The COVEN project: exploring applicative, technical and usage dimensions of collaborative virtual environments

Véronique Normand[1], Christian Babski[2], Steve Benford[3], Adrian Bullock[3], Stéphane Carion[6], Nicolas Farcet[1], Emmanuel Frécon[4], Nico Kuijpers[5], Nadia Magnenat-Thalmann[6], Soraia Raupp-Musse[2], Tom Rodden[7], Mel Slater[8], Gareth Smith[7], Anthony Steed[8], Daniel Thalmann[2], Jolanda Tromp[3], Martin Usoh[8], Gidi Van Liempd[9], John Harvey[10] and Nicos Kladias[11].

[1]Thomson-CSF LCR, Domaine de Corbeville, 91404 Orsay Cedex, France.
[2]Swiss Federal Institute of Technology, Lausanne, CH-1015.
[3]University of Nottingham, Dept. of Computer Science. NG7 2RD - UK.
[4]Swedish Institute of Computer Science, Box 1263, S-164 28 Kista, Sweden.
[5]TNO-FEL, Box 96864, The Hague 2509 JG - The Netherlands.
[6]University of Geneva, MIRA Lab, 24 rue Général Dufour, CH - 1211.
[7]Lancaster University, Computing Dept. LA1 4YR - UK.
[8]University College London, Dept. of Computer Science, Gower Street, London WC1E 6BT. UK.
[9]KPN Research, Box 15000, 9700 CD Groningen, The Netherlands.
[10]Division Ltd., 19 Apex Court, Woodlands, Bristol, BS12 4JT - UK.
[11]Integrated Information Systems Ltd., 72-74 Salaminos Street, Kallithea, Athens, GR - 176 75.

Contact: vero@thomson-lcr.fr

Abstract: COVEN (COllaborative Virtual ENvironments) is a European project that seeks to develop a comprehensive approach to the issues in the development of Collaborative Virtual Environment (CVE) technology; COVEN brings together twelve academic and industrial partners with a wide range of expertise in CSCW, networked VR, computer graphics, human factors, HCI and telecommunications infrastructures. After two years of work, we are presenting the main features of our approach and results: our driving applications, the main components of our technical investigations, and our experimental activities. With different citizen and professional application scenarios as driving forces, COVEN is exploring the requirements and supporting techniques for collaborative interaction in scalable CVEs. Technical results are being integrated in an enriched networked VR platform based on the dVS and DIVE systems. Taking advantage of a dedicated Europe-wide ISDN and ATM network infrastructure, a large component of the project is a trial and experimentation activity that should allow to build up a comprehensive understanding of the technical network requirements of such systems along with their usability issues and human factors aspects.
More information may be found at http://chinon.thomson-csf.fr/projects/coven/.

1. Introduction

Multi-participant distributed virtual reality systems are emerging as a new class of tools for supporting groups in collaborative work situations. The essence of Collaborative Virtual Environments (CVEs) is the use of natural spatial metaphors, the support for peripheral awareness, and the integration of participants and data within the same and common spatial frame of reference. Combined with high representational capacities, this creates the potential to support a broad range of co-operative applications in the varied fields of design, visualisation, simulation, training and education, as well as entertainment.

In spite of the exploding activity of the CVE research community, still a number of issues remain to be solved before we can fully reap the promises of this technology. In addition to technical and scientific challenges, a better understanding of the concepts underlying the use of this technology needs to be reached. This calls for co-ordinated, long-term research initiatives allowing to explore the different dimensions of these systems.

An important role in stimulating industrial R&D activity in Information Technology and Telecommunications within Europe is played by the European Commission, with its funding under the current fourth Framework Programme of Research and Development. Several VR-related projects were launched in the past three years, funded in part by the European Commission; among these projects, COVEN (COllaborative Virtual ENvironments) is a project of the Advanced Communications Technologies and Services Programme which specifically addresses the development of CVE technology.

COVEN is a four-year project that was launched in October 1995. The overall objective of the project is to comprehensively explore the issues in the design, implementation and usage of multi-participant shared virtual environments, at scientific, methodological and technical levels.
This is performed through a set of concurrent and related threads of work focused on different issues at application, system, network and human factors levels. The project builds upon a core of networked VR technology and experience gathered around our dVS, DIVE and MASSIVE systems. A significant aspect of COVEN is its grounding in the development and trial of several application scenarios, which allow for extensive experimental evaluations to be conducted regarding network performance and usability issues. The continuity of the project over time (which allows a three-fold cycle of development), its grounding in applications and trials, together with the rich diversity of the project consortium constitute the main assets of this project.

