel. Thus the screen gets very cluttered with text chat boxes. The problem and solution are covered in this comment:

*There were far too many text chat sessions from Dive - you only have so much screen estate. I had to give up on the separate text chat tool while Dive was running cos I couldn't keep everything in shot. Maybe the text wants integrating into the bottom of the main Dive window rather than being in a popup on it's own?*

### 6.4.2 Interaction and Navigation

The main difference between Dive and dVS in interaction is that in Dive, selection, manipulation and mouse navigation are all accompanied by a "focus line". This is a red line drawn between the avatar initiating the interaction and the subject of the interaction. This facility alleviates some interaction problems which we would have expected through our experience with the dVS trials:

> *Yes, when I was actually watching them, I could see their manipulation focus line (i.e. a red line that is drawn between the avatar and the manipulated/interacted object, each time an object is interacted.*
>
> *Regarding actions "in the world": Yes. I could see when somebody clicked on something (the red ray), but sometimes you could not tell who was doing the clicking, since the ray originated from a point outside of your view.*

Although the navigation techniques in Dive and dVS are different, they seem to be equally usable. One problem that occurred more often in the Dive worlds than the dVS ones was problems associated with collision detection with the walls. Two of the Dive worlds were quite densely populated so collision was occurred more often.

> *As long as you navigated yourself in to the correct position to be able to keep your eye on someone, it was ok. Within the office world it was much harder, as I found myself often walking through the walls and loosing sight of most things.*

Given the difficulty in navigation through dense environments one suggestion is to make the navigation more automatic with some constraints imposed by the system. This quote indicates that navigation itself imposes quite a significant workload on the user, thus they can't give their complete attention to the task.

> *Maybe the system could take care (more than now) of navigating the avatar letting the user concentrate on other things.*

### 6.4.3 Collaboration

In general it was found quite easy to identify the different users in the virtual world. Each site chose a main colour for their avatar, and each person had a name tag. In this it was similar to the identification used towards the end of the dVS trials. However one item made it easier to identify people - the audio "waves" that emanate from the user when they speak.

> *Yes, even having two people in the same color, by looking at the 'waves' it was possible to 'recognize' who is who and who is talking.*

However the use of inappropriate names in one session hindered interaction although the names did distinguish the users:

> *Yes - all the ucl people had the same or very similar names above their bodies. I was nervous to go up to them because of this (just like being at party when you don't know anyone)*

One suggestion for future trials was that teams would have share one similar coloured garment but would be distinguished by different colours on other garments as well:

> *…(e.g. everybody from SICS would have blue trousers, but the top of the body would be painted according to the personal taste of the person behind the body icon).*

With such large groups though, getting an overview is hard since the group formed covers quite a large area in the virtual environment, and the avatars in a group can obscure lines of sights between members of the group quite easily. A explicit user command "Look at …" is provided in the standard interface:

> *When i a big group as this time (8 people) it is hard to have a good view of everyone at the same time. If you move away people get tiny and are hard to hear.*
>
> *I used the "Look At" command on a couple of occasions to locate new people*

A couple of the usability questions deal with the issue of participants being interrupted in the real world when participating in the virtual environment. Indeed it can be quite difficult to tell if an avatar is active or not:

> *No its very hard as people is not interacting all the time, sometimes they just stand still and observe or they might have left their computer....*
>
> *We picked up Jolanda at one point and there was no reaction - therefore she wasn't there!*

If a participant knows that they will be absent for a while, then they can give hints to the other users via audio or text. A novel approach to this is indicated by the following quote (the person laid his avatar horizontally on the floor):

*One person (I think it was Adrian) got interrupted and laid down his body icon to mark this. I did the same once. I felt that this was a very good idea, because you would know that he was "off" the environment.*

Sometimes it was still found to be quite difficult to establish interaction. If one avatar approaches another with the intention of interacting, there is not way for the "approached" avatar to discourage the interaction through body language. If they want to turn away they would have to break away from some other interaction they were involved in. Indeed the actions that people do make with their avatars take on quite complex meaning:

*When somebody seems 'alive' and slowly moves away from me, maybe even passing stright through me, I do get rather annoyed - that seems like a deliberate ignoring. If they suddenly moved away that would suggest they were not in control of their movements.*

*Nope, and i don't think it was possible to even tell whether you ever had their attention. I would start a conversation with someone and before you know it they had vanished.*

*There seemed no sensible way to meet someone and ask questions. It was hard to get someone's attention and even then they either didn't hear you or wandered off before you could talk to them.*

## 6.5 Conclusions

Progress had been made in supporting groups of users in collaborative tasks. The Dive trials have solutions to some issues, such as audio and interaction indication. Other problems have been highlighted, and one of the considerations of the final implementation phase should be on how to best present underlying functionality that does exist (such as group text communication) to the user.

One pressing issue that was raised is how person to person interaction can be established in the absence of immediate facial or body feature feedback.

In general though, it is certainly possible to collaborate in the environment despite the relative lack of richness in the environment. Figure 12 illustrates this quite well. All the participants are "looking away" whilst another participant hides an object in the world.

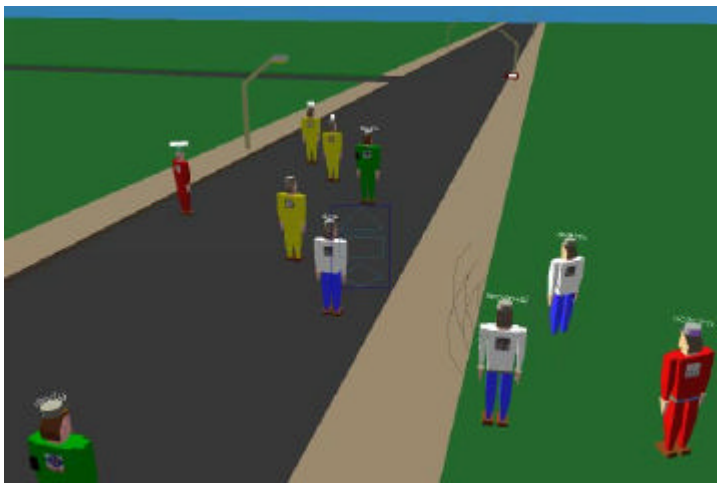


*Figure 12 Emergent collaborative behaviour*

# 7 Summary of Small-Group Experiments

## 7.1 Introduction

This chapter describes a series of experiments that were designed to investigate what happens when a small group of strangers meet together to carry out a joint task in a virtual environment. The issues of interest are:

- The relationship between co-presence and presence;
- The relationship between both types of presence and immersion;
- The relationship between both types of presence and accord within the group;
- The extent to which leadership is conferred by immersion.

Three experiments were carried out. The first was a local trial at UCL with three participants, one of whom was immersed. The second was a small-scale trial carried out over the COVEN network involving groups of three participants each using a desktop system. The third, the main trial reported in this Section, again used a network, with one immersed person.

## 7.2 Experimental Framework

All the small group experiments have used the same task scenario but each experiment has had different isplay, interaction and network conditions. The initial experiment made a preliminary exploration of the questions raised in the introduction. It used a local-area network and the subjects performed a task twice, once in the real environment and once in the virtual environment. The second experiment proved the use of the COVEN network platform and demonstrated the feasibility of carrying out an experiment over a wide-area network. The third experiment used the COVEN network again, but re-introduced one immersed person.

