



• D3.3, Usage Evaluation of the Initial Applications

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

CEC Deliverable Number:	A040-UCL-CS-DS-P-033.b1
Contractual Date of Delivery to the CEC:	1 st September, 1997
Actual Date of Delivery to the CEC:	3 rd October, 1997
Title of Deliverable:	Usage Evaluation of the Initial Applications
Work package contributing to the Deliverable:	WP3
Nature of the Deliverable:	R**
Author(s):	UCL, IIS, Nottingham, Lancaster, Thomson, SICS

Abstract:

This document reports upon the formulation, results and conclusions of a usability evaluation of the initial COVEN demonstrators. Given the lack of prior work on evaluation of Collaborative Virtual Environments (CVE), we derive a framework for evaluation that addresses the particular constraints and needs of the area. From this we established three strands of work which enabled us to both evaluate the applications and the methodologies chosen. These strands are: a usability inspection, observational studies of network trials and auxiliary case-controlled experiments focused on the evaluation of the factors to the 'presence' concept. From the results of the different studies we produce a set of requirements for the development of the COVEN applications, and also propose a refined set of methods for CVE usability evaluation.

Keyword list:

Collaborative Virtual Environments, Usability Inspection, Presence, Cognitive Walkthrough, Heuristic Evaluation

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

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

Executive Summary

The first phase of the COVEN usability evaluation activities addresses the evaluation of the initial versions of the COVEN demonstrator applications. Our first task was to lay the foundations for the COVEN usability evaluations through clarifying a framework for our activities: identifying our goals, constraints, and general approach; this was done taking into account the conflicting concerns of usability engineering and scientific inquiry on CVE concepts, as well as our specific constraints and needs at this stage of the COVEN project. Within this framework, three threads of evaluation work were then launched and developed so as to answer two main goals: 1) provide support to ‘clean up’ the design of the initial applications; 2) further develop a fundamental understanding of the CVE technology concepts.

- usability inspections of the initial applications so as to uncover the main design flaws and allow to clean up the design.
- observational evaluations of participants performing tasks in networked trials, so as to better understand behavioural characteristics and underlying concepts.
- auxiliary case-controlled experiments focused on the evaluation of the factors to the ‘presence’ concept.

These evaluations have yielded a considerable amount of results at all levels. The inspections allowed to identify a significant number of issues at system, interaction and application levels, providing highly valuable input for the application and system designers to improve their design. The observations and case-controlled experiments allowed us to refine our understanding of the behaviour of participants in CVEs, and of the fundamental concepts underlying the usage of CVE technology.

Finally, our experience in this first phase of the COVEN evaluations are feeding a reflection at methodological level; the lessons learned should allow us to further refine our methods in the perspective of the second iteration of the COVEN usability evaluations, and to usefully contribute to the research on this subject.

This document presents the work that was undertaken in this initial stage of usability evaluations, and summarises the results that were achieved.

Contents

1. INTRODUCTION	5
2. FRAMEWORK FOR USABILITY EVALUATION	7
2.1 CONFLICTING CONCERNS: USABILITY ENGINEERING AND SCIENTIFIC INQUIRY	7
2.1.1 Usability engineering view point	7
2.1.2 Scientific inquiry view point	10
2.2 SPECIFIC FACTORS IN CVE EVALUATION	10
2.2.1 CVE as a place; inside and outside view points	10
2.2.2 The distributed nature of the system under evaluation	11
2.2.3 Evaluating prototypes	11
2.3 USABILITY EVALUATION IN THE COVEN PROJECT LIFE	11
2.4 PROCESS IN THE FIRST USABILITY EVALUATION STAGE	13
2.4.1 Evaluating core application usability	13
2.4.2 Further developing a fundamental understanding of the CVE technology concepts	14
2.4.3 Refining a methodology for CVE usability evaluation	14
3. APPLICATION OVERVIEW	16
3.1 BUSINESS APPLICATION	16
3.2 CITIZEN APPLICATION	17
4. USABILITY INSPECTION	20
4.1 METHODS	20
4.1.1 Heuristic Evaluation	20
4.1.2 Cognitive Walkthrough	21
4.1.3 Severity ratings	21
4.1.4 Logistics	22
4.2 RESULTS	23
4.2.1 System Problems	24
4.2.2 Interaction Problems	25
4.2.3 Application Problems	27
5. NETWORK TRIALS	32
5.1 NETWORK INFRASTRUCTURE	32
5.2 TRIAL PLATFORM	32
5.3 TRIALS AND TASKS	33
5.4 GENERAL EXPERIENCE	34
5.4.1 Communication	34
5.4.2 Interaction Metaphors	36
5.4.3 Other Services	36
5.5 BUSINESS APPLICATION EXPERIMENT	37
5.5.1 Task	37
5.5.2 Results	37
5.5.3 Conclusions	39
5.6 CITIZEN APPLICATION EXPERIMENT	40
5.6.1 Task	40
5.6.2 Results	40

5.6.3 Conclusions	43
6. AUXILIARY TRIALS	44
6.1 EFFECT OF BODY MOVEMENT ON THE SENSE OF PRESENCE	44
6.1.1 Presence Model	44
6.1.2 Experimental Design.....	45
6.1.3 Results	47
6.1.4 Conclusions	47
7. APPLICATION RECOMMENDATIONS	48
7.1 RECOMMENDATIONS FROM INSPECTION.....	48
7.1.1 General recommendations	48
7.1.2 Detailed suggestions for design alternatives	50
7.2 RECOMMENDATIONS FROM NETWORK TRIALS	52
8. METHODOLOGICAL RECOMMENDATIONS	54
8.1 INSPECTION EXPERIENCES	54
8.1.1 Using the COVEN Inspection Method	54
8.1.2 Suggested improvements to the Inspection Method.....	56
8.1.3 Suggested Improvements to the Inspection Topics	57
8.2 NETWORK TRIAL EXPERIENCES.....	57
8.3 AUXILIARY TRIAL OUTPUT	57
8.3.1 Measures of Subjective Presence	57
8.3.2 Virtual Ante-Room	58
9. CONCLUSIONS.....	60
10. ANNEXES	61
10.1 USABILITY EVALUATION METHODS	61
10.2 GLOSSARY OF USABILITY EVALUATION TERMS	62

1. Introduction

Usability evaluation studies are needed to ensure that the perceptual and cognitive capabilities and limitations of human beings, as well as the requirements of the specific tasks under consideration are being used as driving criteria for system design; that the system hardware and software deliver the Collaborative Virtual Environment application in a cost-effective manner; and that the CVE application represents a significantly better way of doing old things or of doing new things that were not possible with more classical technology.

A methodology for CVE usability evaluation?

At this early stage in the development of CVE technology, the tendency in the community is generally to place the focus on technology development, with very few resources devoted to the development of applications of a reasonable size (toy systems only). In relation to this, evaluations of some aspects of CVE applications can be found in the literature (in particular, physiological and cognitive VR experiments), but no comprehensive approach is ever reported, nor is actually usability evaluation a common practise in the CVE community.

In this context, considering the specificity of CVE applications, a usability evaluation methodology for CVEs does constitute a full research topic. CVE applications are very specific in that they are 3D, multi-sensory, immersive and multi-participant systems. Insight can be taken from methodological results of the HCI community, which yielded a considerable amount of practises, methods and techniques to system usability evaluation; the CSCW community is also providing particular evaluation tools (observational approaches) to evaluate technology-mediated co-operation and communication between people. Starting from this basis of experience and tools, a small community of people is addressing the evaluation of the usability of CVEs, among which the COVEN group.

The COVEN initial usability evaluation phase:

The first phase of the COVEN usability evaluation activities addresses the evaluation of the initial versions of the COVEN demonstrator applications. Our first task was to lay the foundations for the COVEN usability evaluations through clarifying a framework for our activities: identifying our goals, constraints, and general approach; this was done taking into account the conflicting concerns of usability engineering and scientific inquiry on CVE concepts, as well as our specific constraints and needs at this stage of the COVEN project. Within this framework, three threads of evaluation work were then launched and developed so as to answer two main goals: 1) provide support to ‘clean up’ the design of the initial applications; 2) further develop a fundamental understanding of the CVE technology concepts.

- usability inspections of the initial applications so as to uncover the main design flaws and allow to clean up the design.
- observational evaluations of participants performing tasks in networked trials, so as to better understand behavioural characteristics and underlying concepts.
- auxiliary case-controlled experiments focused on the evaluation of the factors to the ‘presence’ concept.

These evaluations have yielded a considerable amount of results at all levels. The inspections allowed to identify a significant number of issues at system, interaction and application levels, providing highly valuable support for the application and system designers to improve their design. The observations and case-controlled experiments allowed to refine our understanding of the behaviour of participants in CVEs, and of the fundamental concepts underlying the usage of CVE technology.

Finally, our experience in this first phase of the COVEN evaluations are feeding a reflection at methodological level; the lessons learned should allow us to further refine our methods in the perspective of the second iteration of the COVEN usability evaluations, and to usefully contribute to the research on the subject.

Organisation of the document:

This document presents the work that was undertaken in this initial stage of usability evaluations, and summarises the results that were achieved. In view of the amount and variety of the tasks undertaken, the main body of this document is structured into seven sections:

- Section 2 gives a general presentation of the framework within which we placed our usability evaluations.
- Section 3 provides a quick overview of the applications that were evaluated.
- Section 4 describes the usability inspections that were conducted.
- Section 5 describes the evaluations that were performed on network trials.
- Section 6 describes the auxiliary trials that were performed in a case-controlled manner.
- Section 7 emphasises the recommendations that were made toward the application designers, out of the evaluations results.
- Section 8 highlights some methodological recommendations extracted from our experience in this first evaluation stage.

In the Annex part of this document, we are moreover providing synthetic background information on usability evaluation methods, as well as a *glossary* of usability evaluation terms.

N.B.: the contents of this document are further detailed in a set of three internal COVEN documents: ‘Framework for usability evaluations’, ‘Evaluation strategy’, and ‘Combined inspection report’.

2. Framework for Usability Evaluation

In the absence of an existing dedicated CVE evaluation methodology, the design of a framework for the COVEN usability evaluation studies was our first task in this workpackage. Our objective was to precisely define our goals, constraints and approach for the usability evaluation studies, to describe the structure of our usability evaluation process model and to clarify the way it is incorporated into the development process of the COVEN Platform and Demonstrators. A first version of a general framework was built and then instantiated in the context of our first evaluation stage in the COVEN workplan; this was used as the reference frame for designing and articulating our different evaluation activities and experiments in this first evaluation stage of the project.

This section provides a summary of the rationale and contents of this framework, which is further detailed in the ‘Framework for usability evaluation’ internal COVEN document. Some background is provided on the driving forces that are motivating and shaping the COVEN usability evaluation activities: the different, conflicting concerns of usability engineering and scientific inquiry. Our general process for usability evaluation in the overall COVEN project is then presented, and the instantiation of our initial evaluation process is described.

2.1 Conflicting concerns: usability engineering and scientific inquiry

The design of this framework was influenced by two main considerations: usability engineering and scientific inquiry. On the one hand, the COVEN project is developing Demonstrator applications intended to demonstrate the added-value of the CVE concept from the end-user and the customer point of view; this requires an evaluation of the overall usability of the CVE applications, together with elements of a cost/benefit analysis - this places our evaluation activities in the general frame of *usability engineering* methods. We are addressing the evaluation of the benefit of CVE applications, as assessed through the overall usability of our Demonstrators.

On the second hand, the CVE technology on which these Demonstrators are built is in its early stages, with in particular the human factors impact of its specific features still poorly explored. Investigating the human behavioural aspects which affect performance and satisfaction in CVEs is precisely one of the objectives of COVEN; this requires focused exploratory studies of specific phenomena - this clearly places our evaluation activities in the general frame of *scientific inquiry*. These focused studies should provide insight into the specific and unique features of CVEs, especially with regards to presence, co-presence, awareness of ones virtual self, and of the other participants; this comes as a mandatory foundation-level component to our overall evaluations of the benefit of CVE technology (Tromp, 1996).

The main challenge in defining our usability evaluation framework is to have these two concerns and their related sets of evaluation techniques cohabit and feed each other in a relevant, fruitful and well-controlled manner. The final result of these evaluations needs to be an informed description of the usability benefits and shortcomings of CVE applications, as generalised from our experience with the Demonstrator applications.

2.1.1 Usability engineering view point

Our activities in COVEN are in keeping with the general pattern of user-centred system development. This well-documented approach developed in the HCI community (see e.g. Norman (1986) or Nielsen (1993)) relies on usability engineering activities at all stages of the development process: project planning phase (user needs elicitation and task analysis), application building phase (iterative design) and demonstration/validation phase (user validation).

Whatever the target stage within the system development, usability evaluation requires a structured and well-documented approach so as to ensure the quality, utility and pertinence of the collected results, to allow comparisons over several stages of evaluations, to control the general cost-effectiveness of the usability activities themselves, and to pool the results for future reference. A reasonable amount of literature and experience has been gathered until now on this subject, providing general but also practical guidelines and recommendations; our approach in COVEN is to adapt such state-of-the-art user validation framework elements to our specific context.

In particular, we have used the ‘Handbook for practical usability engineering in IE projects’ (Melchior et al., 1995) produced by the ELPUB project of the Telematics Application Programme. This handbook provides the practical description of a general framework for user validation focused on the needs of Information Engineering projects (multimedia services, electronic shopping); this framework is generic enough to provide a sound basis for structuring our activity. Our main task was to consider each element of this framework and instantiate it within the COVEN context, creating a framework adjusted to the specific objectives and constraints of the COVEN project.

Usability engineering process

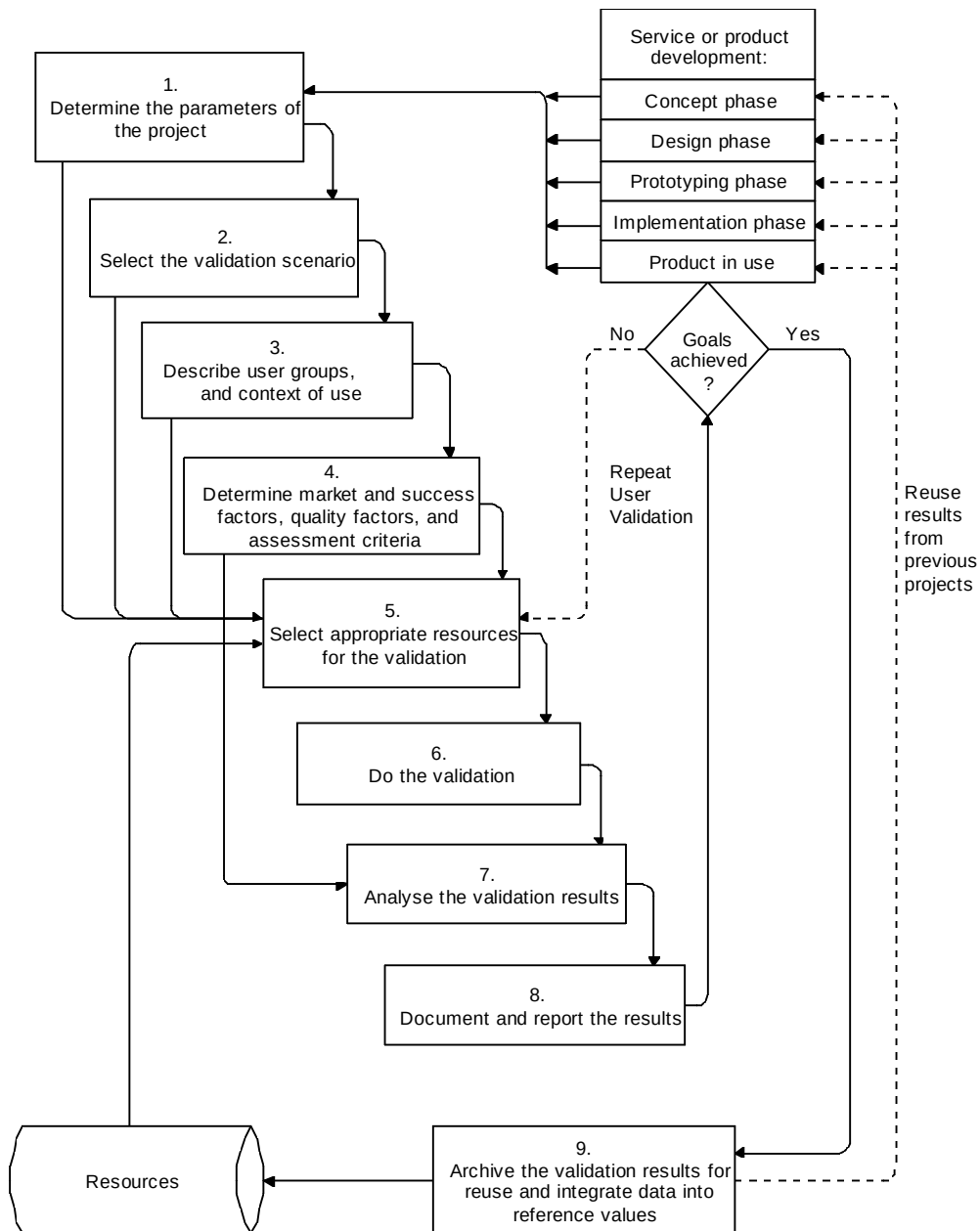


Figure 1: Structure of a standard usability evaluation process, reproduced from (Melchior et al., 1995)

Figure 1 presents the structure of a standard usability evaluation process, which was used to organise the COVEN usability engineering activities. Usability evaluation methods classically involve usability inspections and usability experiments¹ aimed at the evaluation of the overall system usability or at more focused usability issues.

¹ N.B.: synthetic background information on usability evaluation methods is provided as an annex in section 10.1.