The COVEN consortium

One of the unique aspects of COVEN is the fact that it brings together partners out of very different technological and cultural horizons: as depicted in Figure 1, the COVEN consortium consists in twelve academic, industrial, telecommunications and government partners with a wide range of expertise in networked VR, computer graphics and animation, CSCW, human factors, HCI, multimedia and telecommunications infrastructures.

Thomson-CSF Corporate Research Lab. is the co-ordinator of the project, providing expertise in HCI, Natural Language Processing and VR simulations. Financial management of the project is partly handled by Arax Ltd.

Networked VR systems expertise and technology are provided by Division Ltd (the dVS™ commercial system) and Swedish Institute of Computer Science (the DIVE system). Additional expertise is brought by University of Nottingham with their MASSIVE system, together with their experience in network trials and usability experimentations.

University College London is bringing expertise in computer graphics, rendering optimisation techniques and human factors experimentations. University of Geneva and Swiss Federal Institute of Technology are bringing their expertise and technology in the modelling and animation of virtual humans.

University of Lancaster is bringing expertise in HCI and CSCW.

KPN Research is providing application design expertise together with telecommunications experts. TNO-FEL is bringing expertise in distributed simulations and the development of VR simulations.

IIS Ltd. is bringing expertise in multimedia design and modelling; they are also providing their knowledge and contacts in the Greek tourism market area.

Structure of the paper

This paper provides a picture of the COVEN initiative at mid-term, that is after two years of work. This overview of our multi-facet, comprehensive exploration of CVE development issues is structured into five main sections. Section 2 presents the applications that are driving and supporting our different technical investigations and experimental work. Sections 3, 4 and 5 present the main technical components in our work: the COVEN networked VR platform, and specific work on collaborative interaction techniques and scalability enabling techniques. Section 6 provides an overview of our experimental work.

2. Driving applications

Central to the COVEN project are the applications that drive the development of the technology through their requirements, demonstrate our concepts and ideas, and allow for extensive experimental evaluations of usage and network performance issues.

In order to capture different kinds of collaborative situations and to explore different types of requirements, several application scenarios are developed, as presented in section 2.1. The main challenges in the development of these applications involve design choices, usability concerns, software development and performances issues. These are evoked in section 2.2, together with our approach to address them and our effort to elaborate guidelines out of this experience.

2.1 Application scenarios

Both professional and general public application scenarios were defined, reflecting the varied interests of the COVEN consortium. Scenarios for professional usage are addressing virtual conferencing, a business game, interior arrangement and travel rehearsal. A citizen-oriented scenario is addressing vacation planning.

2.1.1 Virtual Conferencing

The Virtual Conferencing scenario was developed as a generic tool allowing to explore the basic requirements of communication within CVEs and to develop a set of supporting generic functions reusable in the other application scenarios.

Besides the audio channel, CVE users need to exchange information such as short text messages, images, documents, software or 3D objects. Communication can be realised by handing over objects from one user to another. However, to distribute information to multiple users, and to be able to address users, even if they are not present in the same virtual room, a communication system consisting of messages, senders and receivers is needed. By incorporating these communication means in a shared virtual environment, we realised a Virtual Conferencing application where users can meet, present their own overhead sheets and start a discussion.

2.1.2 Business Game

A Business Game scenario was developed as an extension to the Virtual Conferencing application so as to further explore the communication needs in negotiation situations.

This scenario supports the co-operative training of managers who are each responsible for the production and sales of a given product within the same company. Each participant can inspect and manipulate a 3D spreadsheet parameterised with a simple economic model which gives him detailed information about his own product. The participants try to achieve a common goal: maximisation of the total profit. A 3D spreadsheet visualises quantitative information about a certain product, viz., the production and stock capacities, sale, and demand in each month. These values are represented by bars, as depicted in Figure 2. Users can only manipulate the production bars representing the production of their own product per month.

Conflict situations can occur in the case of high stock costs and a peak demand in a certain month, requiring negotiations between the players.