### 7.2.1 Task Scenario

Groups of three strangers meet for the first time in an office environment. The subjects have avatars coloured Red, Green or Blue, and refer to themselves and the others by colour throughout the experiment. They meet in a small room, and have to find their way together to another room that has a series of puzzles written on pieces of paper stuck around the walls. On each piece of paper is a set of words or phrases each prefixed by a number. The subjects have to rearrange all the words corresponding to the same number in order to form a saying such as “A critic is a man who knows the way but can’t drive a car”. There are twelve puzzles on the posters and there is a time limit of 15 minutes in which to complete as many as possible.

This task was chosen since is requires a good deal of collaboration in order to be successful. It is not easy for one person to remember all the words needed to form a sentence since the words are displayed in random order. The task can be easily partitioned if each subject takes a different area of the room. Thus the task requires a lot of audio communication, and the inclusion of avatars makes the task easier since participants can make reference to posters by looking at those posters.

### 7.2.2 First Experiment

The first experiment was an exploratory study to generate hypotheses about how people would conduct themselves in such a highly collaborative task[38]. For this purpose, each trio carried out the scenario twice, once in the real environment, and once in the virtual environment which was a copy of the real environment. For the virtual version, one participant was immersed, and the two others were using desktop machines. The scenario was implemented in Dive3.2 software. There were 10 groups, all of which did the virtual task followed by the real task.

### 7.2.3 Second Experiment

The second experiment proved the feasibility of carrying out an experiment over a wide-area network [44]. 4 groups completed the task that involved the virtual task only. Each participant used similar equipment and each used a desktop interaction metaphor. The scenario was implemented in the Dive3.2 software.

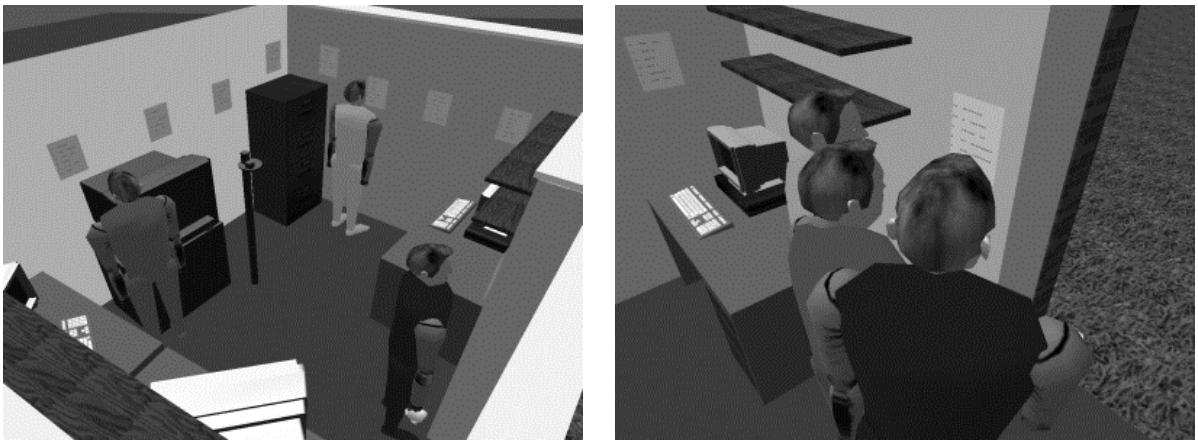


*Figure 13: (a) Overview of the task room with Red, Green and Blue participants examining different walls (b) Red, Green and Blue participants examining one of the posters*

## 7.3 Experimental Design

### 7.3.1 Overview

Given the distributed nature of the experiment, a rigorously timed schedule was prepared and carried out. The participants brought along a personality questionnaire that they had completed beforehand. The participants then spent 4 or 5 minutes training to navigate using the desktop or immersive system. Once this was completed, the task was explained and the participants were told their colour instructed to refer to each other by these colours.

Once each site had reached this stage, the multi-user session was started in co-ordination between the sites. The participants entered the environment and were left to introduce themselves and complete the task. Note that the subjects are not told how to solve the puzzles, just that there are puzzles which they have to solve. Thus after introductions a few minutes was usually spent deciding what the puzzles were and how to solve them.

Once the session time was up, the experimenters (who had been communicating with a text chat tool), would give the participants the questionnaire described in the next section and then de-brief the subjects in a short interview.

### 7.3.2 Post-trial Questionnaire

The questionnaire was used to elicit information regarding the areas of interest outlined in the introduction: the relationship between co-presence and presence, the relationship between both types of presence and immersion, the relationship between both types of presence and accord within the group and the extent to which leadership is conferred by immersion.

Self-reported presence was assessed by six questions each on a 7-point scale, with the higher score indicating higher presence. The questions (paraphrased here) required a response to the extent to which:

1. there was a sense of being in that room which has pieces of paper with the riddles;
2. there were times during the experience when the virtual place became the reality and the real world of the office in which the whole experience was really taking place was forgotten;
3. that the virtual place is thought of as somewhere visited rather than just images that were seen;
4. the sense of being in the virtual place was stronger than of being in the real world of the office;
5. the structure of memory of the virtual place is similar to the structure of memory of other real places (in terms of the visual memory, whether the memory is in colour, the vividness, realism, size in imagination, panoramic, and other such structural elements);
6. the virtual experience was overwhelming (so that they did not often remember that they were really in an office using a computer).

In line with previous experiments where these questions have been employed the overall measure was conservatively taken as a count of how many scores of ‘6’ or ‘7’ were given out of the six questions.

Read each item carefully and decide the degree to which the statement is characteristic or true of you. Then place a number from 1 to 5 in the correct space according to the following scale.

1 = the statement is not at all characteristic of me.

2 = the statement is slightly characteristic of me.

3 = the statement is moderately characteristic of me.

4 = the statement is very characteristic of me.

5 = the statement is extremely characteristic of me.

| I .... | Decide 1-5 |
|---|---|
| 1. often feel nervous even in casual get-togethers. | |
| 2. usually feel uncomfortable when in a group of strangers. | |
| 3. am usually not at ease when speaking to a member of the opposite gender. | |
| 4. get nervous when talking to a teacher or boss. | |
| 5. often feel anxious and uncomfortable at parties. | |

*Figure 14: Questions to assess social-phobic response*

Reported co-presence was assessed by eight questions variations on the theme of being with the other people. These were again rated on a 7 point scale, and required a response to the extent to which:

1. there was a sense of being with the other two people;
2. the computer interface seemed to vanish and there was direct working with the other people;
3. the experience was more like working with other people rather than interacting with a computer;
4. other people were forgotten and concentration was on the task as if the individual was the only one involved;
5. the experience was like some other real experience of working together with people to solve a problem;
6. a sense of being with other people rather than just experiencing computer images;
7. the experience resembled being together with others in a real-world setting;
8. there was a sense of other human beings interacting with the individual rather than just a machine.

Overall co-presence was measured as the number of ‘6’ or ‘7’ scores out of the eight questions.