2.1.2 Scientific inquiry view point

CVE technology is in its early stages, it is a relatively young, multi-disciplinary science (Kalawsky, 1993). Like any new science, it aims to discover new concepts to develop, such as presence in a virtual environment, virtual embodiments, and working together in a virtual space. These concepts become constructs in a model of human behaviour and this human behaviour is explored in order to come to operational definitions for further empirical research.

The development and demands of social behaviour in CVE will only become clear by people using the software, and by exploring the needs of these users. Examples of constructs which have been identified so far are presence, and co-presence, immersion, awareness, spatial phenomena, wayfinding, subjective views, believability, collaborative work, etc.

Many of the constructs which need to be addressed to fully exploit to capabilities of CVE software and hardware are either unknown, or unexplored, and untested. Thus, because CVE development is still in its early stages, we often need to employ an exploratory approach to identify the aspects of human behaviour which affect performance and satisfaction. Sociological ethnographic techniques, psychological exploratory techniques and case-controlled experiments are particularly suitable at this phase of the empirical cycle of inquiry.

Ethnographic CVE investigation entails extended CVE participant observation with and within the CVE. The rationale is that a prolonged period of intense immersion in a culture best enables the ethnographer to experience the world of his or her subjects, and hence to grasp the significance of their language and actions. For CVE development and evaluation this means that ethnographic inquiry is aimed at trying to interpret the hidden assumptions of the designers, their pioneering behaviours, their experiences and wishes with the CVE technology, by describing the field, and registering the phenomena. By focusing on uncovering these hidden assumptions there is a lot of design information to be found, which can help to identify user needs and requirements.

Psychological exploratory investigations are not primarily aimed at describing phenomena, investigations of the field are made to articulate and select possible hypotheses. Phenomena are recorded, and ordered, in order to establish relationships between them and to come to would-be hypotheses. For CVE explorations this means that the researcher lets the subjects 'speak for themselves', in order to gather as much concrete data as possible. The concrete data is analysed to find the elements which are causing or related to the phenomena observed. This will then help the CVE researcher to decide what data are to be used, what is to be measured, and roughly what relationships are to be studied in future evaluations of performance and satisfaction with CVEs.

Case-controlled experiments are used to investigate the validity of a particular claim or hypothesis. This is typically used to investigate the impact of system characteristics on concepts such as the 'sense of presence' that is afforded to the participant, in an attempt to further characterise and understand the dimensions of these behavioural concepts.

2.2 Specific factors in CVE evaluation

In addition to the exploratory nature of CVE technology, a number of factors may be identified, which further characterise the specificity of CVE usability evaluation, and the constraints within which evaluation activities are taking place.

2.2.1 CVE as a place; inside and outside view points

A major specificity of CVE evaluation with regards to 'classical' HCI evaluation is that the evaluation task of meas-

uring human performance can take place on two levels. We are interested in human behaviour and performance with the application; at the same time, we are also interested in human behaviour and performance inside the CVE, answering questions about perception of 3D computer generated space, navigation, (co-) presence, and (group-) awareness, etc. Thus, in general observations and experiments can and need to be performed both from outside the CVE and from inside the CVE.

2.2.2 The distributed nature of the system under evaluation

We have to face a number of difficulties related to the geographical distribution of users, common to the evaluations of CSCW systems in general. A CVE allows multiple users to interact simultaneously within the CVE in real-time, regardless of the physical location of these users. One of the implications of the distributed character of the CVE application and its users, is that it becomes more difficult to conduct proper controlled experiments.

A typical concerns for experimental set-ups is that they should be as similar as possible for each subject in each condition of the study. In order to be able to claim with confidence that the observed difference in behaviour is attributable to a specific different implementation of a construct in the CVE, the researcher needs to rule out any influences on the user, other than the desired ones. This means that the environments of the users should be as similar as possible, the researcher has to behave as similar as possible with each subject in the study, questionnaires should be answered at similar times as possible, subjects within one group should be as similar as possible, subjects within one group should receive similar treatment, etc. When conducting distributed usability studies this becomes a complicated task, because the researcher can not be in all places simultaneously in order to guarantee a similar treatment of all subjects. This needs to be taken into account when assessing and generalising the findings.

2.2.3 Evaluating prototypes

Some constraints are related to the fact that the evaluations are performed on prototypes that are not fully functional products, even though the COVEN developments should lead to stable and documented applications featuring a reasonable number of functions.

Because the production of CVEs is taking place within the developmental phase of the production, the final product is often a prototype or demonstrator. Boundaries of the existing technology are pushed to create new ways of doing things, and as a result even more new things become possible. One of the side-effects of doing evaluations on a prototype which is still under development, is that there may remain defects in the functioning of the application, and there are few opportunities for using proper representative subjects from the population of intended users for the usability studies - generalisation of findings will have to be carefully performed.

Another consequence of evaluating prototypes is that it is often not feasible within the time and effort available to create two (or more) different situations for an experiment. This means that the independent variables can not always be manipulated. For instance, a researcher may have found that having a personal shadow in a CVE may assist orientation and wayfinding. In order to find out what kind of shadow is most effective the researcher needs at least two versions of the CVE, each one with a different kind of shadows, and preferably one CVE without any shadows at all, for the control group. These three versions of the CVE constitute the manipulation of the variable 'shadow'. The group that does best on the orientation and wayfinding points to the CVE with the best shadow. Obviously this is an informative, but labour intensive way of gathering knowledge, which may not always be possible.

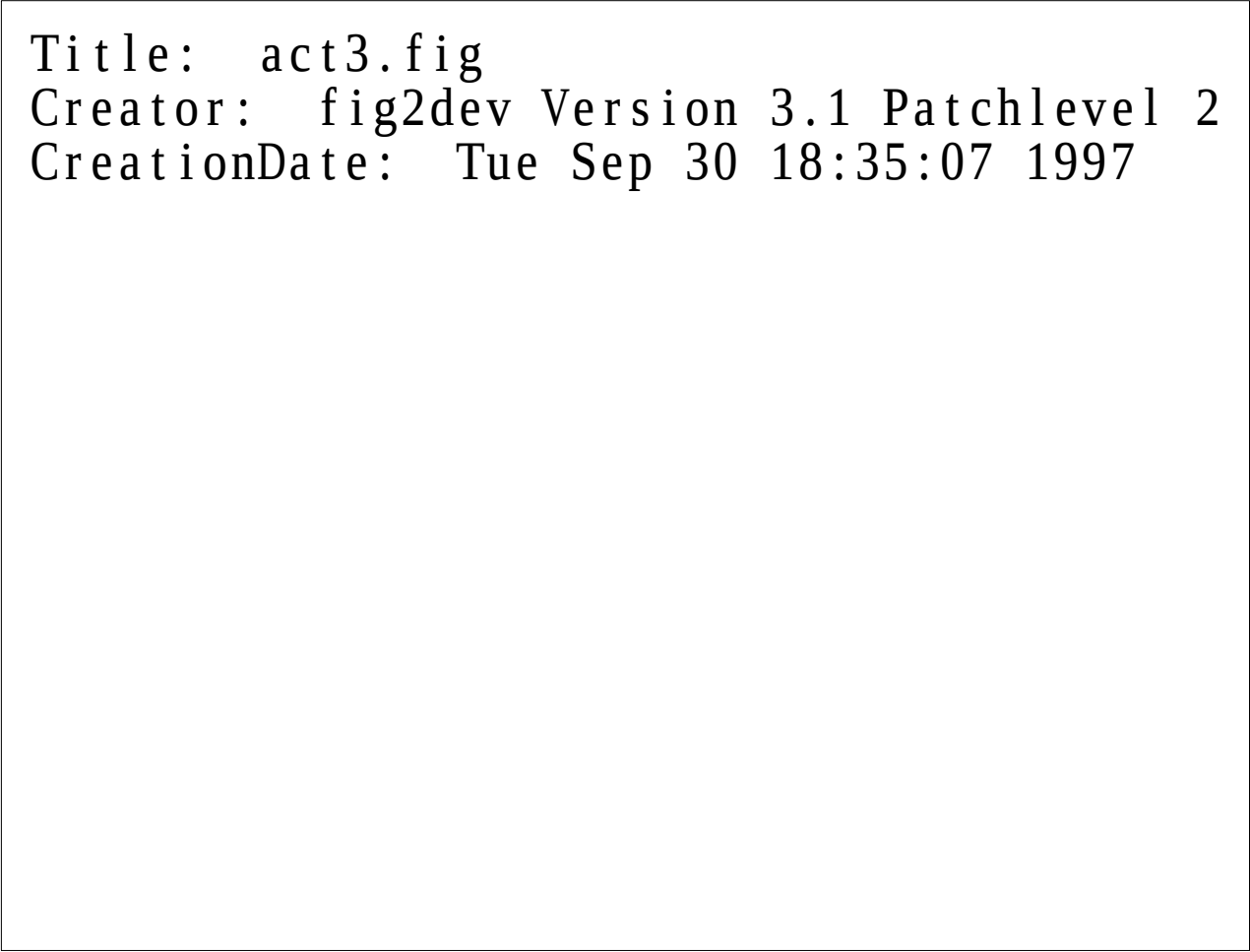
2.3 Usability evaluation in the COVEN project life

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 needs to be 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 this initial stage of the applications, the available prototypes clearly feature a partial set of functions; some bugs remain; documentation effort is underway, with first elements of user manuals available. Evaluation of the initial version of the applications mainly focuses on the evaluation of serious usability problems, with an objective of *cleaning up the design*; a secondary objective is also to set-up a line of work on CVE technology concepts that will further develop in the other evaluation stages,

It is planned that the evaluation of the second version of the applications will focus more on the demonstration of the *added value* of the technology; the evaluation of the final version of the applications, while further assessing the added value, should include elements of a customer-oriented *cost/benefit analysis* of the application concepts.



Title: act3.fig
Creator: fig2dev Version 3.1 Patchlevel 2
CreationDate: Tue Sep 30 18:35:07 1997

Figure 2: Usability evaluation stages in the COVEN project life

Figure 2 provides an overview of the usability evaluation activities (activities 3.2, 3.4 and 3.6) within the COVEN workplan, and of their relationships with the application developments activities (workpackage 2) and the network performance assessment activities (activities 3.3, 3.5 and 3.7).

2.4 Process in the first usability evaluation stage

Our objectives in these initial evaluation activities span three dimensions of work: core application usability, CVE technology scientific inquiry, and a reflection at methodological level.

2.4.1 Evaluating core application usability

Evaluation of the initial version of the applications mainly focuses on the evaluation of serious usability problems, with an objective of *cleaning up the design*. This is achieved through a ‘classical’ usability engineering approach.

Intended users for the evaluation results:

Intended users of this evaluation mainly are the designers and developers of the COVEN applications. This evaluation indeed aims at uncovering flaws in the system design and implementation, and improve our understanding of design guidelines so as to prepare for the coming design iterations. It is foreseen that small-size design adjustments may be performed on the prototypes between two evaluation sessions, depending on the results of the evaluations performed during this period. Major re-designs will only be addressed in the second stage of application developments (second semester 1997).

Target application users and context of use:

The COVEN applications aim at different user groups: general public for the Rhodes Traveller application; business persons in different communities for the Virtual conferencing and 3-D spreadsheet applications. The foreseen context of use for the Rhodes traveller 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.

Focus and quality factors of the evaluation:

The focus of this first stage of evaluation is on functional and information content, functionality of the user interface, and collaboration support; quality factors under assessment are:

- Adaptedness of the application to the intended task domain.
- Efficiency of task performance.
- User problems and design deficiencies.
- Error frequency.
- Learning cost.

The outcome of the evaluation will be an overall identification of advantages and shortcomings of the application design regarding each quality factor, which will be fed into the next design iteration as a basis to improve the applications

Approach:

Inspection methods are particularly well suited to perform a quick and ‘cheap’ clean-up of a design, as described in (Nielsen, WWW). We propose to apply both heuristic evaluation (Nielsen, 1994) and cognitive walkthrough (Wharton et al., 1994) approaches, each approach focusing on different usability aspects of the design: heuristic evaluation focuses on the general usability of a user interface, while cognitive walkthroughs mainly address ease of learning.

Drawing from the HCI and Human Factors expertise of the consortium, usability inspections were thoroughly performed during this first evaluation phase, involving four usability experts. The method and a summary of the evaluation results are described in Section 4 of this document.

2.4.2 Further developing a fundamental understanding of the CVE technology concepts

At the scientific inquiry level, the objectives in this first evaluation stage are to extend the research of some of the COVEN partners through setting up a line of work that will extend and develop as the COVEN applications mature in the further stages of the project.

Claims under focus:

CVE technology entertains a number of specific claims which naturally constitute the focus of part of our evaluation activities. These claims may be summarised as:

- enhanced social interactions and mutual awareness, as a supporting factor for the quality of collaborative work support.
- intuitiveness of the 3-D metaphors (CVE affordances), favouring ease of use and exploratory learning, as an element of low learning costs.
- enhanced ‘presence’ feeling, as a factor to task performance efficiency.
- attractiveness, enjoyability of the CVE user interfaces.

The objective of this evaluation work is to further analyse these claims, to explore their validity and assess the CVE technology aspects that favour or invalidate these claims.

Approach:

As evoked in section 2.1.2, the evaluation of the fundamental CVE technology concepts requires a trial-based approach that can take several forms. We propose to start up an ethnographic/observational component of work to be further developed as the COVEN applications mature over the project length; and to further develop an existing research on the concept of presence through case-controlled experiments.

Observations were performed along with the COVEN network trials that were fortnightly organised as part of the network performance assessment work package, gathering 4 or 5 users within the COVEN application CVEs over 1997. These are described in Section 5 of this document.

In addition to the network trials, a local auxiliary experiment was performed in a case-controlled manner; this experiment and its results are described in Section 6 of this document.

2.4.3 Refining a methodology for CVE usability evaluation

Finally, a third general objective is to use the experience gained in this first cycle of evaluation to further refine our methodological approach. The aim is to further prepare the next evaluation iterations, but also to generalise lessons from the COVEN experience toward the scientific community of CVE researchers.

3. Application Overview

The applications are described in detail in deliverables 2.2 “Initial Citizen Travel Application” and 2.3 “Initial Business Travel Application”. We give a brief overview of both applications in this section and also present breakdowns of the applications into task trees.

3.1 Business Application

User’s initial Goals to attend a meeting that involved presentations, user communication and operation of a spreadsheet

Anticipated users project members and other VR professionals interested in remote conferencing.

Potential users general business people for the purposes of remote conferencing and training

The main goal of the user is to interact with the various objects in the two rooms, and to communicate, and negotiate with the other users in the room. The application consists of two rooms, connected by a door. One room contains a stock-market game which can be played by three simultaneous users, where each user has a personal board in front of them and a view of the two other users and their respective boards. In the middle of this arrangement is a large green bar which provides feedback to the users about the success of their combined manipulations of the game.

Through the door one finds the other room of the application, which is a presentation room, equipped with several conferencing tools to be used by the users. These conferencing tools consist of four blueboards, on which users can type messages which will be displayed on each of the other blueboards, a holoview presentation projector on which slides can be displayed. Slides are available in the room as objects (as of now normal users can not bring slides into the application), to make them visible on the holoview they are inserted in the bandybox, an overhead-projector, shaped like a red bin. The angle of the holoview can be altered with a control device which is also an object in the room.

There are several other objects in the rooms, which the user can not really interact with, apart from picking them up and moving them about. These objects, such as four chairs, a table, and a radiator are (for the time being) there for decoration purposes only. Another feature, a personal text-input interface is always available to the user, and displays the message from one user to all other users in the room, at the bottom of their screen. Each new line wipes out the previous line, the last line will remain in view. Furthermore, the tasks of communication, interaction, and negotiation, are central to this application and presentation by voice and image display is in addition central to the tasks which the user can perform in the presentation room.

All users have to navigate the rooms, and the door, manipulate the objects in the rooms, communicate with each other. Communication involves recognising each other, getting each others attention, exchanging spoken and typed messages, discussion shared objects, and negotiating turn-taking when presenting, or interacting.

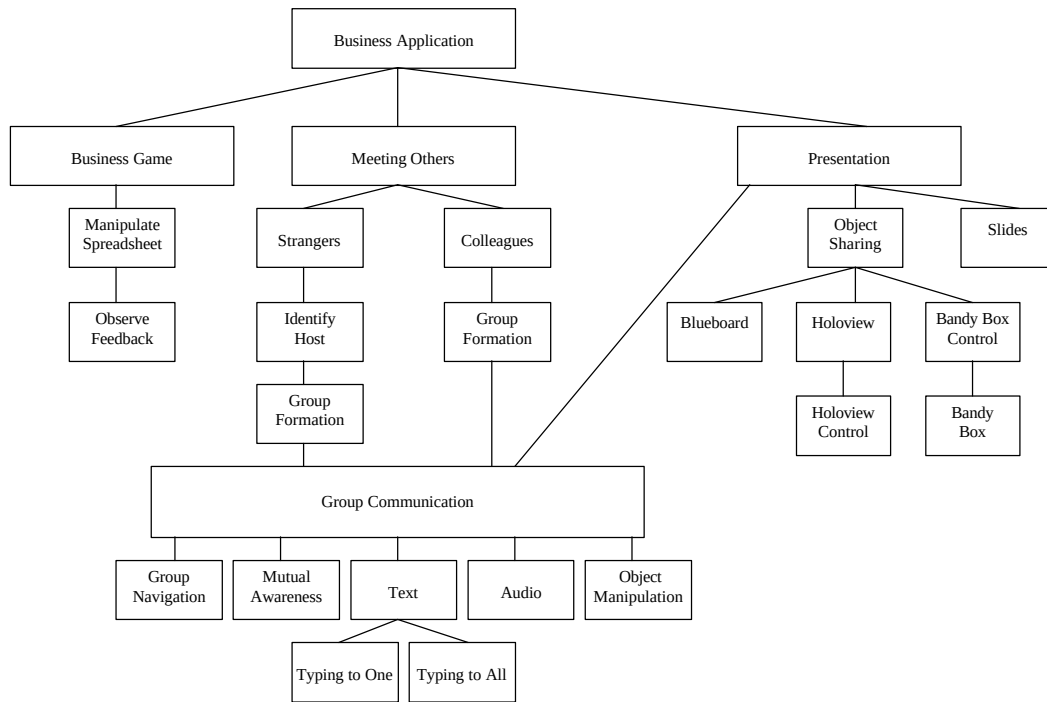


Figure 3: Task Breakdown of the Business Application

3.2 Citizen Application

User's initial Goals is planning a vacation

Anticipated users who will use the citizen application are expert users (i.e. they will be familiar with dVS/dVISE to some extent) in the first place.