2.1.3 Interior Arrangement

An Interior Arrangement scenario was developed to explore the needs of a collaborative 3D design task; rich interaction means were designed, combining 3D manipulation tools and spoken natural language techniques (refer to section 4.3); subjective viewing capabilities (see section 4.2) were also explored as a basic support to collective interaction and individual task requirements.

Consider a newly built open plan office, which will be populated by a number of office workers. Initially, the office floor is empty, but a number of artefacts will be added to it to create a working office environment (desks, chairs, filing cabinets and so on). Each office worker will be provided with a number of artefacts, some of which are solely owned and some of which are shared (e.g. bookshelves may be shared between two colleagues). The users must co-operatively arrange the location of dividers (temporary walls) and shared desks, while also ensuring that they are confident with the final layout. The final arrangement must comply with a number of constraints, such as walls, steps, availability of power and communication ports and sufficient thoroughfares for people to get to their desks and for others to pass through the office with minimal disturbance of colleagues. Figure 7 provides a snapshot of elements of this application.

2.1.4 Travel Rehearsal

A Travel Rehearsal scenario was developed so as to explore the specific requirements of very large virtual environments.
Learning a route in a city is a complicated task and the actual experience of the city can be fraught with anxiety because of the fear of making mistakes and becoming lost. For a journey such as from London Heathrow to the University College of London, there are many choices to be made about the method of travelling. These range from a relatively straightforward alternative such as getting a taxi, to the more complex task of finding the correct bus routes and changing at the correct stops. Persons inexperienced in making such a journey usually use guides of some sort: physical maps, transport route maps, timetables and directions. Assuming that they have some knowledge about how to travel, the London Travel Application allows them to practise and gain a more concrete experience in making decisions and recognising key points.

The major components of the London Travel Application are: a model of the area, including static geometry and appearance, and dynamic properties such as cars and pedestrians; a model of the travelling possibilities in the defined area; and some guiding documents. The model of the city of London can be infinitely large, depending on scale and detail. In general, low level of detail is required, but on locations where key decisions are to be made, more detail is necessary. Modelled areas will at least include the campus of the University College of London and surrounding areas (see Figure 9), and some relevant parts of the airport and the city. The application will demonstrate how very large virtual environments can be presented to the user, using appropriate scalable rendering techniques as described in Section 5.1.

2.1.5 Virtual Travel Agency

Our citizen-oriented scenario is a large Virtual Travel Agency application that addresses two main objectives: explore and demonstrate the capabilities of CVE designs for general public access to rich and varied information (in this case, on touristic destinations); support the social dimension through allowing participants - who typically have not met before - to get to know each other, form groups and move through the virtual environment together.

Derived from a general survey of a panel of Greek travel agencies, the functionality offered by the application was defined so as to meet the main information demands of tourists. The application offers sophisticated information access functions: besides the usual flyers with textual information about holiday destinations, the customer can view a slide-show supported by audio or pay a «virtual visit» to find out what his favourite holiday destination really looks like, and interact with local experts or guides. One holiday destination is currently incorporated: the Greek Island of Rhodes[1]. The application supports the formation and navigation of groups of participants within the environment. A vacation plan may be defined through picking destinations and accommodations. Figure 3 provides a view of the virtual flight to Rhodes.

2.2 Status

Application development in COVEN is performed within a three-cycle software engineering process involving the design and evaluation of three successive versions of application prototypes. At the time of writing, the first development stage was completed, yielding initial versions the most mature of which were submitted to usability evaluations as reported in section 6. We are now in the middle of the second development stage, which should see the identified design flaws amended and the applications further developed and consolidated, integrating the specific work on collaborative interaction techniques (see section 4).

In the third cycle of the project, the developments will more specifically focus on extensions demonstrating and testing the scalability enabling techniques that are described in section 5.

2.3 Toward guidelines for application development

The experience of developing the COVEN applications highlighted a number of deficiencies in the available supporting platforms, and allowed to provide argumented requirements regarding programming facilities, data distribution, platform support for subjective viewing, access rights, management of groups, media support, avatar functions, to name a few. These requirements were offered as input to the platform and interaction techniques work packages described in sections 3 and 4.

The main challenges in the development of these applications relate to design choices and usability concerns, software development and performance issues. Part of our effort aims at generalising the lessons learned from our practical experience into elements of recommendations (COVEN D2.6, 1997).