Accord was assessed by a series of questions relating to enjoyment, harmony and cooperation of the group. These questions were rated on a 7 point scale, and concerned the extent to which:

1. the experience was enjoyed in a similar manner to a previous real meeting that was enjoyable;
2. there was a desire to meet either of the other two people again;
3. the group was in harmony during the performance of the task;
4. the person felt comfortable with each of the other two people;
5. there was a desire for the group to form again;
6. the other two people were cooperative;
7. the other two people made the individual feel embarrassed or self-conscious.

Overall accord was measured as the average response across all of these questions.

There were two questions that related to leadership, one directly and the other indirectly. Each subject was asked to score all three subjects on the degree to which that person "was the 'leader' or main organiser" in the meeting that had just concluded. The three scores, one for Red, Green and Blue had to add to 100. In addition, there was a similar question concerning who did most of the talking. An overall leadership score was constructed for each individual by taking the average of the assessments of the other two members of the group. A similar method was used for 'talkativeness'.

In order to take into account the impact of personality on the results there was a final question that assessed the extent to which the individual generally experienced social-phobic responses in everyday life. This employed the Interaction Anxiousness Scale [23] where fifteen statements are given, and the subject has to respond in a range of 1-5, where the higher score corresponds to a higher social-phobic response. For example, the first five statements and the possible responses are shown in Figure 14.

## 7.4 Technical Description

### 7.4.1 Platform

We used the dVS/dVISE 5.0 system from Division Ltd[7] to implement the scenario. While dVS supports audio from the application to the participants, there is as yet no direct support for audio communication between participants, hence the Robust-Audio Tool (RAT) v.3.0.23 [15] was used alongside the main application.

The model consisted of approximately 11,000 polygons and ran at 20-30Hz on all client machines.

### 7.4.2 Workstation Configuration

Four main workstations were involved in the trials, with auxiliary machines used to display the experimenters' chat sessions.

The participant machine at UCL was a Silicon Graphics Onyx with twin 196 MHz R10000 processors, Infinite Reality graphics and 192M of main memory.

The participant machine at Nottingham was a Silicon Graphics Indigo$^2$ with a 200MHz R4400 processor, High Impact graphics and 192M of main memory.

The participant machine at IIS was a Silicon Graphics Octane with a 195MHz R10000 and 128M main memory.

The final machine was the dVS server and audio server at UCL. This was a Silicon Graphics Inidigo$^2$ with a 200MHz R4400 processor, High Impact graphics and 64M of main memory.

The participant at UCL was using an immersive system employing a 2 tracker Polhemus Fastrak system, Virtual Research VR4 helmet and a 3D mouse with 5 buttons. The helmet has a resolution of 742x239 pixels or 170,660 colour elements for each eye and a field-of-view of 67$^o$ diagonal at 85% overlap. The only actions the participants had to learn was how to navigate about the environment, by turning their bodies and pressing buttons on the 3D mouse to move them forward or backward in the direction in which they were looking.

The other two participants were using a desktop system using the full 1280x1024 screen and a standard 2D mouse with three buttons. The participants navigated about the environment by pressing the middle mouse button and dragging left and right for rotation about a vertical axis, and forward and backward for translation in the direction in which they were facing.

### 7.4.3 Network Configuration and Logging

The experiment was carried out over a network used for weekly trials between the COVEN partners. Each site was connected over ISDN to UCL.

The dVS master server ran on romulus, each client downloads the initial state from the server and then peer to peer connections are made between the sites for the delivery of events. An audio server ran on romulus and sites connected using the point to point connection capability in RAT.

Network logging was performed at each site using the tcpdump command, a filter being used to capture only relevant traffic. For each session exhaustive pair-wise comparisons of the tcpdump files for all three sites were made. For each packet from the first file a match was made to a packet from the second file, and similarly a return packet from the second file shortly after the original packet was matched to the first file. This enabled the time differences between these packets to be computed for both files, and so generate the round trip time independently of the local clock times (there is no guarantee that the clocks are synchronized between sites). The sample size for each analysis is the number of such calculations made per pair of tcpdump files. The mean round trip time was then determined for each session, together with the standard deviation for the round trip times.

Logs were also taken from dVS that recorded all movements and actions within a session. This allows us to play back the VR session and observe from any vantage point.

## 7.5 Results of the Statistical Analysis

Fourteen subjects were eliminated from the questionnaire results because of incomplete questionnaires. Of these eight were the Green subjects, located in Greece. There is obviously considerable difficulty in conducting a study across people with different native languages, and this is probably reflected in the relatively high proportion of non-native English speakers with incomplete questionnaires.

### 7.5.1 Co-presence and presence

As found in the previous experiments, though with more confidence in this experiment because of the greater variety of questions, and greater attempt to clearly distinguish the two concepts, there was a significant positive correlation between the two. This relationship is practically important whatever the reason for the correlation. If one influences the other, or if they are both independently caused by the same underlying factors, it is useful to know that benefiting one may positively influence the other.

### 7.5.2 Presence and immersion

There was no reported significant difference between reported presence and co-presence between the immersed and non-immersed subjects. Again this is in line with the previous results.

### 7.5.3 Presence and Accord

As found previously, individual accord is positively associated with both types of presence - the more that the experience was enjoyable, the group in harmony, the greater the degree of presence and also the greater the degree of co-presence.

### 7.5.4 Leadership and Immersion

In the first study there was a very clear relationship between leadership and immersion - the immersed subjects tended to have the advantage in the virtual session, and this advantage was not reported in the corresponding real session. In the second study there were no immersed subjects, and no special pattern of leadership emerged.

In this study the situation is more complex, and still there is no answer that can be given with confidence to this question. The Green subject (in Greece) was almost never the leader. There is strong evidence that the subjects in Greece experienced the dual problem of language difficulties and also having the poorest audio response. Recall also that eight of the twenty Green subjects had been eliminated from the questionnaire data.

There was a very high correlation, as found in the earlier studies, between leadership rating and degree of talkativeness ($R^2 = 0.82$). In fact 85% of the variation in overall leadership score can be accounted for by the degree of 'talkativeness' (positive correlation) and the extent of social phobia (negative correlation): independently of any other factors, the more talkative and the less social discomfort a person generally experiences, the more likely they were to have been rated as leader, which is to be expected in real-life situations and is reproduced in this virtual meeting.

Leadership and degree of talkativeness may be considered as different aspects of one overall leadership variable, computed as the average of the two. Using this as a response in a regression analysis, 40% of the variation in the response can be accounted for by the 'colour', gender and social-discomfort as explanatory variables. The statistically significant relationships are that the Green subjects were significantly less likely to be overall leader than the other two. There is no significant difference between the Red (non-immersed) or Blue subjects. Female Blue subjects were significantly less likely to be leaders. For Red subjects only, a higher degree of 'social phobia' is correlated with a lower overall leadership rating. Table 5 summarizes these relationships - the actual numerical values are less relevant than the type of relationship indicated. It is clear that whatever gender, the Green individuals are significantly less likely to be overall leader. In the Red group there is a negative correlation with the social_phobic response. In the Blue group, females were more likely to emerge as leader, other factors being equal. However, the relationship is complex, and clearly more studies are required in order to throw further light on this. The question of whether immersion confers leadership advantage remains very much open.