Potential users of the application include the general public (travellers), travel agents, hotel operators, other tourist service providers (souvenir vendors, etc.), archaeologists, local experts, etc. The application provides a variety of functionality.

Users are will be able to:

- access large quantities of heterogeneous spatial data in the form of text, images, audio, and (in a future version) pre-recorded video
- experience virtual visits to archaeological sites and (in a future version) hotels
- access on-line information (news, weather forecast, currency rates, etc.) via the World Wide Web (limited integration so far)

The main goal of the traveller is to plan his/ hers vacation. The application, a virtual tourist guide, offers the functionality of any ordinary travel agency plus a lot of extra functionality made possible by using VR-techniques.

The user can retrieve information about a specific destination but also general information like holiday-assurances and packages. Beside the usual flyers with textual information about holiday destinations, the customer can view a slide-show supported by audio or even pay a “virtual visit” to find out what his favourite holiday destination really looks like. Unfortunately it is not possible within this application to book a trip and pay directly.

Another important functionality is meeting, interacting and collaborating with other people. People can enter the

room as group but also as individual looking for a group. Group aspects like communicating, mutual awareness and group navigation then become important.

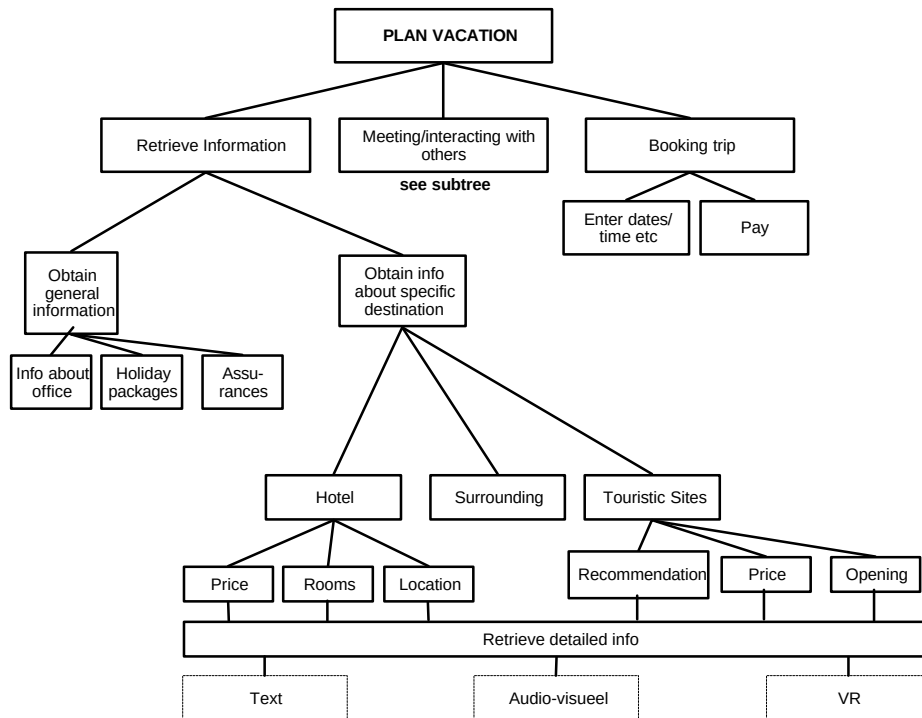


Figure 4: Citizen Application Main Task Breakdown

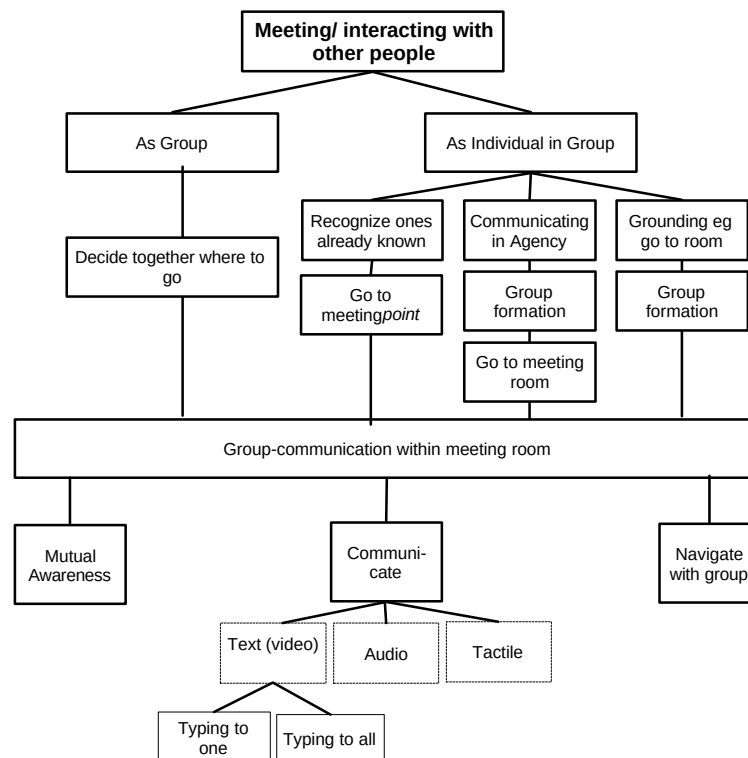


Figure 5 Citizen Application Subsidiary Task Breakdown

4. Usability Inspection

4.1 Methods

In the “Framework for Usability” document which was summarised in Section 2 we identified the Heuristics Evaluation and Cognitive Walkthrough methods (Nielsen and Mack 1994) as the most useful for the application inspection. We outline our application of these methods in this section and give an overview of the logistics involved.

4.1.1 Heuristic Evaluation

Nielsen proposes ten Usability Heuristics as the basic focus of his heuristic evaluation approach (from <http://www.useit.com/papers/heuristic/>):

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

Adjustments are needed on this set of heuristics since some of the definitions above are definitely specific to 2D graphical interfaces (especially and ‘Aesthetic and minimalist design’) or to full-scale products (‘Flexibility and efficiency of use’ and ‘Help and documentation’). We suggest the following adjustments:

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

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

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

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

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

4.1.2 Cognitive Walkthrough

The questions used for the cognitive walkthrough were not changed prior to the inspection. See (<http://www.cc.gatech.edu/computing/classes/cs3302/documents/cog.walk.html>). However revisions have been suggested for the next iteration of the evaluation (see Section 8.1).

- *Will the users be trying to produce whatever effect the action has?*
Are the assumptions about what task the action is supporting correct given the user's experience and knowledge up to this point in the interaction?
- *Will users be able to notice that the correct action is available?*
Will users see the button or menu item, for example, that is how the next action is actually achieved by the system? This is not asking whether they will know that the button is the one they want. This is merely asking whether it is visible to them at the time when they will need to invoke it. An example of when this question gets a negative supporting story might be if a VCR remote control has a hidden panel of buttons that are not obvious to a new user.
- *Once users find the correct action at the interface, will they know that it is the right one for the effect they are trying to produce?*
This complements the previous question. It is one thing for a button or menu item to be visible, but will the user's know that it is the one they are looking for to complete their task?
- *After the action is taken, will users understand the feedback they get?*
Assuming the users did the correct action, will they know that. This is the completion of the execution/evaluation interaction cycle. In order to determine if they have accomplished their goal, the user needs appropriate feedback.

4.1.3 Severity ratings

Given the prototypical nature of the COVEN demonstrators, the severity ratings can be regarded purely as guidance for development rather than actual recommendations for a plan of action. The severity of a usability problem is a combination of three factors (from <http://www.useit.com/papers/heuristic/>):

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

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

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

It is clear that these severity ratings will need to be consolidated at final report integration time (when combining the reports from the different evaluators). Indeed, as noted by Nielsen, it is difficult to get good severity estimates from the evaluators during an evaluation session when they are more focused on finding new usability problems. Also, each evaluator will only find a small number of the usability problems, so a set of severity ratings of only the problems found by that evaluator will be incomplete.

4.1.4 Logistics

Four evaluators undertook the combined inspection, each covering both applications. These inspections were done in July 1997 using versions of the application prepared for a demo at the “21st Century: the Communication Age” exhibition at the European Parliament in Brussels on June 18th 1997. Each evaluator performed the inspection in their own laboratory and independent from the others. It was expected, based on guidelines given in the ELPUB document that the inspection of each scenario would take 1.5 hours with 30 minutes for the actual task performance and 60 for writing up. The majority of the effort was expected to be devoted to the preparation of experimental material, integration of the individual reports in to a consolidated document.

The background of the evaluators is given in Table 1.

Table 1: Background of the Usability Inspectors

<i>Institute</i>	<i>Usability Expert</i>	<i>Background description</i>
THOMSON	Veronique Normand	Usability inspection, HCI
NOTT	Jolanda Tromp	CVE evaluation methodology, Presence research
UCL	Anthony Steed	Computer science, Presence research
KPN	Judith Dijkhuis	Usability expert

The accompanying document “Experimental Plan” gives further information about the work these researchers have been involved in and their particular research interests.

4.2 Results

In order to extract the usability issues, three scenarios were prepared for each applications. These were decomposed into their component tasks and the actions that were required. The scenarios and task lists can be found in the accompanying document “Combined Inspection Report”.

Each inspector then filled in the cognitive walkthrough and heuristic evaluation for each scenario and application. The comments were combined into a single set of tables which form the basis of the Combined Inspection Report. In this section we present a summary of the usability issues found.

In this Section we present summary of the usability issues found. Generally the problems discussed can be broken down into three categories:

1. System problems including lack of functionality, performance and display quality.
2. Interface problems that concern the actions of navigating, and picking of objects.
3. Application specific problems concerning the actual actions and meaning of objects within the environment.

System problems are not often immediately apparent but can result in less than optimum strategies being taken in interface and environment design. For example, perhaps the most serious problem encountered was navigating through the doors in the environment. Though ostensibly an application problem, since the door is simply an object with a behaviour, the problems were exacerbated by three factors. Firstly the fact the door opening caused the scene on the other side to be loaded results a short stall as the relevant information is loaded. Secondly the script to control the door runs on a central server, so the remote clients have to send a signal which delays the door opening. And thirdly, synchronisation problems between the client view, and the collision detection (which also runs on a central server) meant that although the participant might see an open door, they could not enter because they had not got the correct response from the collision system.

Since both the initial demonstrators are for use on desktop systems, it is not surprising that there are some control problems since this is a known problem area. Essentially problems arise because an input device with few degrees of freedom is used to perform six degree of freedom tasks. These can involve problems with interface modalities, such as there being several control modes depending on which mouse and modifier buttons are depressed or the actual

mapping of user motions into three dimensional transformation. These are aspects of the interface that are plainly not obvious to a naïve user,

Control problems thus tend to be pervasive across applications since they are constrained by the capabilities of the underlying VR toolkit. Although the efficiency with which a particular task depends upon the interaction metaphors, there are more general problems with the participant's understanding the purpose of the application components. In the context of VR research this has often been referred to as a problem with the affordances of the objects in the environment. There is a balance to be struck between making the objects realistic in appearance so that they may be recognised, and making functionality apparent to the user. A number of particular problems have been noted in the inspections.

In the following sections, we use the above three categories to classify the problems reported in the inspections. This is partly due to the qualitative differences between the categories, but also because each category of problem is being tackled by a different part of the COVEN project. For instance, system problems can be used to direct the development of the COVEN platform under WP 1, interface problems should be addressed by WP 4, and application problems are the responsibility of WP 2.

4.2.1 System Problems

Under this heading we list all problems which occurred because of system constraints, be they due to existing system design or known deficiencies. Essentially these problems are beyond the scope of the application authors themselves, and are often apparent in many VR toolkits, not just the COVEN platform. Thus in considerations of application evaluation they can be factored out since they are universal problems. Table 2 lists the reported system problems.

Table 2: System Problem Classification

Problem	Description
Application Start Up	The application start up procedure is fairly complex and time consuming. The “slide show” that accompanies the start up is unnecessary. Occasionally a participant will start near the floor rather than at normal height.
Scene load stalls	General problem of the rendering slowing/stopping when new scene components are loaded, either by changing zone, or objects being loaded for the first time.
Centralised Script Evaluation	This leads to a few side effects, such as events appearing to be out of sync in different modalities and slow response to actions such as selection.
Centralised Collision Detection	An instance of the above which occurs fairly often and can lead to confusing effects upon collision when navigating.
Generalised Undo	No general way to undo actions such as selections or navigation if they were accidental. If these have semantic effect, such as playing the CD, this is a distinct disadvantage.
Object Locking	No concept of ownership or indications of action
Selection Highlighting	All objects are selectable, which might lead one to believe they had some function, but can also just look slightly odd, when, for instance, the whole room has been selected.
Independent Selection	Selections are exclusive.
Audio Conferencing	Apparent disassociation between VE events and positions and audio channel.
Unique Collision Detection Reporting	Each object only reports collisions with one other object, which has a number of side-effect, especially in the use of the bandy box.
Multimedia scripting	Difficult to schedule of 2d multimedia presentations, in particular there would appear to be some deficiencies in the control of timed events.
Feedback	Inconsistent feedback on object interaction. .

None of these problems are unique to the COVEN platform, though the wide-area distributed nature of the applications does exacerbate many of them.

Scene loading problems, can be partially countered by efficient modelling (see Deliverable 2.6 “Guidelines for making a CVE”) which minimises the sizes of the texture and geometry information. In particular, using binary object formats such as dVS BGF format, or VRML’s proposed compressed binary format (Taubin *et al.* 1996), will improve both file size and loading and parsing speed. This can alleviate some of the problem by reducing the stall, but this technique only works when files are loaded from local disk copies (as in dVS), rather than across the network (as in VRML browsers, or DIVE). Locally storing files is one step towards caching the required files, and pre-loading them based on predictions of required scene state. Such pre-caching and loading has been done for large scenes with static geometry, and the mechanisms exist in VRML97 (VRML 1997), though browser support for such features is not yet complete.

The centralised nature of certain services with dVS, particularly script execution and collision detection, leads to a problems over wide-area network, since critical signals will be affected by latency. Centralised evaluation provides certain desirable qualities, such as ordering and consistency of scene events. The trade-off is that signals are received after a significant delay, resulting in possibly inconsistent behaviour on the client machine. One example is collision with room walls. Since the collision signal is not received immediately, the viewpoint can pass slightly through a wall before being pulled backwards once the signal is received.

One of the functions developed in the COVEN experimental platform is support for subjective views of the world. One of the problems with selection highlighting (the fact that highlights are visible to all participants) can be solved with such a mechanism. Also since subjective views are supported at the database level in DIVE some forms of object locking can be built in. The other aspects of highlighting, such as the fact that every object can be selected, are built in to the platform. This should be a policy that can be enabled or disabled since is it not desirable in all circumstances. In general it would be useful if the platform or database could enforce stylistic rules on the objects, such as selection only possible if movable, or picking to work generate an audible error as a general policy. At the moment, some such rules can be added but at cost of significant programming effort.

The trials use a separate package, UCL’s RAT (Hardman *et al.* 1995), for audio communication. There is no integration of the audio and the VE display, and this has led to a couple of problems. A tighter integration between the two is required to provide audio scoping, loudness modification and spatialisation. Certain toolkits, including the COVEN experimental platform have integrated such abilities successfully.

One capability of 2D toolkits that is not accessible within 3D toolkits is a generalised undo mechanism. Since the simulation involves a combination of continuous and event traffic, undo mechanisms are difficult to construct except at the database level where each change can be time-stamped and “rolled-back”. Whether this is actually feasible or desirable in a networked environment is questionable since there are many independent sources of events.

4.2.2 Interaction Problems

As discussed in the introduction to this chapter, interaction problems concern the mapping of interaction devices through an interface into the virtual environment. 3D control metaphors can place a significant cognitive load on the user because of the number of actions that must be performed. In particular the user must be able to control the viewpoint and manipulate objects within the environment. Indeed in a desktop system, and in particular in the system used for the evaluations, manipulation of objects is usually broken down into two tasks: selection and positioning. The actual metaphors used for these tasks can affect the application design, so we highlight those particular problems, because they have serious implications for VE design in general. General guidelines for world design within the constraints of the interaction metaphors will be comfounded by the diversity of control metaphors that exist for 3D systems, especially through the wide-spread use of different types of VRML browsers. Table 3 lists the reported

interaction problems.