As we are faced with the open and still unexplored questions of what a ‘good’ CVE interface may be, and of how to design such interfaces, the insight gained from our practical design and evaluation experience (see section 6.3) is proving invaluable. On-going work aims at elaborating a core body of guidelines derived from an analysis of user perception and user mode of learning within a CVE; this is complemented with an extension of basic HCI usability factors and specific considerations on the use of 3D metaphors.

From software design and implementation viewpoint, we identified a set of generic services for collaboration: mutual awareness, communication, embodiment, subjective views, collaborative manipulation of objects, roles and rights of participants and groups of participants, group navigation. By facilitating these services to application developers, the objective is to reduce the time and effort needed for CVE application development.

Finally, we were faced with a number of performance issues related to modelling, which lead us to clarify guidelines for the CVE modelers so as to tune the models to improve the performance of rendering, collision detection and intersection.

---

1 Rhodes was chosen because of the high amount of tourists visiting the island every year, and the accessibility of data via one of the COVEN partners.

3. The COVEN Platform

The COVEN project needs to undertake research into advanced VR techniques while simultaneously placing stable applications before user in trials. These two drivers place potentially conflicting requirements on a supporting platform for the project. Essentially, a COVEN platform needs to provide open access to researchers to allow them to develop new features and techniques while simultaneously providing stable managed code that can be used for extensive trials by users.
Rather than force a single software platform to address this tension between the project's development requirements and research needs, the COVEN platform recognises the need to maintain a separation between research investigation and application development. The COVEN platform achieves this through closely associating two separate classes of distributed VR systems and providing a clear migration path between the two. This allows the COVEN platform to provide support for the overall construction process from initial research through to the eventual application development and use. In addition, each VR system can focus on a particular set of research and development needs.

- *The Development System:* This is provided by Division's dVS™, a leading VR product. This is used as the final delivery platform for the COVEN applications given its status as a VR product.
- *The Research System:* The DIVE distributed virtual reality system from the Swedish Institute of Computer Science provides the experimental research platform for exploratory prototyping and initial development work. This is complemented for specific investigations by the use of MASSIVE from Nottingham University.

The project supports a migration path between the research systems and the development systems. This allows initial developments to be easily integrated into final applications (see Figure 4). Migration between platforms is supported by common data formats and protocols for real-time inter-working.

3.1 The DIVE system

The Distributed Interactive Virtual Environment (DIVE) is an experimental platform for the development of virtual environments, user interfaces and applications based on shared 3D synthetic environments. DIVE is especially tuned to multi-user applications, where several networked participants interact over an internet (Hagsand, 1996).

As the research system of the COVEN Platform, DIVE provides the frame for part of the COVEN exploratory developments reported in this paper, in particular: the subjective views and way-finding techniques of Section 4, and the scalable rendering techniques of Section 5. Extensions to DIVE within the context of COVEN thus involve system support to these exploratory developments, but also more general consolidations of the platform including enhancements to programming facilities, support of standard formats and tools (especially, integration with the WWW), and support of continuous media; the development of the DiveBone structure also aims at facilitating the COVEN network trials.

3.1.1 General architecture - main features

DIVE is based on a peer-to-peer approach with no centralised server, where peers communicate by reliable and non-reliable multicast, based on IP multicast. Conceptually, all peers share a common state that can be seen as a memory shared over a network. Processes interact by making concurrent accesses to the memory. The DIVE architecture is described in (Carlsson & Hagsand, 1993a), (Carlsson & Hagsand, 1993b) and (Hagsand, 1996).

Consistency and concurrency control of common data (objects) is achieved by active replication and reliable multicast protocols, based on ideas originating from SRM (Floyd et al., 1995). That is, objects are replicated at several nodes where the replica is kept consistent by being continuously updated. Update messages are sent by multicast so that all nodes perform the same sequence of updates. The peer-to-peer approach without a centralised server means that as long as any peer is active within a world, the world along with its objects remains “alive”. Since objects are fully replicated (not approximated) at other nodes, they are independent of any process and can exist independently of their creator.