*Table 5: Regression of Overall Leadership (0-100 score) on Gender, Social Phobic Response and Individual Colour $R^2 = 0.40$*

| | Male | Female |
|---|---|---|

| | | |
|---|---|---|
| Red | 51 - 0.7* social_phobic | 33 - 0.7* social_phobic |
| *Green* | *19* | *12 (NS)* |
| Blue | 27 | 41 |

## 7.6 Discussion

Participants in a CVE have to establish and maintain a set of social norms or etiquette of social conduct during the interaction, using a much lower bandwidth than available in real world settings. In order to establish and confirm ones perceived trustworthiness as a competent computer-mediated collaborator, a participant has to be perceived by the other participants as acting according to the social norms [35][32]

There are several relevant social norms for the establishment of cooperative attitudes, which have been found to govern computer mediated interactions aswell. Firstly, phatic communication, (i.e. the exchange of stereotyped phrases and commonplace remarks to establish and maintain a feeling of social solidarity and well-being) is important during three major phases of the interaction: the opening phase, the medial phase, and the closing phase [22]. Secondly, spatial regulation (i.e. the arrangement of single or combined average body-size related spaces around and between people and objects, signifying temporary or permanent micro-territories; where each cultural tradition has its own micro-territorial sizes and arrangements) governs the judgment about the appropriateness and our competence at navigation and positioning of ourselves in shared spaces [36]. Thirdly, proxemic shifts (i.e. patterns of interpersonal distance in face-to-face encounters) accompany and influence changes in the topic or in the social relationship between speakers (i.e. situational shifts). Fourthly, co-verbal behavior (i.e. non-verbal communication accompanying verbal communication) has an important role in the understanding of social interaction. The most fruitful co-verbal behaviors for understanding and observing social interaction are head nodding, face looking, smiling, head touching, and speaking, including simultaneous speech [26]. Finally, there are clearly defined cues that govern turn-taking in a conversation. An auditor may claim the turn when the current speaker displays a turn-signal. A turn-signal is composed of a set of six behavioral cues, found variously in intonation, context, syntax, paralanguage, and body-motion [8]. Taking turns in a conversation and the alternation of action and inaction, and the subsequent rotation of performance among individuals in a group is the most salient feature of group dynamics [26].

In CVEs, participants are only aware of each others voice and virtual embodiments, and the co-verbal cues which can be observed from these. The interpretation and display of proxemic shifts, and the observation of the spatial regulation rules are influenced by the ability to fine-tune navigation and positioning of ones virtual embodiment. Conversation, turn taking, and the display and perception of co-verbal cues are influenced by the performance and reliability of the audio-channel. And last but not least, the success of the collaboration is heavily influenced by the success of the phatic communication, which in its turn is dependent on the performance and perceived reliability of the audio-channel and the participants' ability to coordinate their embodiments. A rich CVE interface, with effective virtual embodiments and easy to use navigation controls should help participants to reach what Goffman (1959) has called the “working consensus” of the interaction, especially in those situations where the role structure is not obvious to the participants.

Novice participants bring their expectations of social norms from real world interaction to the CVE. Because of the inherent problems and difficulties of the unorthodox CVE interface and technology, participants have to rapidly adapt their expectations to the limitations of the situation. However, it is difficult - especially for novice participants - to judge which particular aspects of the CVE interactions can or must be attributed to limitations of the technology, and which to a participant and their possible display of a lack of respect for the social norms, and thus their trustworthiness.

There are several stages during the CVE interaction at which the success of the collaboration is at risk (see Figure 15). In order for the participants to be able to acknowledge each others presence, they have to be able to either hear or see each other. In order to be able to fine-tune their position in the virtual space, they have to be reasonably competent at navigation. In order to be able to contribute to the meta-planning of the collaborative task, they have to be relatively undistracted by concerns of audibility, visibility, and sociability. In order to collaborate on the activity at hand, they have to be certain that they are audible and visible to the other participants when they think they are and vice versa.

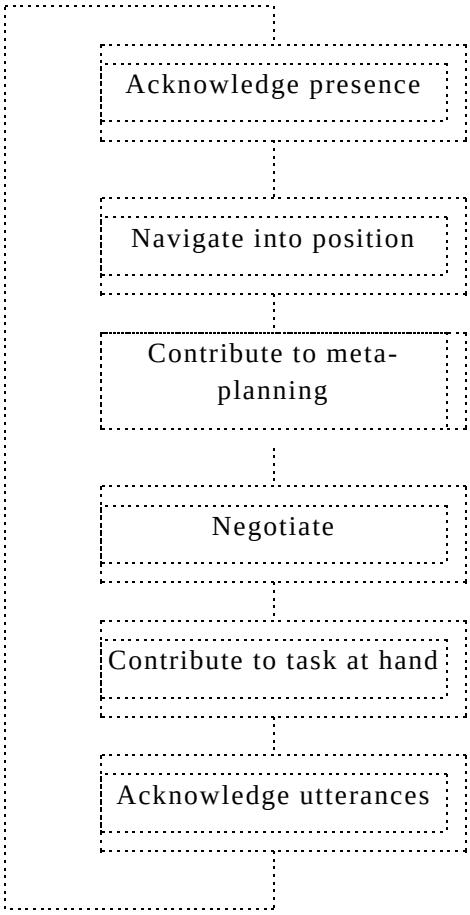



*Figure 15 Interaction flow in a (CVE) collaboration task*

### 7.6.1 Example from Video-Data

Blue and Red want to give up working with Green, because Green's sound is judged too unreliable for collaboration purposes.

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?

### 7.6.2 Solutions

Navigation and fine-tuned positioning can be improved by creating 'trenches' that automatically bring a participant to the optimal path or viewing distance. All objects have and optimal viewing distance associated with them, which increases with the number of participants trying to access the same area. Participants have an optimal collaboration distance associated to their embodiments. The autopropelling and autotracking properties of the furrow or trenches and personal space distancers propel the participants to their goal automatically. Participants can interrupt this automatic movement by a proportionally greater co-movement from their input device. With long intensity user-control participants are simply gravitating inside the gravity wells of the surrounding others and objects; like a slow-motion pinball in a pinball-machine.

Field of view can be subjectively enlarged by stretching the outside edge of the workspace in the outward direction only. By making the walls unpenetrable but at the same time flexible to pressures in an outward direction, the participant can increase their field of view until they can encompass the number of other participants and relevant shared objects in one view.

Automatic behaviours to initiate phatic communication and signal turn-taking could be a sequence of automatic actions which occur the moment a participant enters the VE. The avatar could automatically search for the optimal place in the VE space where the participant will have the best view of the other participants. If this automatic sequence is uninterrupted by the participant the avatar could start waving automatically, go up to the nearest other active avatar and smile, when still uninterrupted it could go to the next nearest other avatar, etc. An utterance from a speaker could be automatically accompanied by slow nodding of the head, smiling, slowly turning the head from one side of the view to the other, etc. Participants intending to collaborate could make their avatars automatically assume positions oriented at an angle of about ninety degrees, where they may turn their heads to interact in face-to-face relation. Or the avatars could automatically form small triangles, squares, or circles, depending on the number of participants in the group.