Table 3: Interaction Problem Classification

Problem	Description
Interface “level”	There can be a confusion about the manner in which an action is performed, at a 2D interface, by a mouse action, by a action within the VE or by using the keyboard. All modes are used at the moment,
Modal Control	Mouse control is inherently 2D, so many modes are used. There is a implicit mode in the area on the screen in which a manipulation is made.
Un-separable Control Dimensions	When moving forwards at speed, the view turns unless the mouse is dragged directly upwards which is difficult to perform
Manipulation Relationship	The cursor does not have to be over the object when moving it, which can lead to confusion.
Navigation/Manipulation Disjoint	Objects can not be held when navigating since both are activated by the mouse. This make it difficult to move objects long distances, or between rooms.
Granularity of Manipulation	Precise object manipulation is hard because of the nature of the dragging motions involved.
Object Manipulation	Moving objects over long distances takes a lot of mode swapping and dragging options.
Manipulation Constraints	Because the direction of object manipulation depends on the area of the screen the drag occurs in, it can be confusing when a constraint is also applied since it is difficult to tell whether the objects is moving or not. Some extra feedback is necessary.
Automatic/Object Centred Navigation	In certain cases objects are being inspected or destinations for navigation are known, automatic or objects centred techniques would seem to be appropriate.
Navigation Velocity	Navigation occurs at a fixed speed, and this can appear slow.
Navigation Constraints	Application of navigation constraints (flying on the level verses free direction) seems arbitrary and is not a user controllable option.
Re-orientation/Re-location Methods	There are no general tools to align oneself to the world, or return to a sensible position. Co-ordinates can be typed in directly, this relies on the user knowing suitable co-ordinates
Selection at a Distance	May not be a desirable facility although it means that certain tasks may be performed more quickly. It might lead to more mistakes since not just objects within range are candidates for selection. It has ramifications for group awareness.
Selection/Manipulation Choice	Not always obvious whether an object has a behaviour on being selected, on being manipulated, or can actually be picked up. This is not consistent, and also prevents some objects being moved, since the perform their action first.
Continuous Navigation Control	The behaviour of the navigation when the frame rate is low can be confusing. The motion can appear jerky anyway, but when the scene freezes unpredictable results can occur.
Multiple Interaction Metaphors	There are three ways of moving, by mouse navigation, form input in the body control, and occasionally in-scene controls. In addition in the citizen application is it possible for another person to move you by operating the carpet.
Direction of Travel	Since navigation is in the direction of look, moving around an object is hard since the user must look away from their focus of interest.
Appropriate Viewpoints	In some situations it is not obvious where the best viewpoint to perform an action is. Especially when manipulations have to be made.

The most fundamental problem with the interaction metaphor as used in the COVEN demonstrators are the multiplicity of layers at which interaction occurs. There are aspects of all of the following in use in the applications:

1. Keyboard input
2. 2D Widget interaction

3. mouse driven 3D control

4. 3D Widget interaction

This is exacerbated by the lack of integration between the displays of the various interaction components.

However this is a fault found with most VE interaction systems. Integrating all controls into one coherent structure is difficult on a desktop display because of the relative difficulty of providing easy to use 3D control, and it is difficult in an immersive system because of the difficulty of providing text work arounds.

Generally it was thought that the 3D control metaphors could be improved. Especially with respect to the number of modes involved when using the mouse. An alternative approach taken by DIVE and at least one VRML browser is to have the modes represented by icons on the screen and reducing the reliance on mouse modes. This allows both object manipulation and navigation with a single mouse press.

The real need is for controlled studies on the best control metaphor for mouse based interfaces. Several techniques exist, including control metaphors for object centred navigation as suggested by one of the inspectors. However previous related research has shown that the best interaction metaphors depend critically on the task being undertaken and there is no one best style (Steed 1996, Steed and Slater 1995). Providing customisable interfaces is a concern that might be also addressed.

Two limitations that should be addressed with some urgency are the navigation behaviour when the frame rate is variable and the lack of any ability to carry an object when using the desktop interface.

4.2.3 Application Problems

Little attention has yet been applied to design criteria for effective virtual environments. There is a conflict between generating realistic worlds, which have been shown to increase “presence” and aid performance in certain tasks (Slater *et al.* 1996) and generating fantastic worlds which explore the possibilities of VE technology but require participant familiarisation. Essentially the conflict is over resolving what objects portray actions within the environment. Making objects realistic endows them with affordances that can not always be satisfied within the environment. Extreme examples include participants enquiring whether they can sit on virtual chairs, but more common examples include participants expecting light switches to work, or tables to be movable.

Problems in this category often involve inconsistency within the interfaces. Although this was one of the original categories in the heuristic evaluations, in VEs it has broader connotations here since consistency must be supported across multiple hosts with different viewing parameters and machine capabilities.

Both applications use the functionality of the core *COVEN services* as described in Deliverable 2.1, 2.2 and 2.3. In addition, there are generic types of object and issues raised in both applications. We have thus have four major categories in the classification: problems inherent in the support of COVEN Services, problems common to both applications, problems with the citizen application, and problems with the business application.

The COVEN services as implemented in the initial applications are:

- Support for *mutual awareness*, with possibly varying degrees. This involves awareness of the presence of other participants, but also recognition of the identity, role, perception of the current activity of the other participants.
- Support for *communication* between participants. The users must be able to communicate with each other using audio and possibly other media including text, video and gestures.
- Support for *resource management*. Resource management also includes that each user has a fair chance of accessing an object by introducing some kind of waiting queues, or, since the “laws of nature” can be different in a virtual world, multiple instances of an object.
- Support for *participants roles and rights*. Users may need certain rights in order to manipulate an object.
- Support for *object manipulation*. Object manipulation includes geometric and edit operations as well as

- object handover.
- Support for *group navigation* within the environment.
- Support for a *global map*.

Both of the applications support the majority of these services in a basic form. Discussion about the breadth of support can be found the relevant deliverables.

Table 4 lists the reported COVEN service problems, Table 5 the problems common to both applications, Table 6 the business application problems and Table 7 the citizen application problems.

Table 4: COVEN Service Problem Classification

Problem	Description
Communication - Message Control	Need to know or find out the name of an object or person before sending a message. There is no error feedback when an incorrect name is entered.
Communication - Message Display	Messages get truncated after two lines. There is no hard limit on the length of a message when entering it nor a signal that the end of the line is approaching. This is exacerbated by words disappearing when the name of the sender is pre-pended to the text.
Communication - Blue-Board Display	Once an object has been copied to the blueboard it can be manipulated and moved away. This potentially leads to several copies of the same object existing in the world which might cause confusion.
Communication - Audio	Many problems exist with the audio since it is not fully integrated with the VE system. There is no feedback about the quality of your own audio when received by others. The RAT controls also seemed slightly un-intuitive.
Communication - HoloView Display	The fact that the HoloView display is flat doesn't immediately suggest that 3D objects can be displayed. The size of objects once presented can be surprising.
Communication - HoloView Control	The remote control for the HoloView is fairly abstract and its purpose is not at all obvious. It is only useful for an immersed participant behind a desk since a desktop participant can select and manipulate the object from any distance.
Communication - HoloView Commands	Text input is hidden, in that it is never apparent when a text message is required or even that it can be done
Object Manipulation - Selection Consistency	All objects are selectable, they can change colour to indicate action, but this must be consistent. Most objects (those without any role in the application) should not be selectable.
Object Manipulation - Picking Consistency	Many objects can be picked in the environment, but this does not always indicate that this has any use or meaning given the semantics of the application.
Mutual Awareness - Actions	There is little feedback as to what the other people are doing in the environment. In particular when a person speaks or moves an object there is no indication from their avatar that they are doing anything.
Mutual Awareness - Avatar	A full body is presented, but it doesn't move very much which might confuse people (they might think their body is "broken"). In addition it would be very beneficial to be able to set ones body colour once inside the application.
Mutual Awareness - Positions	In general with the desktop display it can be difficult to remember the positions of others since there is a lack of peripheral vision. Since peoples' voices are of a constant volume no matter how far away or which room they are in it can be confusing when trying to remember that they are remotely situated.
Group Navigation -Group Formation	The mechanisms for group formation seem to be too mechanical and unwieldy. More transparent support is needed for joining and leaving groups. The role of group leader would seem to complicate the interface issues since they have sole control of the Notebook.
Global Map-	Although the Rhodes Zone itself can be considered a global map since it can be experienced at several scale, there is no overview to provide a context at smaller scales.

Table 5: Application Common Problem Classification

Problem	Description
Orientation/ Relocation Tools	There is a need for better re-orientation tools since it is easy to get lost in the environment. To some extent this can be done in the body dialogue but it is not at all obvious what the required values should be.
Object Collision	Object intersection is preventable, and makes sense for most objects.
Room Functions	It is not obvious once in the applications what or who might be found behind the doors. Some sort of labelling would seem appropriate.
Affordance Issues	The realistic style of the rooms and furniture doesn't hold throughout application. For example the HUD controls and teleporter seem to break the metaphor for the design of the rest of the environment.
Door Opening	Not immediately obvious how you open doors. The choices would appear to be: automatic opening on approach, open on select, open on select and drag, open on rotation handle. It needs to be obvious and consistent. Also the door can slam in your face and sometimes you get through the door only to be "pulled" back.
Object Consistency	Related to affordances, but in particular the maintenance of object class distinctions, door sounds are consistently applied.

Table 6: Business Application Problem Classification

Problem	Description
Application Context	The business application starts up initially in the room with the business game. If there are undertaking a conferencing task it might more sense to start in the other room. A 3 rd "neutral" room (such as the VTA in the Citizen application) might be appropriate.
Business Game Rules	There is no information in the business game room about the objectives or rules of the game.
Desk Choice	In the conferencing room there are several desks and no hint whose is whose or even if this is important.
Bandy Box Control	The bandy box controls are complex and its capabilities are not obvious. There is little feedback when facing the box on whether an error has occurred or the object was successfully sent. It would seem to be impossible to reset the Bandy Box to show no object. Reliance on text controls means that it is impossible to use with immersive display.
Slide Control	The sequence of actions to operate the slides is not obvious. Focus must be indicated by selection which is not obvious. Also not obvious which objects are actually slides and which are plain objects.
Slider Bars	The handles on the slider bars are difficult to see. There is no feedback when limits of slider bar movement are reached.

Table 7: Citizen Application Problem Classification

Problem	Description
Remote Control	The functionality of the remote control is a little obscure. It seems to duplicate some functionality available elsewhere. The control clutters the display somewhat, especially when a notebook is added.
CD Selection	The CD titles are hard to read. The layout would seem to limit the number of slide shows that can be supported.
CD Player	The controls on the CD player are hard to see and activate. Not obvious that selecting the CD automatically plays it rather than a button control on the player.

Slide Show	A video tape would seem more appropriate if pictures are going to be displayed. There does not appear to be a way to pause or wind back the presentation once it has started. There are also problems with synchronisation between video and audio.
Teleport Buttons	The controls for the teleporter are not obvious since they are unlabelled. Indeed one has to open the teleporter first before these controls become apparent.
Teleport Machine	No hint of the teleporter's function or mode of operation. It is not obvious that the participants must stand inside the teleporter. Indeed this is hard to gauge on the desktop system. Error conditions aren't handled yet.
Notebook	The functionality of the Notebook is not obvious. It has several buttons and a window that seem to be inoperative. The user will probably not expect the Notebook to attach to the visor upon select. If one then tries to move the Notebook in order to reduce screen clutter it is dropped. Some functionality is duplicated with remote control.
Carpet Controls	The carpet controls have two main drawbacks. The participants can be moved independent of their own controls, and they might suddenly find themselves in mid air when they are afraid of heights.
Site Icons	It is not obvious that the site icons have to be clicked upon to enter. This is a third transformation type control (c.f. walking through doors or teleporting). The user might expect more information before teleporting.
Hotel Icons	The hotel icons are presented without a reference scale in order to compare or look up information.

Again many problems stem from the different control modes that are used in different situations in the applications. For example BandyBox employs two very different control types: text to specify the observer by name, and collision to specify the object to copy. An argument could be made for providing only in scene 3D controls for this object, maybe by selecting the observers on a list, by scoped distance or by group icon.

In general gracefully applying several styles of interaction requires strict demarcation of the roles each can play in the environment. For example, text input might solely be used for message contents, the message addressee could be specified by graphic control, or indicative gesture, and this would be applied consistently across all objects that require some form of user interface.

Another major problem area is the lack of guidelines or constraints for the presentation of widgets like objects within a 3D space. Since interfaces can be built solely from collision and selection information, there are few constraints on what objects can have behaviours attached to them, and thus little consistency between objects. For example, the citizen application takes a literal approach for the interface to the CD Player (the CD Player has an active stop button on a textured panel that looks like a standard CD player), but a more abstract approach for the design on the teleport.

This can be thought of as one facet of a more general problem of the implied affordances of objects. As described in the introduction to Section 4.2.3, there is a conflict between making the objects realistic in appearance so that they are easy to recognise, but then having to support interfaces that do not always have a real-world equivalent. For example, the teleporter might have a real world analogy of an "express lift" but its appearance and behaviour does not immediately suggest this. Conversely the Notebook is recognisable as a ring bound book, but its behaviour is as a button holder and token of group leadership.

Reports of affordance issues are covered by the cognitive walkthrough question "*Will users be able to notice that the correct action is available?*". The realistic appearance of objects makes this an object recognition task in most cases, and unless appropriate hints are given for objects that have unusual behaviour, it is obvious that tasks will be complicated by the representation. This then poses the question of how best to provide help to the user when confronted by an interface they can not recognise. The approach of giving audio hints for the function of "remote control" is appropriate when the user's focus is directed at an object. However in a room sized environment with many objects and users audio descriptions might confuse unless focus of interest can be secured. In a 2D window interface this is

easy, but in a 3D environment where freedom of movement and viewing is allowed, focus is difficult to be sure of unless interaction takes place, and this interaction might not take place, unless it is obvious what the object is for and whether it is relevant.

One approach to this difficult problem is to include human agents that can demonstrate functionality in place and in context. The ability for an autonomous agent or another user to demonstrate functionality would thus seem critical for these applications. Unfortunately the only information the current avatar conveys is placement, direction of view and user name. It is not directly possible to see which object a user has selected, or which one they are moving because the standard user interface does not animate the arm, and in any case there is no limit to the distance from which an object can be selected or manipulated.

5. Network Trials

21 network trials have been held to date to using the COVEN platform and most of these have used the COVEN applications. At this stage the trials were mainly an exercise in sorting out networking problems and testing the stability of the platform and applications. As such the participants in the trials were expert users since they needed to be able to solve networking and software problems during the fortnightly or sennightly meetings. We concentrate in this section on results about the actual applications. A fuller description of the results of the networking trials including analysis of resource load, application stability and networking set-up will form the basis of the upcoming deliverable D3.4 “Networking Assessment of the Initial Applications”.

Comments about the applications are drawn from two sources: questionnaires administered after each trial, in which the participant described the tasks undertaken, problems encountered and opinions on the applications themselves, and two in-depth studies of tasks within the COVEN applications. Section 5.1 describes the network set-up and parties involved in the trials. Section 5.2 describes the use of the applications in a network setting. Section 5.3 outlines the tasks undertaken at each meeting. Section 5.4 summarises the results of the questionnaires completed at the end of each trial and Sections 5.5 and 5.6 report on the two in-depth trials with the COVEN applications.

5.1 Network Infrastructure

Five partners have taken part in the trials to date. Figure 6 shows their geographic positions.



Figure 6: The different sites involved in the COVEN project, with the sites involved in the network trials represented with red dots

All five sites have taken successfully in the trials, see Section 5.3 for the actual list of participants in each session.

5.2 Trial Platform

In the network trials the COVEN demonstrators are used alongside other collaborative applications that support audio and textual communication. The COVEN demonstration platform does not currently have integrated audio

support, so the Robust Audio Tool (RAT) from University College London was used (<ftp://cs.ucl.ac.uk/mice/rat/current>). The demonstrator applications do support textual communication, but because the applications were being tested and could not be relied upon to work 100% of the time, a back-up text channel based on a modified version of the ytalk (<ftp://ftp.cdrom.com/FreeBSD/packages-current/All/ytalk-3.0.2.tgz>) software was provided. Figure 7 shows a typical layout of the resulting windows.

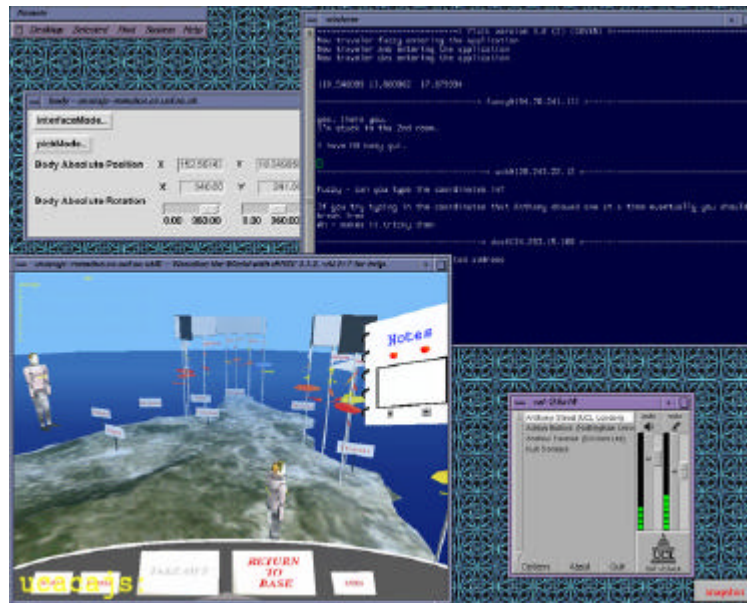


Figure 7: View of the Rhodes Zone of the Citizen Application during a network trial session

Clockwise from top left the windows are:

- ytalk session with 4 participants
- RAT interface showing the names of the participants. The highlighted name indicates that that person is speaking.
- the dVISE window showing a view of Rhodes with the hotel icons and two visible participants. The participant whose view is shown is the group leader, so they have the Notebook attached to their view as can be seen in the top-right.
- the dVISE body control window.

Each of the network trial participants would have a similar view, though the layout of the windows is not constrained in any way.

5.3 Trials and Tasks

The tasks undertaken at each trial were concentrated initially on setting up and running the default dVISE “tutorial” world. Once the networking was stable they moved to the Business Application initially and then to the Citizen Application. Finally a single trial was held to test the feasibility of moving the Network trials to the latest version of dVS/dVISE. The 21 trials to date are summarised in Table 8.