For sessions involving peers located in different local networks, DIVE has long relied on the existence of the MBone (Kumar & Vinay, 1995), the IP Multicast backbone, a structure for interactive multimedia communication over the Internet. Recently, independence from the MBone was gained by developing the DiveBone, an application-level backbone which can connect sub-islands of the MBone and/or single local networks. The DiveBone will ease traffic filtering and analysis using application-level semantics during simulations involving many geographically dispersed participants.

3.1.2 Integration with the World Wide Web

DIVE was fully integrated with the World Wide Web. This integration has an impact at several levels:

- Any file or document needed during a DIVE session can be accessed using the main WWW transfer protocols, namely the Hyper-Text Transfer Protocol (HTTP) and the File Transfer Protocol (FTP).
- DIVE supports many WWW formats internally: three-dimensional formats (VRML 1.0, AC3D (Colebourne, 19XX) and DIVE's own internal format (Avatare et al., 1995)); ambient and directed audio (SND format); two-dimensional bitmaps and movies (GIF, JPEG, PNG, SGI Movie Format); HTML and GZIP compression.
- DIVE can visualise virtually any other document on the WWW using MIME compliant mechanisms.

3.1.3 Programming facilities

In a typical DIVE world, a number of actors, i.e. the representations of human users, leave and enter dynamically. Additionally, any number of application process exist within a world. Such applications typically build their user interfaces by creating and introducing necessary graphical objects. Thereafter, they "listen" to events in the world, so that when an event occurs, the application reacts according to some control logic. Events can be user interaction signals, timers, collisions, etc. DIVE offers three distinct ways of building applications, which were further developed during the COVEN project:

- The dynamic behaviour of objects may be described by interpretative scripts in DIVE/TCL (Frécon & Hagsand, 1995a), a super-set of the Tool Command Language (Ousterhout, 1994), with specific DIVE commands. Scripts can be evaluated on any node where the object is replicated. A script is typically triggered by events in the system.
- Applications can be written in plain C, releasing all the power — and complexity — of a general distributed VR platform. Typically, C applications will also be triggered by system events.
- External applications can communicate with a running DIVE application through the DCI interface (Frécon & Hagsand, 1995b), an open and simple communication protocol, not to be confused with the multicast protocol used by the DIVE applications to communicate. DCI has been used to interface different languages, e.g. OZ, Java and TCL.

3.1.4 Continuous media support

Finally, DIVE was extended to support live audio and video communication between participants. Sounds are spatialised and video streams can be textured mapped on any object of the virtual scene. DIVE offers a wide range of encoding — and compression — methods to adapt to different network requirements.

3.1.5 Perspectives

Started as a lab tool in 1991, DIVE has evolved into a well-developed system running on many platforms[2]. Though, much is still to be done. Further developments will especially continue addressing network scalability issues by moving more and more intelligence into each peer in order to decrease traffic, e.g. support for synchronised animation and generalised dead-reckoning at the scripting level; this work will in particular be performed through integrating some of the scoping techniques presented in Section 5 below. Furthermore, new emerging standards such as VRML97 will be taken into account.

3.2 The dVS system

dVS is a commercial system...
As development system of the COVEN Platform, the dVS system was used to develop the main part of the applications described in Section 2[3].

General description of the platform (short) at start point of COVEN; highlight on shortcomings, specific objectives within COVEN. Summary of the approach and results so far. Perspectives.
1 page - Division.

4. Collaborative interaction

COVEN is addressing the requirements of collaborative interaction in CVEs, with a specific focus on four main issues:

- embodiment, that is providing users with appropriate virtual bodies so as to convey key communication information on presence, location, identity, but also activity, including actionpoints (i.e. object zones where they are manipulating).
- subjective viewing, that is allowing each user to tailor his own representation of the scene so as to match e.g. his preferences or personal task requirements.
- spoken interaction, that is providing interaction modes complementary to 3D direct manipulation all the while respecting VE immersion requirements; and exploring the impact of the collaborative dimension onto this form of interaction.
- collaborative way-finding, that is offering tools supporting VE knowledge exchange and building within a group.

4.1 Embodiment

The COVEN developments regarding embodiment have focused on realistic human representations. While realism is clearly not the only answer to the embodiment issue (see e.g. Benford et al., 1997), this choice was motivated by the requirements of some of our applications (in particular, the Travel Rehearsal simulation-type scenario and the citizen-oriented Virtual Travel Agency). Our work mainly consisted in extending our libraries of dynamic models of humans, and integrating them into the COVEN platform. A particular piece of new work addresses the modelling and animation of autonomous crowds.