## 7.7 Conclusions

There was much additional data collected from network logging, audio and video, and post-experimental de-briefings of the subjects. This remains to be analyzed, part of it will be reported in Del3.6. The results of the statistical study based on the questionnaire responses offers further support for the relationship between co-presence and presence, between co-presence and group accord. The question of immersion and leadership is still very much an open one, though there is evidence here supporting the notion that immersion confers leadership advantage.

# 8 Overview of Auxiliary Trials

## 8.1 Introduction

The purpose of the auxiliary trials is to identify factors that might determine the usability of VE technology and to develop techniques for the evaluation of VE systems and applications. They are more speculative that the focused trials, such as the small group experiments, but are important since CVE are being built on technologies that badly little understood from a human factors point of view. We report upon two recent strands of work which have given us a number of criteria for designing the user-interface for future CVEs.

The first experiment [39][41] shows that the metaphors behind user-interface decisions can have a significant impact upon presence and thus task performance in virtual environments. In particular the experiment show how the method of button pressing in a VE affects the sense of presence. The first experiment also introduces a new method to measure presence. This method is still under development at UCL, but has already been used at KPN and a number of other institutions as part of their experimental method.

The second set of experiments [27] looks at how visual attention varies through the depth plane. This is important for VE research since it shows how the visual system seems to be more responsive to objects further away but of the same retinal image as those close by. Certainly this should inform applications which rely upon "heads up display" type approaches for information display.

## 8.2 Breaks In Presence Experiments

Whether a participant feels as if there are present in the virtual environment is fundamental to many VE applications. We would argue that before certain types of situational awareness are possible the participant must first feel as if they are in the virtual environment being presented. Both objective and subjective measures of the "sense of being in a place" have previously been used to assess virtual environment (VE) systems [10][18][37][40] though they have limitations. Many objective measures, such as testing fear reaction to looming objects, interfere with the task being undertaken and thus can only be used in controlled experimental situations. Subjective measures such as questionnaires can not be administered during the experience without taking the participant "out" of the experience. We wish to be able to assess presence continuously during the experience through a non-invasive method.

We have noted in several previous experiments that there notable are events that we call "breaks in presence" that occur during the experiment. Examples are the subject hitting their hand on the tracking equipment, witnessing parts of the virtual body passing through other objects, or failing to comprehend the effects of an interaction metaphor. Such occurrences are often accompanied by a question to the experimenter such as "Where am I now?" which indicate that they have lost a continuous sense of presence in the environment. In fact it is our belief that during such moments the participant is present in the real world rather than the virtual world. This is similar to the effect of being brought back to the real world during a movie when disturbed by a loud noise or some such event. The number of frequency of such events can tell us much about the periods when it was possible for the person to be present in the virtual.

### 8.2.1 The Method

The notion of a person's presence switching between two states is based upon the study of ambiguous figures in gestalt psychology [24]. A person's perception of a figure can switch between two states and these switches are quite noticeable by the person. In a similar manner we know that participants notice the switches between virtual and real environment, though not the switches from real to virtual.

We give the subjects two tasks, one for the virtual environment, and one for the real environment. The real world task is simply to report that they are in the real world. The virtual task is the task which the virtual environment is designed for. That is, there are no extra instructions to respond to events within the virtual world.

We log each of the participant's statements that they are in the real world (a break in presence or BIP). From this we can construct a measure of overall presence by first noticing that a person who never reports a BIP was either present all the time, or none of the time. The mapping between number of BIPS and presence score is shown in Figure 1.

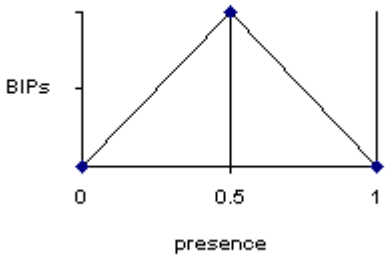



*Figure 16 Mapping between Presence and Number of BIPS*

Since the mapping form BIPS to presence is not single valued, a single discriminator is required to determine whether they or present more often than not in the environment. We use a value determined by subjective, reported, sense of presence as taken from a post-experience questionnaire.

### 8.2.2 An Experiment

We have used the BIPs method in experimental situations and have found a strong correlation between this new method and the results of our standard presence questionnaire.

As an example, we performed an experiment to look at the effect of forcing hand-object interactions in the virtual world, rather than using a button on a 3D joystick to interact with the world. Figure 2 shows a virtual environment that consists of a copy of our laboratory with an adjacent virtual field with a 3D chess board. The participant starts in virtual laboratory section of the environment and is trained by the experimenters to use the equipment. They then make their way outside in to the virtual field where they have to observe and remember a set of moves on a 3D chess set (see stereo pair image in Figure 3).

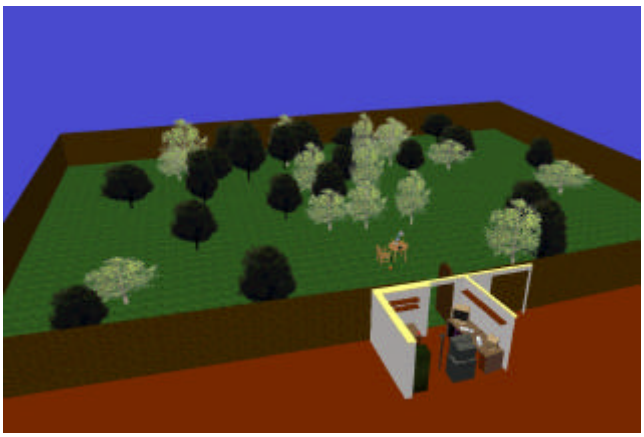


*Figure 17 Overview of the experimental environment*



*Figure 18 Stereo pair showing the 3D chess set*

The sequence of chess moves was either triggered by the participant touching the next piece to move with their virtual hand, or pressing a button on the 3D mouse. The environment was implemented using Division's dVISE4.0 on a Silicon Graphics Onyx with Infinite Reality Graphics. The display was a Virtual Research VR4 helmet, and the person was tracked using a pair of Fastraks.

The results showed a high degree of correlation between the amount of hand movement the person made and the sense of presence. The reasoning is that body movement exercises the participant's model of how their body and the environment interact and reinforces the correspondence between visual and proprioception information.

### 8.2.3 Discussion

The presence counter method is obviously a useful tool since it not only gives us an overall measure of presence, but the BIPs themselves indicate aspects of the experience that should be fixed. Most BIPs occur because of some failure of consistency in the virtual world, or a conflict between real and virtual world cues. However they can arise for other reasons such as boredom or frustration. The following list contains a few reasons for BIPs that were reported in a short post-experience interview:

> "I was supposed to be in a grass plane, but when I moved my feet I realized it was a plank under my feet (in the real)."
> "Weird things happening that are obviously not real (e.g., the chess set)."
> "Embarrassment."
> "Attention wandered after realizing that the chess sequence was iterative."

The BIPs themselves give us a large number of elements of the experience to work on. Some of them as certainly tractable, such as the behaviour of the virtual chess set where pieces would levitate, though overcoming the participant's embarrassment is a little harder.