Table 8: Summary of network trials to date

Date	Duration	Participants	Activity
10 12 96	60 mins	Nott, UCL, Div, TNO, KPN	dVS3 Tutorial
08 01 97	60 mins	Nott, UCL, Div, TNO, KPN	dVS3 Tutorial
22 01 97	30 mins	Nott, UCL, Div, TNO	dVS3 Tutorial
05 02 97	30 mins	Nott, UCL, Div, TNO	dVS3 Tutorial
12 02 97	90 mins	Nott, UCL, Div, TNO, KPN	dVS3 Tutorial
19 02 97	60 mins	Nott, UCL, Div, TNO, KPN	Business
05 03 97	90 mins	Nott, UCL, Div, TNO, KPN	Business
19 03 97	60 mins	Nott, UCL, Div, TNO, KPN	Business
02 04 97	60 mins	Nott, UCL, Div, TNO, KPN	Business with 4 people Citizen with 3 people
16 04 97	80 mins	Nott, UCL, Div, TNO, KPN	Business
07 05 97	105 mins	Nott, UCL, Div, TNO, KPN	Business
14 05 97	60 mins	Nott, UCL, Div, TNO	Business with 4 people Citizen with 4 people
28 05 97	120 mins	Nott, UCL, Div, TNO, KPN	Business
04 06 97	50 mins	Nott, UCL, Div	Business
25 06 97	45 mins	UCL, Nott	Business
09 07 97	90 mins	Nott, UCL, Div, TNO, KPN	Business
16 07 97	90 mins	Nott, UCL, Div, TNO, KPN	Business Trial (§ 5.5)
23 07 97	135 mins	Nott, UCL, Div, TNO, KPN	Business and Citizen
06 08 97	100 mins	Nott, UCL, Div, TNO	Citizen
03 09 97	120 mins	Nott, UCL, Div, TNO	Citizen Trial (§ 5.6)
23 09 97		Nott, UCL, Div, TNO	dVS4 Demos

One further networked session took place as a demonstration for the “21st Century: the Communication Age” exhibition at the European Parliament in Brussels on June 18th 1997. This involved several hours of demonstration within the applications, where they proved very stable.

5.4 General Experience

After the 21st network trial, participants were asked to complete a short questionnaire to get a feel for how the interfaces for the various applications were working and how best to support collaboration. These did not include any questions about the applications themselves. Initial evaluations of the applications for the networking trials are presented in Sections 5.5 and 5.6.

Specifically questions were asked about the components of the interface and the major tasks that must be completed in the course of a collaborative session. Figure 7 showed the main components of the applications during the network session. These are: ytalk text window, RAT interface, dVS 3D scene and dVS body control dialogue. Various other windows would be used including SGI audio panel, web browsers and shell windows, but these were usually to perform tasks that did not involve collaboration. With the exception of the dVS body control dialogue, the other three main windows support collaboration and some sort of representation of the other participants.

5.4.1 Communication

Participants were asked their opinion on the reliability and usefulness of the three communication media: text, virtual environment (VE) and audio. In general the audio quality was thought to be good, though it would vary greatly depending on local or remote network load. It was thought to be essential for collaboration since it was easier to express emotion and since it did not distract from the visual cues presented in the VE. However, in common with the inspection, one participant did find it disturbing that the audio's being mono meant that it did not correspond to the relative spatial position in the VE. Normal conversation was easily supported since drop outs were rarely serious.

However when data such as IP addresses or URLs were being communicated the text channel was essential.

‘The audio is fairly reliable, but it can’t be relied upon as the sole source of communication.’

Given limitations of audio equipment, most participants had to leave RAT muted most of the time. That is, it was operated on a push to talk mode. This has a small cognitive overhead and occasionally meant that phrases had to be repeated since the participant had forgotten to turn the audio on.

‘[T]here was no means of adjusting the minimal floor noise level (i.e. a squelch) I was forced to continuously use the mute button to push to talk because the headmounted microphone would pick up too much noise and end up transmitting continuously.’

Generally everyone found the text channel to be stable except in the fact that it showed up improper networking configurations to a greater extent than RAT or dVS. As noted before, people found it essential for passing data because although the message might be delayed, it was guaranteed to arrive. Also useful were the abilities to cut and paste text from the ytalk window into other programs, and the persistence of the messages on the screen for a few minutes until they scrolled off the top of the screen. Given the perceived robustness, the text channel was regarded as essential as a back-up in case the audio or VE could not be initialised.

‘The text channel is the safest way to ensure that all the users can receive a message.’ And ‘The text channel, in my opinion is the best means of communication. It takes a minimal amount of network bandwidth and allows an unambiguous means of communication.’

‘The text channel is a godsend! It is extremely reliable and is often used in preference to the audio.’

‘It provides a mechanism to convey information to the other people, and you are sure that they see what you are typing and know what you want to say. With Audio there is no certainty that you have been heard. With the text channel you see exactly what everyone else sees. It is also a good way of exchanging URLs or other such information which can be cut and pasted. It is low bandwidth and provides very useful support to audio.’

The VE was considered to be reliable once it had been started. Starting was not always easy since it would test the networking and bandwidth available. Only a few glitches were experienced and most problems were to be found with the behaviour of the applications as described in the subsequent section.

‘Apart from some bugs in the zone switching within the application, I’d say that dVS is rather stable between us four.’

Although the communication channels were independently successful, the next question highlighted a couple of problems with using all of them at the same time. Notable amongst these was the lack of screen space in order to display all windows simultaneously. Most kept the body control window open since it was useful for re-orienting the body and resetting positions, and the RAT, dVS and ytalk windows were all used constantly. Attention was thus divided between 3 or 4 windows. The main reason for keeping the audio and text windows open was that this was essential in order to be able to tell who had typed a message or who was speaking since this feedback was not available in the VE.

‘Ideally you should only use one window for the application. Now we have three communication media, that are three different applications, you must have three windows open all the time. This is a bad thing, because you want to have the graphics window full size and you don’t want to worry about the other windows. I didn’t use the toolbox.’

‘I would only occasionally use the body dialogue, so that was iconised frequently. I occasionally iconise ytalk

if I need to pop up another window for some diagnostics since it takes up a lot of space.'

5.4.2 Interaction Metaphors

Four questions queried the usability of the interaction metaphors for navigation, object selection and manipulation since this was an area that provoked many comments in the inspection. Most found it relatively easy to control the VE once they had become accustomed to it, but the two who reported that they found it “a bit difficult” were reflecting their opinion that alternative metaphors would be easier to learn and more efficient to use.

Comments were made about the inconsistency of object reactions to selection or manipulation. In particular there were problems using the doors due to the fact that only one person could select an object and had to specifically relinquish that select by deselecting it or selecting another object. In combination with the fact that the door opening time was quite short and not precisely co-ordinated with the collision detection, it lead to the participants creating a “door policy” where an ordering for passing through the door would have to be agreed so that each person could pass through in quick succession.

'Selections are a problem. Anyone can manipulate your immersive control panel. I would like to be able to see who is manipulating an object. The concept of groups and group leaders needs a re-think I believe.'

5.4.3 Other Services

One specific change that was made mid-way through the trials was the addition of body colours to aid the recognition of the other participants. In the original set-up each participant had a name tag, but it wasn't always readable from the side or at a distance. Although the colours chosen were not universally appreciated, all but one participant found that their addition aided recognition. This might be because although the relationship between colours and participants might not have been remembered, it did make it easier to track moving people. Two participants expressed an interest in having even more control over the appearance of their avatar, especially since the geometry for body and face is the same for each participant.

'Yes because the text labels are hard to read anyway, but are impossible to read from the side. Because the new bodies are so bright it makes spotting people a way off very easy, but then they are bright to the point of being ugly.'

'No. For me the name suffices. My feeling is that there IS an enhancement, if you can use different geometries for different persons.'

'[...] It would be cooler if we could change trouser colour and shirt colour independently [...].'

The final questions asked for requests for future extensions. Two of the participants requested a more advanced teleporting system to allow them to switch between zones.

'[...] Fast teleportation to a given zone [...].'

One suggested an automatic door system that would move you through the door once you had indicated you wished to go through.

'[...] automatic transport mode (e.g., click on an open door and you are guided through it) [...].'

An important suggestion was that the application designers constrain objects so that no one participant could arbitrarily edit their properties. At the moment all objects can be edited so that, for example, the walls of the rooms can be deleted or re-coloured.

There is too much scope for chaos at present. There should be a set of interactions that a particular type of participant can perform. This not only save the VE from being disrupted, but also prevents anxiety on behalf of the user as they are less likely to feel worried about what they do in the VE.'

Overall the participants found the applications easy to use after the first few trials. Future work would be needed to enhance the integration, both in presentation and functionality, between the different communication channels.

5.5 Business Application Experiment

The purpose of this experiment was to evaluate the effectiveness of the communication services provided within the business application. The participants were to undertake a simple negotiation task that involved them using all the communication channels available. Five sites took part in this trial as shown in Table 8.

5.5.1 Task

The task involved the participants trading cards with the aim of obtaining a specific set of cards. Each participant started off with four cards which they had to trade one by one using any of the communication channels in order to set up the swap. It was possible for each participant to reach their target given the distribution of cards, so any one of the participant's completing would not halt the game. One of the participants acted as an umpire. The umpire's responsibility was to record transactions that took place since there were no actual tokens being passed around.

The game was scheduled to last 10 minutes, though it took the best part of one hour on the day in order to set-up the network system because of a file server crashing. Feedback was obtained using a web based questionnaire which elicited comments on ease of use and manner of communication between the game participants and the umpire.

5.5.2 Results

There were effectively three different manners of communication: RAT audio communication, ytalk text and dVISE text to all and dVISE text to named participant. Each participant was potentially fully aware of the communication taking place with the exception of dVISE text which could be sent to a named recipient instead of the whole group.

5.5.2.1 Effectiveness of Text

Although there were two distinct text channels, they were lumped together by the participants since they had similar characteristics. Most communication was made through text because of its perceived reliability and the fact that a message would "persist" (i.e. be readable on the screen) for a short period. In particular messages typed on ytalk were likely to be visible for several minutes, whereas messages made through dVISE would remain on the recipient screen until another message was received.

[I communicated] Mostly using ytalk, because I could guarantee that every member would be able to see what I had so say.

Generally it's [the audio] OK, but it's good to have the text channel to fall back on where clarification of details can be found.

Because it took longer for a text message to be entered, one participant made a point of asking the question via audio first and then recording it in text. The audio would thus draw their attention to the message which was then recorded in case it was not received.

I used rat, [and] ytalk to confirm what I had said in rat and make sure everyone was aware of what was going on.

One difficulty with the dVISE messages was that it was easy to enter an incorrect name resulting in a message that was intended to be private being sent to the whole group.

So when I wanted to address just Kurt, and typed Kuurt: do you have any Ds? this message would be broadcast to all anyway.

5.5.2.2 Effectiveness of Audio

Opinions on the quality of the audio differed between sites, and this may have affected the reported usefulness. In particular one participant commented:

I've noticed that I hardly use rat because I'm never sure whether people will catch the entire sentence. If everybody wants to trade together over the audio channel, you're bound to get confused.

If a number of people are communication via rat then there is contention for the audio facility, it very easy to talk on top of one-another.

There are two problems expressed here: lack of confidence in audio reception and congestion when everyone talks at once. These are both supported by the following quotes:

If a number of people are communication via rat then there is contention for the audio facility, it very easy to talk on top of one-another.

I noticed that the letters B and D get muddled very easily over the audio line. Again it [using ytalk] was the only way to ensure complete unambiguity.

The second quote above indicates that the participants would fall back on the text channel when the content of the message was important. The audio channel would thus seem to be useful for short comments or requests for confirmation.

Using rat was useful when you needed to say something quickly or something in passing but rather than repeat myself asking the same questions using rat it was easy to prompt with a [text] question and leave it on the [ytalk] screen to wait for a reply

No matter how the deal was set up [...] it was usually confirmed by voice since it is difficult to get the same rapid response by text communication.

This immediacy of communication would seem important for the smooth running of the task, but more fundamentally participants found it more expressive.

However, rat offers real live audio communication which means you can express yourself (for example, share in humour, laugh, grunt, hum, etc.) things that lose there meaning using the text channels.

5.5.2.3 Usage of Body Positioning

Although not a direct communication media, there was information inherent in peoples' spatial layout in the environment. Seeing two people standing close together might imply that they were involved in a conversation.

[..] I think a couple of personal messages got lost, because they [the two recipients] were engrossed in another (private) conversation, and ignored them. It might appear rude to interrupt, but there wasn't any indication that they were talking apart their standing near each other.

I felt that I had to stand next to the person I was negotiating with. But I noticed that I was almost the only one.

However just position is neither a necessary nor sufficient condition for two people to be in conversation. It is possible for two people to have a private conversation via dVISE text without standing close to each other but on the other hand, two people standing side by side might be involved in global chat which is apparent to all participants.

Such body position cues might aid or hinder, but they definitely do have an effect on behaviour.

I used ytalk to communicate to Adrian since I wasn't usually near him, since he stood in the corner.

5.5.2.4 Other Issues

One of the problems found in undertaking this trial was the lack of integration between the ytalk text and the dVISE text. The advantage of the dVISE text is that it is integrated into the environment display, so interaction or focus on that window does not have to be broken.

[..] so if you want to use ytalk, you first have to select the ytalk window, and then type your message, while typing and moving in the VE goes directly.

I also occasionally used the text messaging in dVS, but sometimes this was only because my mouse was in the wrong window! (I had intended to use ytalk).

The problem of lack of information about participant interaction was again brought up by this trial. In this case the specific problem was not being able to tell who was speaking or who they were addressing. Having this facility would have avoided the situation of not knowing whether you were about to interrupt a private conversation.

In short, there were little clues who was communicating with who (since e.g. visual info was little help)

Having this facility would have avoided the situation reported in the previous of not knowing whether you were about to interrupt a private conversation.

5.5.3 Conclusions

Two participants managed to complete the task within the allotted time so the communication was certainly sufficient for the task. All the different channels were all used, though for slightly different purposes. The dVISE text to named recipient was useful for people concerned about privacy of communication. It had the advantage that it would appear in the main window and would be easily observed. The ytalk window would eventually get quite cluttered with messages so although it was useful for long, persistent messages it did not have the same immediacy. The audio was found to be less than ideal because of congestion and quality problems. However it added another dimension to the communication because it would engage attention immediately.

The problem of integration between the channels was raised also again. This particular task involved the participants using all parts of the interface (see Figure 7) which meant that attention was divided between several windows and mistakes were made with window focus. One immediate integration would seem to be to make the dVISE text entry more flexible and give it more of the capabilities of ytalk. The most important enhancement would appear to be to allow review of previous messages.

5.6 Citizen Application Experiment

The basis for this experiment was the collaborative undertaking of an information retrieval task in the Citizen Application. Four sites took part in this trial as shown in Table 8. The results are determined from a questionnaire undertaken after the task.

5.6.1 Task

The subjects are presented with a task of approximately one hour. Afterwards they are requested to answer a questionnaire of approximately half an hour. Human behaviour and experience was captured using questionnaires, and video-recordings were made. The responses to the questions were collected using web based HTML forms. The questionnaire consisted of attitude statements with Likert-scales, and open-ended questions. The collection of data involved constructs such as satisfaction, experience, ease of use, group involvement and awareness. In addition the questionnaire was used to ask questions about the results of specific decisions made about user needs, while creating the COVEN services, and about the subjects opinions on possible additions to the COVEN services. .

The subjects were given a task, which consisted of steps to be taken through the application to test all available functionality in a logical order. The task is a slightly amended version of the task-tree developed for the Inspection (see accompanying document “Combined Inspection Report”). All subjects proceeded to follow the task description, but ran into problems with bugs in the application several times. For this reason the task took much longer than anticipated, but the subjects continued the task to the end. The questionnaire was answered by all four subjects within the next week, and one additional subject answered the questionnaire based on previous experience with the application. For the statistically reported results the answers from the last subject are not taken into account, for all opinion based answers, the answers from the last subject are included amongst the quotes reported here, when relevant.

5.6.2 Results

The subjects were asked a large number of questions about their satisfaction of various COVEN services. These results are described and discussed below.

5.6.2.1 Effectiveness of Meeting Room for Communication

The subjects were asked to rate the effectiveness of the meeting room for collaborative communication on a scale of 1 to 5, where 1 = ‘Not at all effective’ and 5 is ‘Highly effective’. All respondents gave the meeting room a score equal to ‘Effective’ ($M = 4$, $std = 0$, $n = 3$). The subjects were subsequently asked to explain why they gave this score. The opinions are presented below:

‘I think it definitely tightens the group band, because all people interested in going to Rhodes, eventually come to that room. This is the place where groups are formed.’

‘You can talk using audio, there are personal blueboards as well as public one on the wall, and if prepared in advance there is the ability to project slides. All you really need to communicate.’

‘You can do all the basic things, except leave mail messages or stick notes on a bulletin board.’

‘Some of the group functions don't work, but apart from that it is good. I only didn't give it a 5 because ideally you would integrate the 3 views totally, and include other collaboration tools like a white board.’

5.6.2.2 Effectiveness of CD Slide Show

The subjects were asked to rate the effectiveness of the CD slide show as a collaboration mechanism, on a scale of 1 to 5, where 1 = ‘Not at all effective’ and 5 is ‘Highly effective’. All respondents gave the same score, equal to ‘Ef-

fective ($M = 4$, $std = 0$, $n = 3$). The subjects were subsequently asked to explain why they gave this score. The opinions are presented below:

'Effective, but in a broadcast manner. There is no ability to interrupt, pause or replay the show, and everyone's audio gets drowned by the voice-over. Maybe it would be better if the lights dimmed as the CD started playing, suggesting that normal interaction should be suspended for the duration of the presentation and resumed afterwards. Either that or else make the presentations interruptable.'

'You get information, but the info don't last. (Slides disappear again, are not annotated by names, comment is heard but cannot be reread)'

'Some problems (stopping/pausing) the application. I would really like to be able to collect some slides into my notebook and thus form a tour for me/my group.'

'Effective in a way, that you can get information while just sitting back and relax, but it tends to get a bit boring after a few seconds.'