4.1.1 Virtual humans

Body Definition:

Based on a general hierarchy manager library, we built a specialised library to model human body hierarchy. This hierarchy is defined by a set of joints, which correspond to the real human main joints.

---

2 DIVE runs on many UNIX platforms (SGI, HP, Sun and Linux) and on Windows NT. Binaries are available for free for non-commercial use at http://www.sics.se/dive/.

3 At the time of writing, only the Travel Rehearsal scenario was developed using DIVE due to its heavy reliance on the research on scalable rendering presented in section 5.

Each joint is composed of a set of degrees of freedom, typically rotation and/or translation, which can evolve between authorised values based on real human mobility capabilities. As depicted in Figure 5, through a set of scaling methods applied to several point of this skeleton, it is possible to obtain a set of different bodies in terms of size (global scaling of body parameters) and in terms of characteristics (local scaling like spin scaling, lateral or frontal scaling).
To generate realistic modelling of human body shapes, we adopted a multi-layered approach (Thalmann et al., 1996). Ellipsoid metaballs are used to simulate behaviour of bone, muscle and fat tissue. Those metaballs are attached to the pre-defined skeleton and organised in order to approximate human's organ position. Body surfaces are then generated in three steps:

- The implicit surface resulting from the combination of the metaballs is automatically sampled
- Sampled points constitute control points of a B-spline patch for each body part (limbs, trunk, pelvis...).
- A polygonal surface representation is constructed by tessellating these B-spline patches.

Dynamic Level-Of-Detail:

The basic idea behind LOD is to adjust model complexity based on target image quality and rendering speed. The implicit surface can be sampled (step 1 above) in a more or less coarse fashion to generate polygonal meshes of increasing complexity. Practically, the sampling of the body implicit surface is done by casting rays in a star-shaped manner for one contour, with ray origins sitting on the skeleton link. By varying the number of sampled contours as well as the number of sampled points for a contour, we can obtain polygonal surface representations ranging from a few hundred polygons to a very fine model made up of about 12.000 polygons.

During animation, the fineness of the body model is chosen automatically, according to its distance from the viewpoint. The distances that trigger a switch in resolutions are scaled with respect to the actual size of the world in which the human model is immersed.

Motion Control:

Once the body is defined in terms of shapes and mobility, a global motion control system is used to animate the virtual humanoids within 3D worlds. A walking motor was defined based on our earlier walking model (Boulic et al., 1990). In order to obtain realistic walking motion, this walking motor was based on anthropometric data. By defining a set of configuration parameters on medical data, it is possible to obtain several types of walking motion just by adjusting one or several parameter's values. To increase capabilities of interaction between bodies and between 3D objects, a set of methods was implemented to control body mobility. Inverse kinematics (to control limbs motion), key-framing, dynamics were included in a motion generator tool (Boulic et al., 1994) (Boulic et al., 1995) which permits to combine all these techniques to obtain a very complete way to control virtual body motion.

4.1.2 Crowd Model

There are two kinds of crowd model which can be simulated within COVEN. The first one represents an autonomous crowd which is formed by different groups and can have different behaviours and reactions depending on the relationship between the groups. In this model, the agents have an autonomous participation within the Virtual Environment providing a sense of crowd presence (Musse & Thalmann, 1997). Figure 6 is showing an example autonomous crowd.

In the second model, the crowd is controlled by an avatar. The agents loose their own individual and group behaviour, and respond to the avatar's rules for defining new goals and behaviours. As a result of this change in the groups parameters, the agents follow the avatar. However, they also respect the basic group behaviours such as flocking, avoidance collision and goal seeking, but in this case, goals are represented by the avatar's positions.

A more realistic sense of crowd is provided when we mix both kinds of crowd models, e.g. by simulating a crowd formed by various autonomous groups and groups controlled by different avatars. This will be further explored in the context of the Travel Rehearsal COVEN application.