## 8.3 Visual Attention Experiments

The purpose of these experiment was to investigate how visual attention differs between proximal (viewer-centered) and distal (environment-centered) space. Essentially the experiments attempt to show that the reaction times to cues at different depth can be affected by the current depth of visual attention. Put simply, this means that participants react more quickly to cues at the same depth to where they are visually attending, than cues at a different depth, despite the fact that the cues are the same apparent size and intensity. This is a 3D depth equivalent of the Posner effect[33]. The Posner effect is observable in reaction times to cues on different sides of the visual field. A subject is asked to respond as quickly as possible to a visual target's appearance by pressing a button. The subject is instructed to look directly ahead, and just before the target appears, a cue appears either on the same side or the opposite side to where the target will appear. The participant will be slower in reacting to the target if the cue is on the opposite side no matter how much they train. It appears that the visual system is distracted and thus attention has a "focus" of interest. Our experiments investigate whether attention can be distracted in depth as well as side, as has been proposed[29] but not conclusively shown to date.

### 8.3.1 Experiment 1: Egocentric/Allocentric Comparison

The purpose of the first experiment was two-fold. Firstly it was necessary to show that despite the technological limitations, it was possible to observe the Posner effect through the use of an immersive virtual environment system. Secondly that there was no difference in reaction time to cues that were fixed relative to the observer (that is head-slaved) or fixed relative to the world. This was important since for the purposes of further experiment it is necessary to make the cues head slaved, whereas most visual events are environment centered, not user centered. Details of the experimental procedure can be found in[27], we only briefly outline them here.

The immersed participant must react to a target (a red ball) appearing in the virtual environment on either the left or right of their visual field. 90 ms before the ball appears, a cue (a yellow arrow) appears on the same or opposite side. The cue stays on for 300 msec and the participant must react as soon as possible by pressing a button. If the cue appears on the same side as the target this is a Valid condition. If it appears on the opposite side it is called Invalid. The ration of Invalid to Valid trials is 2:5. 68 trials take place in a run, and each subject does 4 runs, two each of egocentric and allocentric cues and targets. 12 subjects took part.

The cue and targets all appear either in egocentric or allocentric space. In the first the cues are attached to the head and are nearby. In the latter they are distant and stationary (see Figure 19). The size of the cues and targets is calibrated so that despite distance the cues and targets each project the same image on to the retina (see Figure 20 and Figure 21).

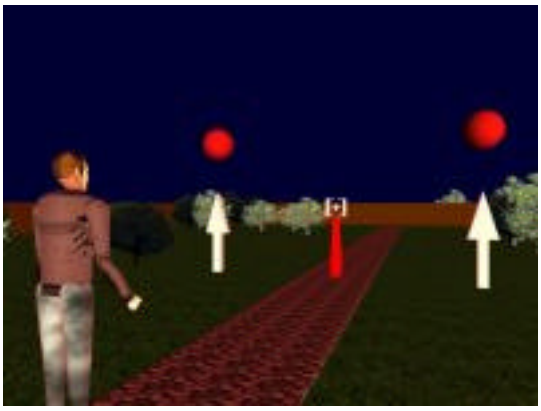


*Figure 19 Participant viewing allocentric cues and targets*

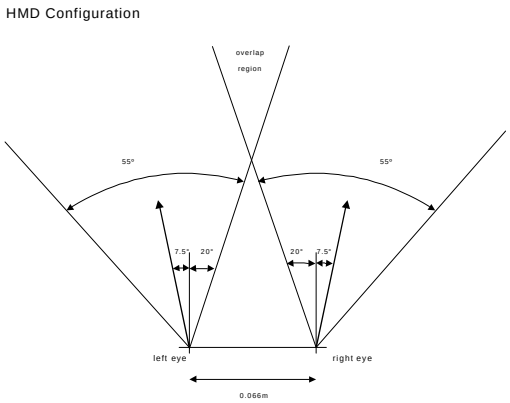



*Figure 20 HMD Configuration*

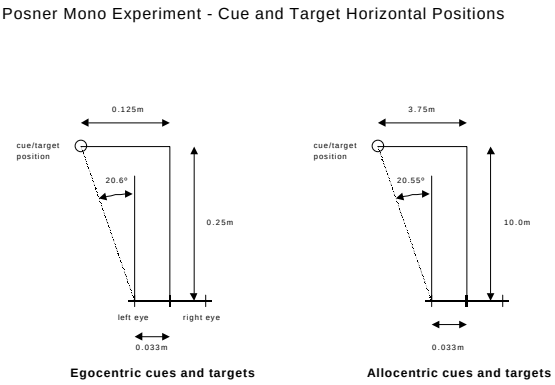



*Figure 21 Posner Mono Experiment - Cue and Target Horizontal Positions*

The results showed that the Posner effect was present and the magnitude of the effect was consistent with previous results. Reaction time being measured at roughly 430 msec, with participants reacting 15 msec quicker to Valid trials than Invalid trials. There was no effect of egocentric and allocentric space. In fact although each subject experiences both egocentric and allocentric trials, few noticed the difference.

### 8.3.2 Experiment 2: Introducing Stereo Disparity

The first experiment had been configured with the cues and targets appearing in the periphery of the display in the non-overlapping (i.e. non-stereo) region. For the second experiment the cues and targets were brought closer together so that stereo disparity cues come into play when determining depth. Bring the cues and targets closer together does however decrease the expected magnitude of the Posner effect. Unlike the previous experiment, the cue and target in a single trial could appear at different depths and well as different sides. This gives 4 possible trial types: Valid, Invalid Side, Invalid Depth, and Invalid Side and Depth. A fifth trial type, Catch, was introduced to reduce the number of errors reported. In a Catch trial the cue is flashed but no target appears. Another condition was also introduced: whether or not the participant had a virtual body. The egocentric cues were placed within reachable distance of the participant. Figure 23 shows an egocentric cue and an allocentric target. The fixation cross is in the middle distance, and the participant is told to focus on this during the trials.



*Figure 22 Participant viewing a egocentric cue and allocentric target.*

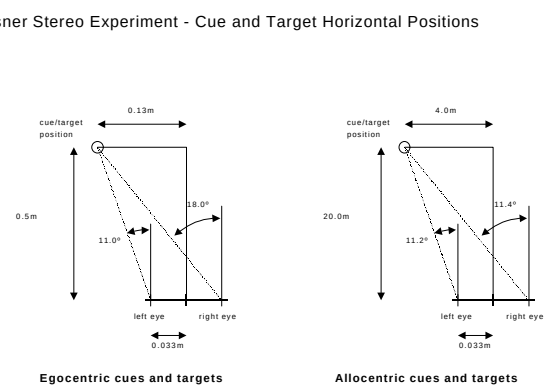



*Figure 23 Stereo Cue and Target Positions*

With the five types of trials, a longer run was required. 400 trials were administered in one long run.

In summary the subjects were 15ms quicker in responding to Valid trials in side than Invalid trials (the standard Posner effect), and were also 8 ms quicker repsonding to Valid depth trials than invalid (the hypothesized Posner-type depth effect). The subjects were also 7.5 ms quicker in responding to far targets than near.

Two interaction effects were significant, Virtual Body/Type of Space/Type of Trial in Side, and Virtual Body/Type of Space/Type of Trial in Side/Type of Trial in Depth. See [27] for full details of the analysis. In summary it appears that the presence of the virtual body affects the overall attention field, in that subjects with a virtual body show a relative bias towards responding to egocentric targets and vice-versa. This effect is extremely important for virtual environment systems were attention cues play a major role in performance. This leads us to suggest that in training situations were situational awareness is important, a virtual body should present since if nothing else it seems to create a field of attention reaction that is more similar to real-life.