5.6.2.3 Effectiveness of Magic Carpet for Group Navigation

The subjects were asked to rate the effectiveness of the group transportation mechanism, the 'Magic Carpet', on a scale of 1 to 5, where 1 = 'Not at all effective' and 5 is 'Highly effective'. The respondents gave the magic carpet a varied score, with an average equal to 'Not very effective' ($M = 2.3$, $std = 1.2$, $n = 3$). The subjects were subsequently asked to explain why they gave this score. The opinions are presented below:

'The magic carpet as a metaphor for group transport is a good one, but as it is now, it looks absurd. The carpet is attached to the viewpoint of the group leader, thus he will never be able to see the carpet. The others will see the carpet through the body of the group leader. I thought the solution used by Gidi in the application made for the Brussels demo, was better. There everybody was able to move around freely and the carpet was controlled by a remote control box.. The carpet as it is now, is only workable for one person.'

'Non volitional motion is a no-no IMHO. It can lead to simulator sickness. You can get round this by making the vehicle Surrounding (so the visual flow below you is of the vehicle, and not letting people get out. Flying in the sky is bad anyway for people who get vertigo, but this isn't apparent on a desktop because everyone looks as if they are huge and actually standing on the ground (another argument for a large floored vehicle?))'

5.6.2.4 Effectiveness of 3D Rhodes Map

The subjects all agreed that the 3D map of Rhodes is a useful aid for collaborative exploration. Although several suggestions were made to improve the usefulness.

'It is definitely a basis for collaboration and discussion, but then you would have to have more detailed information about layout and roads, in order to get discussions like "let's go for this hotel, because it's closer to the beach while still not as near to the noisy center of the town as the other one". I think the 3D map should be represented in different levels of detail: on the level of the whole island, you are interested in the main towns, main roads, main sites of interest. On the level of an individual town, you are interested in hotels and sites and (smaller) roads close to that town.'

'[...] it is useful to see where others are looking and what they are accessing even though this should probably be in a subjective view if they want it to be.'

5.6.2.5 Effectiveness of Hotel Visualisation

The subjects were asked to give their opinion on the hotel visualisation scheme. In many ways this scheme has not

been fully implemented yet as originally planned. The subjects are aware of those plans. The answers are reported below because they give a good first impression as to the effectiveness of the implementation.

‘OK, you can get all sorts of information there, while not getting overwhelmed by it.’

‘Ineffective. The icons used are not intuitive, and e.g. the symbol for money is only relative. ("this one has a smaller pile of money than the other one, so it is cheaper, but is that 1 guilder or 100?"). One would want to see more aspects of a given attribute and be able to select and compare different hotels. E.g., now you can have a picture of the beach near a hotel, but you also want to know the distance to it. What is the name of the beach? (Have I been there before?)’

‘IMHO everyone needs a notebook with a list of the icons and their meanings on it. The icons are OK, but it is not obvious which ones change the picture and which bring up web pages.’

‘I’ve only used this for ten minutes, and that was yesterday. The ability to see a picture on the flag pole and cycle through is was good. I didn’t get a Netscape up (they all went to Anthony, lucky man) so can’t comment on that info.’

5.6.2.6 Effectiveness of Archaeological Information

In general the opinion about the effectiveness of the visualisation scheme of the archaeological sites of Rhodes is that is interesting, although several reservations were made:

‘Does this mean the pillar on the global map of Rhodes? I am not aware of any data visualisations scheme within the archaeological site itself. I think it is good. It differs from the one used for hotels, and gives an impression that it represents an archaeological site.’

‘You can only see that there are sites, not what kind’

‘The view is interesting, but I would like to see a similar model of what it looks like now.’

5.6.2.7 Model of Reality verses Model of History

The subjects were asked whether they prefer to see a historical model of the archaeological sites represented in the application, or a model of the site as it is today. The unanimous opinion was that it would be best to be able to see both. Rationale and suggestions are given below:

‘Seeing the site as it is today gives the user an idea of the holiday they may embark on. Seeing the site as it used to be arouses interest in the holiday itself.’

‘My ideal view would be one where I could see the building in its heyday and then see it decay or change brick by brick :-)’

5.6.2.8 Expert Opinions on Development Assumptions

During the development of the COVEN Platform and the COVEN services assumptions had to be made continuously regarding the design. For instance, there is a trade off between speed of performance of the application and level of detail of the graphics. Each designer had to constantly make decisions about this while creating (parts of) the application. Another assumption which had to be made is the kind of interface tools the users would need. Finally, the most obvious assumption which is being made is that a virtual travel agency can provide added value to the user as compared to the services of standard travel agencies.

When asked to judge what is more important for a multi-user travel application, high quality graphics or a high performance application, the subjects unanimously agreed that high performance is of first importance. The rationale

for this opinion is given below:

‘The detail is somewhat superfluous - if I need to see the high detail texture maps I would get closer to them. Much of the detail seems lost because it appears as small icons.’

‘[...]if performance allows, you definitely should have high detailed graphical objects.’

‘It should have adequate geometries and textures, but I think performance is the key issue.’

‘Slowness very soon gets real irritating.’

5.6.2.9 Added Value of VR Tourist Information

When the subjects were asked to explain if they perceived any added value of a virtual tourist application as compared to standard sources of tourist information, the answers were varied and all interesting. The most elementary view given is that if the application offers the same kind of information as standard brochures is can never be worse, but also a virtual view of the holiday location was seen as a strong memory aid, of being able to create the feeling of being there, and most interestingly of having the additional value of giving a feel for relative locations of places of interest to each other. Here are the quotes:

‘Interacting with the 3d application may help people to remember what they saw as opposed to looking through endless brochures.’

‘I think the use of VR DOES give you more information on your vacation. By that I don't mean information like hotel costs and stuff, but it gives you a better 'feeling' of the location.’

‘You can get a feel for your resort (given a detailed enough model) and with the map of Rhodes you get a feeling for where the hotels are located in relation to one another, and the resorts which is difficult to get from a brochure.’

‘The VR environment should give you an idea of the spatial properties of the holiday location: is this town/hotel close to the airport? What do you see from your hotel? Is the hotel located on a hill? What can you do in the neighbourhood? These properties are not at all well addressed in the present version.’

5.6.3 Conclusions

Because the application had not been fully tested in multi-user mode previously, the subjects ran into problems while transporting themselves as a group to the island Rhodes, and were only able to solve this problem by making use of their extensive expert knowledge of programming in general, and of the application in specific. The execution of the task was lengthened considerably because of these problems, so that on the whole it seems advisable to rerun the pilot with the experts at a later date, with a more finished Citizen Application. In general a large amount of information about the usability of the Citizen Application has been found, and additionally interesting expert suggestions have been collected on ways to improve the application.

6. Auxiliary Trials

6.1 Effect of Body Movement on the Sense of Presence

This section reports on an experiment (Slater *et al.* 1997a, Slater *et al.* 1997b) with a three-fold purpose:

1. To assess the influence of body movements on the sense of presence
2. To study the impact of task complexity
3. To study two independent measures of presence, one subjective and one objective

The implications of the results on these three factors have important implications for CVE technology and most importantly for the 2nd phase of the COVEN evaluation. The study shows that presence is positively correlated with head rotations, i.e. the more the participant has to move around the more present they become. And secondly the subjective and behavioural measures were correlated which suggests an improved method for measuring presence.

6.1.1 Presence Model

The model of presence developed at UCL is introduced in the companion document “Framework for Usability Evaluation”. In summary we consider presence as a psychological factor that can be created when a person is *immersed* within a virtual environment. The immersiveness of a system is a measure of technological factors such as:

Inclusion - the degree to which external environmental factors are excluded

Extent - the range of sensory modalities presented

Surrounding - the degree to which the display surrounds the participant

Vividness - includes the range and quality of the displays in each modality

Matching - match between participant’s proprioception and displayed environment

Plot - the degree to which the environment presents a compelling and coherent story-line

When measuring presence we can aim to measure one or both of their *subjective* or *behavioural* presence.

6.1.1.1 Subjective Presence

Subjective presence is the participant’s own evaluation of their sense of presence. This is done through a questionnaire which elicits responses upon three attributes:

- (a) The sense of ‘being there’ in the virtual environment as compared to being in a place in the real world;
- (b) The extent to which there were times when the virtual environment became the reality - to the extent that the subject ‘forgot’ that they were standing in a laboratory wearing a head-mounted display;
- (c) The extent to which their memory of the virtual environment is similar their normal memory of a place.

Each of these is answered on a rating scale (1 to 7). The actual questions used are summarised in Section 8.3.1 since it is suggested that these are included in future questionnaires when appropriate.

6.1.1.2 Behavioural Presence

The assumption here is that if a task is performed in the same way in a virtual environment as it is in the real environment then the person must be present in the virtual environment. Held and Durlach (1992) and Sheridan (1992) both suggest testing for a startle response to, say, a virtual block threatening to hit the participant’s head.

In a previous experiment (Slater, Usoh and Chrysanthou 1995) a related approach was used to measure behavioural presence. Essentially one object (a virtual radio) existed in both the real and virtual environments. Although the participants would see the radio always in the same place in the virtual environment, it would occasionally make a sound in the real world that was from a different direction. The visual and audio cues would thus appear to come from different directions. The hypothesis was that, individual visual and audio dominance aside, the participant would indicate the radio in the environment in which they felt present. That is, if they felt immersed in the virtual environment they would point to the virtual radio (visual cue), if they did not feel present in the virtual environment they would point at the real radio (audio cue). This was shown to be so after the visual/audio dominance was factored out.

6.1.2 Experimental Design

6.1.2.1 Task

The experimental scenario consisted of a number of plants, each with 16 leaves arranged in a random pattern within a walled garden. The basic task for the subjects was to examine the underneath of the leaves for each plant in order to determine whether the tree was healthy or ill. Figure 8 shows the participants view of two unhealthy trees in the high variation condition.



Figure 8: View underneath two unhealthy trees

In order to control the amount of bending and head rotation, the height and shape of the trees could be controlled so that it was easier or harder to look under the leaves. There were two conditions: in the low variation condition the tree leaves were mostly at or slightly above eye level so little motion would be required in order to look underneath the leaves; in the high variation the tree leaves ranged from being almost at ground level to several metres high.

In order to control the task complexity, two different constructions were given: to count the diseased trees and to count and remember the positions of the diseased trees.

This produced 4 cases, each of which was allocated 5 subjects.

6.1.2.2 Virtual Lab

The third purpose of the experiment was to compare subjective and behavioural measures of presence. We wished to develop the method described in Section 6.1.1.2 where the subject's own bias for conflicting cues can be used to determine which environment they are present within.

A similar method would not work again, since the participant would be moving around the environment investigating cues, so there would be no easy way in which to maintain a fixed relationship between the participant and virtual object. The approach taken was novel in that it involved tasks being undertaken in a virtual copy of the real labora-

tory in which they were undertaking the experiment.

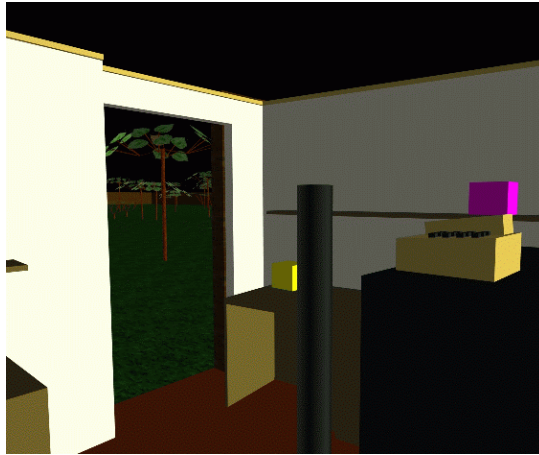


Figure 9: View from the virtual laboratory out into the field of plants

Essentially the virtual environment section of the experiment was arranged as follows:

1. In the real lab the participant is shown a real phone and trained to point at the phone when it rings and call out the colour of any box that they see,
2. Enter virtual lab, familiarisation with controls and main task details. The participant sees a virtual phone in the same position as the one they previously saw in the real lab. Again they are trained to point at the phone and call out the colour of a box whenever it rings.
3. Main task. Participants leaves the virtual lab through a doorway and enters the field to carry out the plant inspection.
4. After a set time, the participant returns to the virtual lab. Again they perform the task of pointing and calling out box colours whenever the phone rings. However this time the audio comes from a different position to the virtual phone.

Note that it is important that the participant is instructed simply to point at the phone “when it rings”, not “when they hear it ring”, nor is the experimenter allowed to clarify when asked whether the real or virtual phone is the one meant.

This procedure is thus similar to the virtual radio procedure, but is novel in the fact that it takes place after and in a different section of the virtual environment to the main task.

6.1.2.3 Sensory Dominance

As mentioned in Section 6.1.1.2 it is necessary to factor out sensory dominance when considering the behavioural presence measure. This consisted of two parts, completion of a sensory dominance questionnaire (Slater, Usoh and Steed 1994) and a behavioural test. The test was a variation of the “cue-target” paradigm, full details are available in the reference paper.

6.1.2.4 Equipment

The virtual environment was built using the dVS/dVISE 3.1.2 from Division Limited. The display system consisted of a Virtual Research VR4 helmet and Division 3D mouse both tracked by Polhemus Fastrak trackers. The graphics was generated by a Silicon Graphics Onyx with twin 196Mhz R10000 processors and Infinite Reality graphics system. The frame-rate for the 32,576 polygon scene was not less than 10 Hz stereo.

Subjects did have a virtual body, but generally did not notice it. The subjects navigated by looking in the direction

they wished to travel and then pressing a button on the 3D mouse,.

6.1.3 Results

We give a summary of the main results in this Section. Complete details of the derivation can be found in the main paper. Note that the “pointing angle” variable is the rotation between the directions of the virtual (visual) phone cue and real (audio) phone cue. The behavioural presence hypothesis would lead to a high sense of presence causing the participant to point towards the virtual cue and thus creating a smaller pointing angle.

- Those that experienced the high variation field generally pointed more towards the virtual phone (i.e. towards the visual cue) rather than the real phone. Given the general assumption behind such objective measures, this would indicate that these subjects were more present. In the low variation field, those subjects that moved their heads more also were found to point towards the virtual phone, supporting this conclusion.
- The subjective presence score is positively associated with virtual distance travelled. It also increases with the combined effect of higher variation field and high task complexity. Those that visited the high variation field subjective presence also increased with higher head yaw rotation. All these suggest that more movement and more difficult tasks are correlated with higher subjective presence score.
- In line with results from our previous experiments it was found that the subjective presence measure is positively associated with the subjective visual dominance score and negatively associated with the auditory dominance scores. Likewise the angle measure is negatively associated with the behavioural visual score.
- The angle score is negatively associated with subjective presence (on a suitable log scale) in a normal regression, taking into account the sensory dominance. The presence score is negatively associated with the angle score on a logistic regression, again taking into account the sensory dominance

6.1.4 Conclusions

The evidence tentatively suggests that:

- Body movement is positively associated with presence
- The angle measure is related to presence and is a useful way of eliciting behavioural presence

In addition the virtual laboratory scenario devised for this experiment has two immediate benefit. Firstly the virtual lab creates a situation where it is natural for the experimenter to talk to the subject since they have not “left” to perform the experiment. Secondly no interruptions are necessary during the main task since all training and final instructions can be given in the virtual lab. The possible benefit of both of these is eloquently expressed in the following quote from one subject:

When I entered the virtual field I really felt I had left the lab - real and virtual. Even thinking about the researchers watching - it was as in they were in the lab I left.

We expect to use this technique for future experiments, especially since the new behavioural presence measure seem to give us acceptable results without interfering with the application task.

7. Application Recommendations

7.1 Recommendations from Inspection

Section 4.2 presented a breakdown of all the inspection issues into system, interaction or application problems. Whilst system or interaction problems are almost all research problems, the application problems represent choices that were made which have alternatives within the scope of the toolkits used. We present a list of suggestions for alternatives and additions that should be more efficient. Whilst some of these suggestions are taken directly from the inspection data, most were generated by subsequent discussions.

7.1.1 General recommendations

Our goal here is to propose elements of a frame for analysing the nature of an application problem, and supporting the problem solving process. A transversal analysis of the application design problems that were uncovered by the inspections leads to articulate the problems and their related recommendations along the general issues of application function design choices, metaphor choice and / or implementation, and general respect of the standard HCI usability factors.

7.1.1.1 Application function design choices

A small number of problems found are ‘deep’, application-level issues questioning the validity of the function design choices in opposition with the user interface nature of the other types of problems. The interest and motivation for a given functional choice may need to be fully re-assessed. Questions that the application designers need to ask themselves are: is this function useful, strictly needed? Does this correspond to a task requirement? Doesn’t this function bear more constraints than benefits to the user and his task? Can’t this function be re-designed in a more efficient way?

In our opinion, example problems of this kind are group formation issues (is the explicit function of ‘group leader’ really useful?), and blueboard communication issues (what is the use of being able to control a personal blueboard independently of the person who is using it? Shouldn’t the blueboard control be the name of its owner?).

7.1.1.2 Metaphor choice; metaphor implementation choices

Due to the 3D-metaphorical nature of the CVE interface, a large number of application-level problems are related to the choice and / or the implementation of metaphors. Of particular interest is the analysis of the exact nature of the problems encountered with the metaphors - this is needed for the application designer to identify the proper solution to the problem: adjust some metaphor implementation details, refine the metaphor design, or totally change the metaphor. We are proposing a tentative frame for analysing the nature and severity of a metaphor problem, and supporting the problem solving process:

A) Is the metaphor implementation fully coherent?

While some disruptions in metaphor implementation are needed² and gracefully performed, others may deeply question the coherency of the representation and of its affordances, threatening the user’s understanding of the scope and

² Realism certainly is naturally not to be looked for; metaphor implementation relies on the choice and exploitation of a

availability of the underlying functions.

Metaphor implementation coherency issues may be decomposed into the following:

- missing representation element or behaviour.
For example, light switches are represented but their behaviour is missing.
- extraneous element or behaviour represented, of no apparent use for the task. Is this element really useful for conveying the general metaphor, or is there a risk for distracting the user?
For example, why represent radiators, why allow the moving of some pieces of furniture?
- unexpected appearance or behaviour.
For example, moving an object through the bandybox surface in order to copy the object to the holoview, mix of realistic and non realistic representations and behaviours.

B) Does the metaphor convey extraneous constraints, which do not exist at application function level? These is a risk that these extraneous constraints impose limitations with no clear benefit to the user and the task.

For example, the size of the teleport machine imposes groups size limitations; the ‘CDs on a table’ metaphor constrains the CD selection process, and imposes limitations as to the number of CDs; the ‘closed rooms’ metaphor, with opaque doors and walls, prevents from being aware of the presence of other people, and of the functions to be found in the environment.

C) Are all the metaphor affordances really useful for the task? Is there a risk for task performance efficiency?

The existence of extra conceptual elements conveyed by the metaphor, of no real use in the application, may be confusing for the user. For example, the desk metaphor in the virtual conferencing room allows (and requires) that a participant chooses his desk and personal blueboard; automatically attaching a blueboard to the user’s view point is an alternative solution outside of this metaphor.

D) More fundamentally, is the chosen metaphor globally and effectively coherent with the function it is intended to represent?

A deep mismatch between the chosen metaphor and its intended function may lead to reject the metaphor choice.

Problems related to the holoview flat screen for displaying 3D objects, the bandybox look of a garbage bin may to some point be assessed as deep mismatch issues.

How to use this frame:

Choosing and implementing a metaphor clearly is a matter of identifying and balancing the affordances and constraints conveyed by a design, so as to reach an effective ratio between the advantages and risks for task performance. Answering questions A to D should allow to characterise the risks attached to the different elements of a metaphor design; these risks need to be balanced against the foreseen benefits of the design, that is the ease of learning, the ‘naturalness’ of the affordances conveyed by the implemented metaphor, etc.

limited number of main features of the source artefact.

7.1.1.3 Respect of standard usability factors

Finally, the usability heuristics used in the Heuristic Evaluation type of inspections are actually basic design recommendations thoroughly used in the HCI community, and fully relevant for the design of CVEs, except for a few adjustments (see Section 4.1.1).

7.1.2 Detailed suggestions for design alternatives

We present a detailed list of suggestions for design alternatives that should be more efficient. Whilst some of these suggestions are taken directly from the inspection data as given in Section 4.2 most were generated by subsequent discussion.

Table 9: Generic Coven Services Suggestions

Problem	Suggestion
Communication - Message Control	Ideally a more visual metaphor for selecting services via a menu. Alternatively a two stage approach to entering a message, where the first stage is specifying the name, which is checked and then the message itself can be entered
Communication - Message Display	Some facility to send longer messages would be useful, combined with better reading facilities and an ability to read previous messages. This might best be combined with the suggestion for a combined Notebook (see suggestion in Citizen Application table below).
Communication - BlueBoard Display	Simplest way would be to blueboard objects copies to be rotated only. Maybe the copies could be drawn slightly transparent to indicate they are not “real” objects.
Communication - Audio	A technological challenge, but experience suggests it is important to know who is speaking which would require some sort of animation of their avatar.
Communication - HoloView Display	The holoview might be better represented as some sort of “container” into which the object is copied rather than a straight copy of the actual object. This makes it more apparent that the object is “projected”. Something simple like a museum display case might be appropriate
Communication - HoloView Control	The holoview position controller is fairly redundant. The control is only useful for an immersed participant, and then it would be useful for ALL objects not just the holoview.
Communication - Holoview Commands	The text command messages are obscure and would be better represented from a menu or by indicative gestures. The concept of bandy box is useful, but a placement of an actual object limits it because it is awkward to manipulate the object to the box. Maybe it could be integrated into a revised text messaging, where the creation of a text message makes a virtual object that can be sent to someone or something.
Object Manipulation - Selection Consistency	Care will need to be taken in order to ensure consistency across interaction styles. In particular no action should be activated by simple selection unless there is a simple and immediate way to undo the command.
Object Manipulation - Picking Consistency	Again consistency should be observed on the behaviour of picking. Every object that has an action highlight in some way to indicate an effect might occur.
Mutual Awareness - Positions	Although it is possible to look up which zone a person is in, this might be better represented on a map (see comment below). This might just reflect which room they are in rather than where in the room.
Mutual Awareness - Avatar	A more expressive body would be useful. Even simple gestures such as an out-stretched arm when moving an object and a pointing finger when selecting an object would be useful.
Mutual Awareness - Actions	See comment above.
Group Navigation - Group Formation	The explicit group formation metaphor currently will not work for large numbers. Perhaps a group could be formed simply by those members in the teleport room. The group could exclude others by locking the door and having an access criteria. The group could simply move to Rhodes whenever they wished. The group leader token would become a single item or page on owners personal notebook rather than a special item. Whether or not the group leader needs to exist is also debatable. More investigation is required.
Global Map-	If no global map is possible, maybe some easy way to get a high up view would be useful.

	Never the less an overview of which zone users are in would be useful (colour coded by group maybe?). See comment below on this point.
--	--

Table 10: Application Common Suggestions

Problem	Suggestion
Orientation/ Relocation Tools	Some form of tool to return to known keypoints would be very useful. Maybe this could be a simple extension of the “Return to VTA button” idea, but with one point for each zone. This would then be incorporated into the zone overview map.
Object Collision	Consistent application of object collision behaviour would be useful.
Room Functions	Again this would seem to suggest that a zone overview map would be useful, with the group or user listing.
Affordance Issues	Given the realistic style of most of the application objects, it would seem necessary to extend this to all the objects.
Door Opening	The door open time need to be much longer. Preferably the doors would open automatically when you are near and looking towards them.
Object Consistency	Consistent application of visual/audio effects when appropriate. Nothing should surprise when it works or doesn’t work.

Table 11: Business Application Suggestions

Problem	Suggestion
Application Context	Maybe a third neutral room could be added so we don’t start in the “wrong” room for the task.
Business Game Rules	More hints need to be given about the rules. This could be posters on the walls or an animated demonstration.
Desk Choice	No suggestion.
Bandy Box Control	As discussed in the coven services suggestions, the physical manifestation of the bandy box is cumbersome. In particular the object collision based approach to setting the object would seem to prevent a clear or reset command being sent to remove the object form the relevant display(s).
Slide Control	Slide objects need to have some visual or audio cue that they are actually slides. Maybe they could look like a set of index cards? This would show how many slides there were and would provide a very easy was of navigating through them (rather than having to rely on the key-board).
Slider Bars	The slider bars seem to need some feedback about their range of motion and how they are moved. Maybe it could look like an abacus?

Table 12: Citizen Application Suggestions

Problem	Suggestion
Remote Control	Combination of the remote control and notebook into a single PDA type object which can then be used for several purposes: transportation, message reading, data/object storage etc. New features could add “pages” to the Notebook
CD Selection	The CD titles are hard to the read at the moment, and the titles are not that descriptive. Maybe a jukebox with a very short description of each one would be useful?
CD Player	CD Player need to have a more explicit control, maybe with play/stop/forward/reverse.
Slide Show	See comment above. The “surprise” factor of the video could be alleviated if the blank video screen contained instructions on how to play a CD.
Teleport Buttons	The teleporter is a fairly abstract device and not that simple to use. A first suggestion would be to make it a lift rather than teleporter, with labelled buttons indicating travel destination. It is complicated by the fact that the Notebook isn’t generate until the teleporter is activated (see previous comments about group formation).
Teleport Machine	See comment above.
Notebook	See previous comments related to group formation and remote control.
Carpet Controls	The carpet should constrain people to remain on it until they leave the group. The boundary should be visible.
Site Icons	Some other method to enter the sites might be more appropriate. Maybe a tour bus could appear to take you there?
Hotel Icons	A key would be useful for the site icons. Another page on the personal Notebook?

The main suggestion from these tables, is that the various head up displays be combined in to a single “PDA” type device that would hold various data and objects on different pages. Although not a simple idea to implement, it seems that it would provide a powerful metaphor. Firstly it would reduce the screen clutter, which can be quite severe at the moment, especially for the group leader when they are in the Rhodes zone. Secondly it makes it very simple to store objects and messages sent to the user for later retrieval. For instance, the Hotel Icon guide could form a single object that is sent to the PDA when that zone is entered.

Several of the items that have been suggested here will require quite extensive development. The group navigation and group formation issues are especially problematic. Groups might better be defined by spatial partition rather than explicit action, since this might lead to a person being left out, an occurrence of which is not undoable at the moment. In the applications as described, there does not seem to be any reason to have more than one group in a zone except for the Rhodes zone, and then the groups might not need to be mutually aware. Separating groups by their being in special zone which they can control access to would seem to be more natural.

Overall the inspectors thought the applications were very successful, though not yet suitable for novices. All the required functionality is provided to some extent, though given the non-existence of guidelines for the design of 3D scenes in order to support application tasks, it is not surprising that many different interaction styles have been applied. To make a more robust application for general usage, decisions will have to be made about interaction and design styles and these will have to be applied rigorously in order to ensure consistency.

7.2 Recommendations from Network Trials

The recommendations from the network all re-iterate the points made during the inspection. Rather than suggesting

alternatives, the results of the network trial indicate which are the most pressing areas for modification in the applications. These following points are taken from the results of the three sets of questionnaires about the trials and applications (Sections 5.4, 5.5 and 5.6). The italics indicate the category in the previous section that the point corresponds to.

- *COVEN Services* Integration of communication media into a coherent and orthogonal interface.
- *COVEN Services* Selection mechanism. selection should not be exclusive, and should not be possible whenever the object is static and has no behaviour.,
- *COVEN Services* Group navigation. The magic carpet metaphor needs revising and updating.
- *COVEN Services* Participant awareness. Better indication of participants interactions and communication.
- *Application Common* Door opening and closing. Since this was the cause of much delay in group navigation between zones.
- *Citizen Application* Keys are required for the hotel visualisation.

8. Methodological Recommendations

One of the ongoing themes of the COVEN usability task was that, given the lack of CVE specific evaluation methods, there was a need to develop new custom techniques. The “Framework for Usability Evaluations” document, summarised in Section 2, derived COVEN’s evaluation strategy from an established methodology, but highlighted several shortcomings. In this Section we summarise our experience with the techniques used and suggest refinements to be used within the next evaluation phase, and indeed as the tentative suggestions for the foundation of a CVE specific evaluation framework. This comprises three areas: inspection experience and recommendations (Section 8.1), network trial experience and recommendations (Section 8.2), and auxiliary trial techniques and recommendations (Section 8.3).

8.1 Inspection Experiences

8.1.1 Using the COVEN Inspection Method

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

In terms of integrating the CW and the HE into one task, the opinions are rather diverse. Two evaluators felt it could be made possible to integrate the two methods by reducing them to a more simplified form. One suggestion is to make one additional column on the CW form which allows the evaluator to categorise each problem along the different heuristics. The main consensus however, is that doing both tasks at the same time creates cognitive overload - there are too many details to keep track of.

A general remark made is that both methods seem to focus mainly on aspects that do not work properly, so that subsequently there is no clear impression of which things do work properly, why, and how. It is suggested that the HE can be used as a way to further analyse and categorise the UI problems which have been identified with the CW, because the HE gives more scope for general impressions

The evaluators were asked how they worked with the forms created for the COVEN Inspection. Two evaluator replied that they took notes during the inspections and edited the report afterwards, one added: ‘[...] I am a lot quicker typing than writing’. Another one replied ‘I wrote the CW answers on the fly, and the HE answers at the end of each major task. I filled in a lot more HE answers at the end as they occurred to me.’ And another one: ‘I was flipping through too many papers while doing the inspection - too many for comfort. I was often confused and spend more time finding the correct page, than I liked.’ A general suggestion is made to create a web based version of the inspection form so that the Inspectors do not have to juggle too many papers, and do not have to enter all information twice.

8.1.1.1 Cost/Effort

In terms of cost and effort incurred by using the Inspection method the evaluators were all surprised by the amount of time it took to complete the task.

‘I estimated about 30 to 45 minutes for inspecting each application. It took about 1h to 1h30 to perform each inspection’.

'I had no idea how long it would take. It took about 2-3 hours'

'Don't remember [if I estimated the time it would take me]. Probably [took me] half a day, though I haven't done all!'

'[I estimated] 30 minutes [...] to do the inspection itself [...]. It took me at least 4 hours to go through the citizen application. I worked on it on two separate occasions.'

Several factors were listed as reasons for the time taken to do the task, roughly divided into system related factors: system crashes one or twice; time wasted around bugs; lack of efficiency of interaction/navigation actions (e.g. door opening); and Inspection related factors: need to redo parts of the scenario several times to check for general system consistency, to check whether a newly encountered problem existed in this other part of the UI, and taking a very long time to actually write the points down.

8.1.1.2 Bugs

The inspectors recognise the fact that certain bugs are invisible to the programmers, because they created the application but did not anticipate the user taking certain actions. Those bugs are very likely to turn up during the Inspection.. Also, bugs may differ in the level at which they occur in the task-tree. Sometimes they make a complete set of actions impossible, these bugs could be called low-level bugs. These bugs need to be weeded out by an initial Inspection by the developers themselves, based on the task scenario. Other bugs may just make a couple of tasks unavailable for inspection, these bugs could be called minor bugs. Low-level bugs should be fixed before the evaluators perform the inspection. A list of known small bugs should be made available to the evaluators.

It is suggested that there could be 'several layers' of inspections, which address incrementally more refined issues moving from a designers', to a programmers', to a users' point of view. Perhaps different inspection forms need to be used during the different stages of the project, by the different people involved. The evaluators see a use for Inspections at any point in the system development cycle, but express the wish to avoid Inspections where an encountered bug distorts the UI.

8.1.1.3 Collaborative Aspects of UI

CVEs differ from 2D interfaces in that they are 3D and made for multiple simultaneous users. Some additions have been made to the standard HCI CW and HE. The evaluators were asked to comment on the usefulness and effectiveness of the Inspection method.

The suggested Inspection method for testing the collaborative aspects of the application are rather cost-intensive in terms of man-hours. The main consensus of the evaluators is that there should be at least three participants present in the application, where one is doing an Inspection who can then set the pace of the walkthrough. In order to save time it is suggested that individual inspectors conduct the inspection first to weed out problems with numbering and bugs, because this is a known source of time-loss. Another suggestion made is that one of the three participants present (apart from the Inspector) could be the programmer of the application. A third suggestion made is that the application could be tested for collaboration by making a video-tape of participant behaviour which is then inspected using pre-defined collaboration heuristics of visible collaborative behaviour.

In terms of the 3D aspects of the CVE interface, the evaluators note that an Inspection is just as useful as for 2D interfaces, but the Inspection as has been used by the evaluators may have to be more adapted to specific 3D problems than it is now. Issues identified so far, which are not properly addressed are: immersive type systems; and broader control dynamics of the world design, looking around, switching attention between real environment vs.

virtual environment, emotional aspects (fun, insecure etc.), presence. These are some of the aspects which are really important and are not addressed yet.

One evaluator remarked:

‘[T]he HE seems very focused on 2D application problems. The C. seems much more transferable to 3D, especially because the application is something one CAN actually really WALK THROUGH. I would like to see some issues such as: What is the first object/action which comes to mind when entering the room, and why? Because we want to guide the user through the 3D application just as much, or perhaps more so than in 2D.’

Problems identified for the 3D aspect of CVE Inspections are:

- Some atomic tasks are harder than others since they involve some degree of dexterity.
- Viewpoint is unconstrained, so that the inspected object may not always be apparent for end users.
- It is not always clear how many actions one can perform on an object, so that there is a large number of branches on the task-tree, which will only be found during the inspection.

In terms of the general usefulness of applying the Inspection method to collaborative applications the evaluators report that very little functionality requires two people to perform it; no task requires two people to act on the same object, there are many communication channels available, and the Inspection might pick up perceived differences in functionality between sites. The main impression from the answers of the evaluators is that the collaborative aspects of CVEs in general and for the COVEN application specifically are not clearly identified in the Inspection method. Problems which have been identified with applying the Inspection method to collaborative systems are:

- Collaboration needs to be tested with 1 inspector and 2 helpers, which is time-consuming, complex to set up, and will require well trained teams.
- Different interfaces will be apparent on each machine (at the desktop level and scene level). E.g. group leaders and group followers have different capabilities.

The suggestion is made that these problems can be solved by making a separate inspection method to run past all collaboration issues, instead of going through all possible branches of collaboration within the application. The task tree could then target specific collaboration moments in the application, which are representative for collaboration activities of the application goals.

8.1.2 Suggested improvements to the Inspection Method

The evaluators found that working with the official task-tree often brought new branches of the task-tree to the surface. Each evaluator had to assign numbers to the individually identified sub-tasks, but the evaluators did not always identify the same sub-tasks, so that integrating the results became a tedious task. The suggestion is made to let one inspector do the first inspection, to let the sub-branches of the task-tree thus found be the official task-tree for the next group of inspectors, this should reduce the number of newly identified sub-branches, which should reduce the time taken to do the Inspection, and the time taken to integrate the results.

It is recommended to make an electronic inspection form, in order to allow the inspectors to simultaneously inspect

and report, which would save time when recording the data. Additionally a request was made for a communication channel for discussing Inspection experiences, after or during the inspection, for geographically distributed inspectors, to mimic team Inspection behaviours when appropriate.

8.1.3 Suggested Improvements to the Inspection Topics

The 3D user problems need to be identified more sufficiently in order to come to clearly defined 3D Quality Factors which can then be addressed in the Inspection. Also collaborative issues should be explored more sufficiently in order to define appropriate Quality Factors.

Since the viewpoint is unconstrained, the current CW question on "visibility" has to be updated, because it needs to address the visibility of objects in the first place.