# 9 Combined Recommendations

This second phase of the usability evaluation has taken a broad look at the platform and application issues involved in the supporting collaborative virtual environments. This included network trials with expert users, controlled scientific trials with naïve users and usability inspection. In this Section we briefly summarize the usability results in order to inform the final implementation phase. In addition we give some recommendations for future development of the methods to be used in the final COVEN evaluation phase.

## 9.1 Combined Recommendations

We summarize the outstanding usability levels at an application and system level. The former is intended to be input to the final implementation phase of the COVEN demonstrator. The latter is a more general indicator of where future research on CVEs might focus since these are difficult problems to solve in the short term.

### 9.1.1 Application Functionality

The tables in Section 8 of Part C of this Deliverable give an in-depth report of the usability issues that were found in the on-line application. Most aspects of the application are covered in some way, though the changes suggested are most cosmetic. We list some important areas for consideration, augmenting the inspection comments with emphasis from the network trials:

1. Identify Receivers - needs to be more obvious. Click on receivers or select from a pull down list.
2. Send Message: Composition - only two lines is not enough. Better error correction and delivery notification.
3. Send Message: Specify mode - better modes for scoping audio. Named zones are usable but a simple locality mode would be better.
4. Audio: Push to talk - a system issue (see Section 9.1.2).
5. Move as Single: Teleporting - HUD teleport maps needs an in-VE icon so that the action is always "available" at the interface.
6. CD Player - larger buttons on the control, and an instructions screen as default on the projection screen.
7. Teleporter - total removal of the teleporter door would simplify usage.
8. Guide Posts - modality of play/stop button. Immediate and obvious visual feedback is required when the button is pressed.
9. Carpet controls - the modal nature of the buttons needs investigation.
10. Hotel Signs- although help is now provided a portable, in-virtual environment key is required.

Note that during network trials it was found that some functionality was not working, or had been disabled in order to support large groups. In particular group teleportation was no longer enabled, and therefore could not be tested. (The inspection used the same version of the application as the network trials).

The most serious application problem is the relative difficulty of using the text communication. It is our suggestion that this be simplified and made menu based so that no text command line are required. A form of receiver buffer is absolutely vital - at the moment received messages over-write previous ones and there is no way of retrieving them. Some sort of scroll back is required. Identifying receivers from a menu would alleviate problems associated with name discovery and misspelling. Mode specification is also tricky. A more detailed design document can be provided in time for the next implementation phase.

A common usability theme is that although the functionality is available somehow in the VE, it is not "visible" at the interface. Some sort of portable interface that presents icons representing interface functionality is desirable. This would act as PDA, with icons representing global functions (talk, teleport), personal functions (reset position) and zone functions (move to teleport, move to CD player, fly to island). The PDA could thus unify the carpet controls and zone dependent functionality. It would remove the need for keyboard interaction. A more detailed design document can be provided in time for the next implementation phase.

One serious problem was introduced during the network trials: a mis-match of plugins to provide the name tags and to identify text receivers. Thus the name of the avatar as provided by their label was not the one to which text messages should be addressed (this was accessed through the ?Who command).

### 9.1.2 System Functionality

Through the Dive trials we have found that the spatialisation of audio have been a boon for collaboration. Even just distance attenuation would be a benefit over the default RAT usage where every source is played at a constant volume. Attenuation cues themselves would allow people to distinguish between speakers. At the moment there is no correspondence inferable from the audio, nor visual cues projected into the environment such as the Dive audio "waves". This issue is probably the greatest hindrance to collaboration in the dVS trials, but unfortunately given the feedback from the first evaluation phase, it is also very difficult to implement.

A further issue with the audio is the lack of reliable measures of fallibility. It would be a great asset to be able to tell if someone had received your audio. Then they would not appear rude if they ignored you. This facility is available in the latest generation of Internet conferencing tools [15], and it would be useful to integrate them into the platform.

From our experience in the Dive trials, there is a need for avatars that express body language. The required representations are possible using the results of COVEN work packages [25]. What is not known is how the functionality of these representations should be presented to the user.

A further issue is that with 16 users and their avatars the rendering load on the client machines rises dramatically. It would not be possible to use much more detailed avatars, since even though they have many levels-of-detail it is common for one user to be in close proximity with several others. We expect that the rendering results from Work Package 8.4, that provide a constant frame-rate will alleviate this problem somewhat.

A serious issue that needs to be resolved with the dVS trials is the help system. At the moment this spawns Netscape browsers from the server machine to clients. This seriously overloads the server machine and causes the whole system to crash. A solution needs to be found where the browser is spawned and updated from each client machine.

## 9.2 Method Development

The problems of actually **doing** usability engineering for virtual environments systems are only just being realised [16][43]. Most of the work undertaken in this phase of the COVEN evaluation has been in developing the requisite methods. The approach taken has been to start from an analysis of CVE tasks (see Section 4), and from this provide design advice and inspection methods (see Section 3). Further more we have developed an observational method that can be used to correlate predicted and actual actions that are carried out in network sessions (see Section 5).

The usability inspection method is now undergoing revisions and testing on other applications. One way in which the development is being informed is by structured interviews with CVE application designers so that the results of the inspection can be tailored to their needs. Currently there are few CVE design methods and most development follows the two rules of making the environment simple to keep the frame-rate up but to make it realistic so that real-world metaphors can be applied. As we have seen from our inspections, the latter rule is far too broad. The presentation player metaphor is now less "realistic" but still comprehensible and easier to use. A next step is to take the results of the inspection and develop design guidelines.

The supporting work for the inspection, the "Theory and Method Description for Spatial and Temporal Activity Charting of CVE Interactions" document and the "Collaborative Actions in CVEs" document are both under development. The former is undergoing testing on videotapes of the recent network trials. The latter is being refined from feedback from CVE designers, experience with the inspection and experience with the review of actual trials.

Finally we expect to carry on development of the small group experiments and presence experiments. These threads of work will hopefully form the basis of a strong theory of presence and co-presence in virtual environments. Further experiments are already being planned.

# References


[1] Bales, R.F., (1951). Interaction Process Analysis; A Method for the Study of Small Groups, Addison-Wesley Press, Cambridge.

[2] Bowers, J., Pycock, J., O'Brien, (1996). Talk and Embodiment in Collaborative Virtual Environments, in: Proc. of CHI96, ACM Press.

[3] COVEN (1997) Usage Evaluation of the Initial Demonstrators, Deliverable 3.3 (A.Steed and J.Tromp eds.)

[4] COVEN (1998) Online Citizen Application, Deliverable 2.4

[5] COVEN (1998) Online Business Application, Deliverable 2.5

[6] Diaper, D. (ed.), (1998). Task Analysis for Human-Computer Interaction, Ellis Horwood Ltd, England.

[7] Division Ltd. (1997) dVS/dVISE 5.0 Software. http://www.division.com

[8] Duncan, S. Jr., (1975). Interaction Units during Speaking Turns in Dyadic, Face-to-Face Conversations, in: Kendon, A., Harris, R.M., Ritchie Key, M., (eds.), Organization of Behavior in Face-to-Face Interaction, Mouton Publishers, Den Hague, pp. 199-213.

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

[10] Ellis, S.R. (1996) Presence of Mind: A Reaction to Thomas Sheridan's "Further Musings on the Psychophysics of Presence", Presence: Teleoperators and Virtual Environments, MIT Press, 5(2), 247-259.

[11] Erickson, F., (1975). One Function of Proxemic Shifts in Face-to-Face Interaction, in: Kendon, A., Harris, R.M., Ritchie Key, M., (eds.), Organization of Behavior in Face-to-Face Interaction, Mouton Publishers, Den Hague, pp. 175-187.

[12] Frecon, E., Stenius, M. (1998) DIVE: A Scalable Network Architecture for Distributed Virtual Environments, Distributed Systems Engineering Journal, 5(3), September 1998, pp 91-100.

[13] Goffman, E. (1974) Frame analysis: an essay on the organisation of experience, New York, Harper and Row.

[14] Greenhalgh, Chris, Steve Benford and Adrian Bullock (1998) Predicting Network Traffic for Collaborative Virtual Environments, to appear at TERENA Networking Conference 98 (TNC98), Dresden, 5-8 October 1998.

[15] Hardman, V., Sasse, M.A., Kouvelas, I. (1998) Successful Multiparty Audio Communication Over the Internet, Communications of the ACM, 41(5), May 1998.

[16] Heath, C., Luff, P., Sellen, A., (1995). Reconsidering the Virtual Workplace: Flexible Support for Collaborative Activity, in Proc. of the European Conference on Computer Supported Cooperative Work (ECSCW'95), Sweden.

[17] Helander, M., Landauer, T.K., Prabhu, P., eds, (1997). Handbook of Human-Computer Interaction, 2nd Ed., Elsevier Science, Holland.

[18] Held, R.M. and N.I. Durlach (1992) Telepresence, Presence: Teleoperators and Virtual Environments, 1, winter 1992, MIT Press, 109-112.

[19] Hindmarch, J., Fraser, M, Heath, C, Benford, S, Greenhalgh, C, (1998). Fragmented Interaction: Establishing mutual orientation in virtual environments, in: Proc. of CSCW'98, Seattle.

[20] Kanuritch, N., Farrow, G, Pegman, G.J., Prof Wilson, J.R., Dr. Cobb, S., Crosier, J, Webb, P., (1997). A Study of the Safety and Usability of Teleoperated Remote Handling Systems, final report UKRL.97.023.

[21] Kaur, K., Maiden, N., Sutcliffe, A., (1997). Interacting with Virtual Environments: an evaluation of a model of interaction, in: Interacting with Computers, Special Issue on VR.

[22] Kendon, A., Harris, R.M., Ritchie Key, M., (eds.), (1975). Organization of Behavior in Face-to-Face Interaction, Mouton Publishers, Den Hague.

[23] Leary, M. (1983) Social Anxiousness: The Construct and its Measurement, Journal of Personality Assessment, 47, 66-75.

[24] Kohler, W. (1959) Gestalt psychology: An introduction to new concepts in modern psychology. NY: Liverright, 1947; NY:NAL, 1959.

[25] Magnenat-Thalmann, N., Thalmann D., Babski C., Carion S. (1997) Report on human modelling and animation in dVS and Dive Environemnt, WP4.4, final deliverable, December 1997.

[26] Markel, N.N., (1975). Coverbal Behavior Associated with Conversational Turns, in: Kendon, A., Harris, R.M., Ritchie Key, M., (eds.), Organization of Behavior in Face-to-Face Interaction, Mouton Publishers, Den Hague, pp. 189-197.

[27] Maringelli,F., McCarthy, J. , Steed, A., Slater, M. and Umiltà, C. (1998) Shifting Visual Attention in a Three Dimensional Space: evidence in favour of a dissociation between attentional systems operating in the proximal and distal space. Submitted for publication.

[28] Melchior, E-M, Bosser T., Meder, S., Koch, A., Schnitzler, F. (1995) Handbook for practical usability engineering in IT project. Project ELPUB105 10107, Telematics Application programme, Information Engineering information section.

[29] Milner, A. D., & Goodale, M. A. (1995). The visual brain in action. Oxford University press, Oxford.

[30] Neale, (1997). A Structured Evaluation of Virtual Environments in Special Needs Education, BEng Honours degree dissertation in Manufacturing Engineering and Management, University of Nottingham.

[31] Neilsen, J. and Mack, R.L. (1994) Usability Inspection Methods. John Wiley and Sons.

[32] Nohria, N, & Eccles, R., (1992). Networks and Organizations, Cambridge, MA, HBS Press.

[33] Posner, M. (1980). Orienting of attention. Quarterly Journal of Experimental Psychology, 32, 3-25.

[34] Robertson, T., (1997). Cooperative Work and Lived Cognition: A Taxonomy of Embodied Actions, in: Proceedings of the ECSCW'97 conference, Lancaster.

[35] Rocco, E., (1998). Trust Breaks Down in Electronic Contexts but Can Be Repaired by Some Initial Face-to-Face Contact, in: Proc. of CHI'98.

[36] Scheflen, A.E., (1975). Micro-territories in Human Interaction, in: Kendon, A., Harris, R.M., Ritchie Key, M., (eds.), Organization of Behavior in Face-to-Face Interaction, Mouton Publishers, Den Hague, pp. 159-173.

[37] Sheridan, T.B. (1996) Further Musings on the Psychophysics of Presence, Presence: Teleoperators and Virtual Environments, MIT Press, 5(2), 241-246.

[38] Slater, M., Sadagic, A., Usoh, M., Schroeder, R. (1998) Small Group Behaviour in Virtual and Real Environments: A Comparative Study. Submitted for publication. Summary available at http://www.cs.ucl.ac.uk/staff/m.slater/BTWorkshop.

[39] Slater, M., Steed A. (1998) A Virtual Presence Counter, accepted for publication in Presence: Teleoperators and Virtual Environments, MIT Press.

[40] Slater, M., Steed, A., McCarthy, J., Maringelli, F. (1998) The Virtual Ante-Room: Assessing Presence Through Expectation and Surprise, Virtual Environments '98, Eurographics Workshop Proceedsings Series, eds M. Goebel, U.Lang, J. Lanauer, M.Wapler.

[41] Slater, M., Steed, A., McCarthy, J., Maringelli, F. (1998) The Influence of Body Movement on Subjective Presence in Virtual Environments, Human Factors, in press.

[42] Slater, M., Wilbur S. (1997) A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments, Presence: Teleoperators and Virtual Environments, 6(6) 603-616, MIT Press.

[43] Steed, A., Tromp, J.G., (1998). Experiences with the Evaluation of CVE Applications, in: Proceedings of Collaborative Virtual Environments Conference (CVE'98), 17-19 June, Manchester.

[44] Tromp, J.G., Bullock, A., Steed, A., Sadagic, A., Slater, M., Frecon, E. (1998) Small Group Behaviour Experiments in the COVEN Project, IEEE Computer Graphics and Applications, 18(6), November/December 1998, pp 53-63.