CSCW literature should be consulted to find information about collaborative issues which crop up from using shared group objects, e.g. different interfaces will be apparent on different users' screens, such as group leaders and group followers who have different capabilities.

8.2 Network Trial Experiences

The general network trials (Section 5.4) ran smoothly since a couple of the partners have experience with similar long running trials. The general approach of using web based forms for task description and feedback means that information gathering is prompt and reliable. The only difficulties with these trials has been the constant modifications that have had to be made in order to support the networking. Whilst this was due to some inexperience with ISDN networking, it has remained a problem since this is a network configuration solely used for the COVEN trials and as such it is only used once a week or fortnight.

For the next set of trials it is suggested that start-up scripts or check-lists be generated, as well as more explicit instructions for the installation and single testing of trial specific objects in order to smooth the initialisation. Since more is known about the behaviour of the applications under load, there is no longer any need to stagger application start-up as has been the norm until now.

The specific application trials (Sections 5.5 and 5.6) also ran smoothly, since they used the same infrastructure as the weekly trials. Given the fact that the users were expert with the system if not the application, there were no problems dealing with or working around limitations in the applications.

8.3 Auxiliary Trial Output

One of the purposes of the auxiliary trial was to make suggestions for refinement of usability evaluation during the next phase. Although it uses a scientific viewpoint, it was felt that the techniques used developed would be beneficial to any single or network user trial.

8.3.1 Measures of Subjective Presence

The COVEN applications involve communication and comprehension of spatial structures, and we have argued in the accompanying document "Framework for Usability Evaluation" that although presence might not be necessary for such application to succeed, it is a highly desirable since it will lead to the participants behaving as they would in the virtual environment.

Investigation of incremental changes to the display systems in order to enhance the sense of presence is thus a reasonable approach to take at both the system and applications level. However it is complementary to the inspection and experiment based usability evaluation since it is certainly possible to construct inefficient and confusing applications that none the less afford a high sense of presence. This is the case as long as the applications present a consistent and coherent world model. If the applications are not consistent or coherent the level of immersion will be less and thus presence is less likely to perform.

UCL has performed a number of investigations of the sense of presence over the last few years. The questionnaire used to elicit subjective presence has been refined a number of times and the current questions as produced for the body movement experiment (Section 6.1) are given below. Only the core questions are shown and these are scattered throughout a larger questionnaire in practise. We suggest that, where appropriate, these questions are integrated into the evaluation studies for subsequent usability inspection since we have shown through repeated experimentation that subjective reporting of presence and objective behaviour are closely correlated.

- 1. Please rate your sense of being in the field amongst the plants, on the following scale from 1 to 7, where 7 represents your normal experience of being in a place.**
I had a sense of "being there" in the field:
 1. Not at all ... 7. Very much.
- 2. To what extent were there times during the experience when the virtual field of plants became the "reality" for you, and you almost forgot about the "real world" of the laboratory in which the whole experience was really taking place?**
There were times during the experience when the virtual field became more real for me compared to the "real world"..
 1. At no time ... 7. Almost all the time.
- 3. When you think back about your experience, do you think of the virtual field more as images that you saw, or more as somewhere that you visited ?**
The virtual field seems to me to be more like...
 1. images that I saw ...7. somewhere that I visited.
- 4. During the time of the experience, which was strongest on the whole, your sense of being in the virtual field, or of being in the real world of the laboratory?**
I had a stronger sense of being in...
 1. the real world of the laboratory ... 7. the virtual reality of the field of plants.
- 5. Consider your memory of being in the virtual field. How similar in terms of the structure of the memory is this to the structure of the memory of other places you have been today? By 'structure of the memory' consider things like the extent to which you have a visual memory of the field, whether that memory is in colour, the extent to which the memory seems vivid or realistic, its size, location in your imagination, the extent to which it is panoramic in your imagination, and other such structural elements.**
I think of the virtual field as a place in a way similar to other places that I've been today....
 1. not at all ...7. very much so.
- 6. During the time of the experience, did you often think to yourself that you were actually just standing in an office wearing a helmet or did the virtual field of plants overwhelm you?**
During the experience I often thought that I was really standing in the lab wearing a helmet....
 1. most of the time I realised I was in the lab ... 7. never because the virtual field overwhelmed me.

8.3.2 Virtual Ante-Room

We also suggest that the technique of using a virtual copy of the real environment the person is situated in as a starting environment is used for both immersive and non-immersive trials. The general technique we dubbed as the "virtual ante-room" (Slater *et al.* 1997b) has several perceived benefits:

- The ante-room provides a gentle introduction to the VE system since their spatial orientation is the same.

- The ante-room is a natural place to give instructions since it is “less removed” from the real environment
- The metaphor of having the participant then leave to another virtual environment is similar to that of experimental rooms in real situations. The transformation may even be beneficial for the sense of presence as shown by the depth of presence study (Slater, Usoh and Steed 1994).
- The ante-room gives a natural place to perform post task tests.

Overall the benefits are quite substantial, though of course it is a fair amount of work to set up in a distributed collaborative setting.

9. Conclusions

We have presented a usability evaluation of the initial COVEN demonstrators in which we have addressed two main concerns:

- Production of recommendations for the future development of the COVEN demonstrators
- Production of a framework for evaluation of CVE technology

Our initial work derived a CVE usability framework from an established framework used for 2D applications. This had to be modified somewhat to suit the particular constraints and facilities afforded by a CVE system.

From this we derived three main threads of work:

1. usability inspections of the initial applications so as to uncover the main design flaws and allow to clean up the design.
2. observational evaluations of participants performing tasks in networked trials, so as to better understand behavioural characteristics and underlying concepts.
3. auxiliary case-controlled experiments focused on the evaluation of the factors to the ‘presence’ concept.

Each of these produced very interesting results which were formulated into recommendations for the applications and suggestions for improved methodologies for the second round of usability inspection with COVEN.

Results concerning the applications were broken down into three areas:

1. System problems including lack of functionality, performance and display quality.
2. Interface problems that concern the actions of navigating, and picking of objects.
3. Application specific problems concerning the actual actions and meaning of objects within the environment.

Each of these areas is being addressed by a different part of the COVEN project.

Results about methodologies were primarily concerned with the application of inspection techniques designed for 2D applications to CVE systems. In particular it was difficult to establish the criteria affecting participant awareness and 3D interaction styles. In addition we presented work derived from an ongoing experimental program that suggests improved techniques for user trials with CVE systems.

10. Annexes

10.1 Usability evaluation methods

Supporting usability engineering methods classically involve three broad categories of methods: empirical methods, inspection methods and analytical evaluation methods.

Empirical methods rely on having a number of users experiment with the system in some form of controlled manner. Possible supporting tools are interviews, questionnaires, and video-recordings of the user behaviour; both inside the virtual environment and with the application. Two particular forms of experiments may be highlighted:

- Case-controlled experiments are carefully designed to support a particular claim or hypothesis. Any experiment has the same basic form. The evaluator chooses a hypothesis to test, which can be determined by measuring some attribute of subject behaviour. A number of experimental conditions are considered which differ only in the values of certain controlled variables. Any changes in the behavioural measures are attributed to the different conditions, leading to confirm or invalidate the tested hypothesis.
- Observational methods rely on the general observation of the user interacting with the system, together with the use of a protocol analysis method such as thinking aloud (users are asked to describe what they think is happening) or post-talk walkthroughs (a transcript of the session is re-played and the user is invited to comment); or ‘simple’ recording of user actions and analysis of patterns and relationships.

Inspection methods are ‘cheaper’ methods that consist in having a number of ‘usability experts’ inspect a user interface against a number of usability criteria, stepping through a task scenario and looking for usability problems, or checking the cognitive dimension in the fulfilment of the task scenario.

Analytical methods rely on mathematical computations performed on a formal specification of the system, tasks and context of use.

Another way to look at these methods is to distinguish objective measurement and subjective measurement methods. A *subjective measurement* is highly influenced by decisions personally made by the evaluators. Inspection methods typically are subjective measurement methods; analytical methods clearly are *objective measurement* methods; empirical methods may use objective measurement procedures, even though the empirical data is by nature of subjective origin.

The CSCW community in attempting to evaluate technology mediated co-operation and communication between different people has turned to *specific observational approaches*. In particular, ethnographic approaches where trained observers study actual work in progress in order to identify key issues for improvements, which can then be input for the design process. Ethnographic observations have become a popular method of evaluating CSCW technologies and other collaborative work settings. In contrast to experimental approaches, ethnographic approaches make no claims as to proven statistical significance of their findings or of objectivity. Instead, they rely on experienced observers offering useful, and often challenging, insights into the ways in which people work with and adapt to technologies. Such observational approaches have already been used in the early evaluation of Collaborative Virtual Environments such as the work performed at University of Nottingham in evaluating the MASSIVE system (Tromp, 1996). Conversational analysis techniques are also used; these support detailed studies of interaction in the workplace, the impact of new computer based technologies on ‘talk at work’, motivating technical design choices, assist-

ing in the analysis of the design process itself, the behaviour in a work-related meeting being conducted in VR.

10.2 Glossary of usability evaluation terms

(based on (Melchior et al., 1995)).

Analytical methods. An analytical method is a method where formal specifications of the system, tasks and context of use serve as an input. The results of analytical evaluation can be seen as the output of a mathematical function which only depends on the formal input specifications. In general analytical methods are objective and access no empirical data. They can be applied very early in the design cycle. The reliability of measures calculated on the basis of these methods is not in question. Analytical methods are often based on simulation: the interaction of a (future) user with the system is simulated.

Assessment criteria are critical values for relevant measures which are the basis for the assessment of an electronic information service or product.

Case-controlled experiments are empirical methods carefully designed to support a particular claim or hypothesis. Any experiment has the same basic form. The evaluator chooses a hypothesis to test, which can be determined by measuring some attribute of subject behaviour. A number of experimental conditions are considered which differ only in the values of certain controlled variables. Any changes in the behavioural measures are attributed to the different conditions, leading to confirm or invalidate the tested hypothesis.

Critical success factors determine the success of the electronic information application for the organisation. Critical success factors can be product oriented (e.g. higher product quality, innovative design), development process oriented (e.g. more efficient and effective development process), standards oriented (e.g. product complies to standards), societal goals (e.g. product can be used by people with special needs).

Distribution of a measure describes the statistical distribution of the measure values (normal, skewed, even, other) which was identified in real use.

Electronic information application is either an electronic information service or an electronic information product.

Empirical methods rely on having a number of users experiment with the system in some form of controlled manner. Possible supporting tools are interviews, questionnaires, and video-recordings of the user behaviour; both inside the virtual environment and with the application.

Inspection methods are ‘cheap’ evaluation methods that consist in having a number of ‘usability experts’ inspect a user interface against a number of usability criteria, stepping through a task scenario and looking for usability problems, or checking the cognitive dimension in the fulfilment of the task scenario.

Measures are operationalised quality factors. Measures can be subjective or objective, direct or indirect, analytical or empirical.

Measurement means a repeatable, objective procedure for generating a measure. The resulting measure(s) are scaled in a known way, and reference values, reliability and validity are known.

A validation method is a repeatable, systematic procedure to produce a given result. The specific aspect of valida-

tion methods as opposed to a general view of method is that user validation always starts with an objective and validation criteria, i.e. a question such as ‘Is design alternative A more efficient to use than B ?’, ‘Does the electronic application fulfil the minimum health and safety requirements?’. A quality general factor such as enjoyability, cognitive workload or efficiency is implied by the validation question, and the resulting measure must be shown to be a valid measure for such a quality factor.

Objective measurement method. A validation method is classified to be objective if it is based on objective measurement procedures. However, the data can be of subjective (empirical) or objective (analytical) origin. A method based on empirical data (experiments, ratings of users) can be objective if there are objective measurement procedures. The results of the application of a subjective measurement method are influenced by decisions made by the evaluator in a high degree.

Observational methods are empirical methods that rely on the general observation of the user interacting with the system, together with the use of a protocol analysis method such as thinking aloud (users are asked to describe what they think is happening) or post-talk walkthroughs (a transcript of the session is re-played and the user is invited to comment); or ‘simple’ recording of user actions and analysis of patterns and relationships.

Quality of use is used synonymously to usability to make clear that quality of use is a concept which consists of multiple dimensions. The dimensions are measurable. An application’s overall quality of use is then determined by specifying the dimensions which are relevant in a certain usage context, adding priorities to the different quality dimensions (this is optional), defining assessment criteria for dimensions.

A **task** is described in terms of the goals a user wants to achieve. More than one user procedure may exist to solve the task.

Usability and other more traditional terms such as user friendliness, usefulness, ease of use, have in common that they are vague and fuzzy terms. They give the impression of just one single dimension. In fact users have different needs and requirements and perform different tasks with an application. An application which is usable by one user may be tedious to use by another user. Usability is a too narrow concept which does not take into account cost / benefit issues. Hence, the term quality of use is used synonymously to usability.

Usability engineering is a well defined process which is performed as part of the application development process. It can be part of the development process of any type of electronic information application. Although each development project is different, the approaches, methods, techniques and activities applied to achieve usability do not vary much.

User-centred design emphasises on early and continuous involvement of users in the design process.

Participatory design and prototyping means making users part of the design team or letting them participating in the prototyping process as “subject matter experts”. It is not reasonable to expect users to come up with design ideas from scratch. However, users are very good at reacting to designs and prototypes. Participatory design is particularly effective when designing customised software in-house.

A **user procedure** is a sequence of commands to be executed to carry out a task or to reach a goal.

A **user validation process** is a set of ordered activities that contribute to a defined objective of a validation project. A user validation process takes place over time and has precise objectives regarding the results to be achieved. A User Validation Process Model is a description of the structure and the elements of a validation process in terms of

stages and steps, dependencies and data.

Quality factors are features by which a product can be assessed such as efficiency of use, task adequacy, cognitive workload, robustness, learning cost, user acceptance. Quality factors are the result of the decomposition of the term “quality of the application”. They are variables which reflect different independent quality aspects of the application. Validation questions must be formulated in terms of quality factors in order to allow meaningful measurement.

Reliability of a measure describes the degree of stability of the measurement procedure. A measure is reliable if the application of the measure yields reproducible results. Factors which could reduce the reliability of a measure could be for example number or attributes of subjects involved in experimental tests or subjective decisions made by the evaluator.

Scale type of a measure can be ordinal, interval, ratio, absolute. It influences the interpretation of the measurement values to a high degree. For example the scale type of measures must be known for a number of statistical procedures to be applied to the values.

Validity of a measure. A measure is valid for a given quality aspect if there exists a correlation of the measurement values and that quality aspect.

References

- R.M. Held and N.I. Durlach (1992) *Telepresence, Presence: Teleoperators and Virtual Environments*, 1(2), MIT Press.
- Kalawsky, R.S. (1993). *The science of virtual reality and virtual environments; a technical, scientific and engineering reference on virtual environments*. Addison-Wesley Publishing Company, England.
- Melchior, E-M, Bösser, T., Meder, S., Koch, A. and Schnitzler, F. (1995). *Handbook for practical usability engineering in IE projects*. Project ELPUB 105 10107, Telematics Application programme, Information Engineering section.
- Nielsen, J. (1993). *Usability Engineering*. Boston, MA: Academic Press.
- Nielsen, J. (1994). Heuristic evaluation. In Nielsen, J. and Mack, R.L. (Eds.), *Usability inspection methods*. John Wiley and Sons, New York, NY.
- Nielsen, J. (on web). How to conduct a heuristic evaluation. On the World Wide Web.
- Norman, D., Draper, S.W. (Eds) (1986). *User Centered System Design: new perspectives on Human-Computer Interaction*. Lawrence Erlbaum Associates, Hillsdale, NJ.
- T.B. Sheridan (1992) *Musing on Telepresence and Virtual Presence, Presence: Teleoperators and Virtual Environments*, 1(1), MIT Press.
- M. Slater, M. Usoh and A. Steed (1994) *Depth of Presence in Immersive Virtual Environments, Presence: Teleoperators and Virtual Environments*, 3(2), MIT Press.
- M. Slater, M. Usoh and Y. Chrysanthou (1995) *The Influence of Dynamic Shadows on Presence in Immersive Virtual Environments*, 2nd Eurographics Workshop on Virtual Environments, M.Goebel (ed), Monte Carlo, January 1995.
- Slater, V. Linakis,, M. Usoh and R. Kooper (1996) *Immersion, Presence and Performance in Virtual Environments: An Experiment with Tri-Dimensional Chess*, ACM Virtual Reality Software and Technology (VRST), M. Green (ed).
- M. Slater, A. Steed, J. McCarthy and F. Maringelli (1997a) *The Influence of Body Movement on Presence in Virtual Environments*, submitted for publication.
- M. Slater, A. Steed, J. McCarthy and F. Maringelli (1997b) *The Virtual Ante-Room: Assessing Presence through Expectation and Surprise*, submitted for publication.
- A. Steed and M. Slater (1995) *3D Interaction with the Desktop Bat*, Computer Graphics Forum, 14(2), 1995.
- Steed (1996) *Defining Interaction within Virtual Environments*, PhD Thesis, Department of Computer Science, Queen Mary and Westfield College, University of London.
- Tromp, J.G. (1996). Working paper for evaluation methods of collaborative VR - development of a toolkit. Internal document.
- Tromp, J.G. (1997). *Methodology of Usability Evaluations in Collaborative Virtual Environments*, presented at the USINACTS video-conference, March 1997, Brussels, Berlin, Madrid, Tampere, document available from author.
- VRML (1997), *The Virtual Reality Modelling Language*, ISO/IEC DIS 14772-1, 4th April 1997, <http://www.vrml.org/Specifications/VRML97/DIS/index.html>
- Wharton, C., Rieman, J., Lewis, C. and Polson, P. (1994). *The cognitive walkthrough: a practitioner's guide*. In Nielsen, J. and Mack, R.L. (Eds.), *Usability inspection methods*. John Wiley and Sons, New York, NY.