4.1.3 Connection to the COVEN Platform

To be able to use the virtual human body definition inside the COVEN platform (dVS and Dive systems) as user representation (avatar) or as autonomous agent, we established a connection between an external process and the COVEN platform based on shared memory technology. This approach allows to use two different kinds of system without deeply combining both of them. The body controller process receives data from the main application which determines the way to animate the avatar: walking motion, grasping motion or keyframe motion can be activated and mixed inside the body controller process which then sends back the resulting animation to the COVEN platform. According to the received information which mainly consists in a set of axis-angle values (some information relative to interaction between the virtual actor and 3D objects can be part of the data flow), the COVEN platform database is updated in order to perform the final display. Because the external process is not directly dependant on the COVEN platform, any 3D shared world application can follow the shared memory specifications in order to be able to establish a link with the external body controller and animate bodies inside its own 3D worlds. Additional details may be found in (COVEN D4.4, 1997). Figure 2 shows the humanoid avatars within the dVS-based Business Game application.

Controlling a set of virtual humanoids (autonomous crowd, guided crowd or user's avatar) will produce a high amount of data to be transmitted from and to all connected clients. This dataflow needs to be optimised in order to avoid an unacceptable slow down of the application. This can be performed through several approaches:

- Applying a compression method on data before sending them on the network.
- Optimising the update rate according to a function which can take a set of parameters based on net load and/or graphics load.

The ultimate solution would be to combine these two approaches; this is the subject of further work.

4.2 Subjective views

Visualisations of virtual worlds by current CVEs are identical for each user, albeit from a different viewpoint. Users cannot tailor their representation of the virtual scene or the degree to which they are aware of other user's activities. This is analogous to the use of strict-WYSIWIS in early 2D interfaces. Research in the area of shared 2D interfaces has shown a strong trend to support individual tailoring of the shared views, and move away from the strict-WYSIWIS abstraction.

The experience of work from early 2D CSCW systems can be integrated into CVEs, allowing individual users to have more control over their user interfaces, and applications to address users on an more individual basis. Support for these features, integrated into a CVE, will support enhanced 3D multi-user applications and let users tailor their user interfaces to suit their working needs. This will provide users with '*subjective views'* of the shared virtual worlds.

The core approach to supporting subjective views is through an access matrix, which extends that developed for a 2D user interface system, SOL (Smith & Rodden, 1995). This extended model, termed Solven (Smith, 1996), defines the representation of individual objects for individual users. The matrix defines an object's view in terms of two independent factors:

- Differing 'geometric' definition (such as different textual or audio definitions).
- Highlighting and de-emphasising abilities (such as making an object appear as wireframe)

These two factors, termed 'appearance' and 'modifier', may be arranged orthogonally into a 'view-matrix' which defines an object's range of possible representations. The appearance of an object is used to describe its 3D geometry. This may be anything from a simple cube to a more complex building. The modifiers are independent effects that may be applied to any of the appearances. An initial set of modifiers include: invisible, transparent, wireframe, dim and bright. For example, consider an object which represents a 'first class honours degree'. This may be represented by either a large '1', a sphere or a mortar board. The modifiers apply orthogonally across each of the representations, allowing one user to view this object as a normal '1', while another sees a transparent sphere.

This implementation of subjectivity has been used in a number of cases within COVEN and other projects (Smith & Mariani, 1997). Within the Interior Arrangement COVEN application, subjective views are used for two main purposes: support for differentiated interaction feedback and user-level customisations. Subjective views are used by the interaction management system to provide different forms of feedback to the users depending on whether they are taking part in the interaction or not - e.g. referents of spoken utterances (see 4.3) will be highlighted for the speaker only.
User view customisation facilities allow them e.g. to selectively view visualisations of draughty or noisy regions. Additionally, users working independently can choose not to see the application's visualisations of other users' activities. For example, in the two screenshots of Figure 7, the user on the left is moving the table using a specific driving tool interface. As the table is in collision (with the filing cabinets) it appears as wireframe to that user, proving that user with feedback of the table's current state. The user on the right has chosen not to view these features, and sees a solid table being moved.

4.3 Speech-based multi-modal interaction

Classically, interaction within VEs is the direct manipulation of 3D representations. While the interest of the direct manipulation paradigm of interaction cannot be questioned, the limits of this paradigm are also quite clear, making it unpractical to refer to abstract concepts, to objects that do not exist yet (typically to create them), to specific properties of objects, and to large and evolving sets of objects. One option is to have recourse to 'traditional' interaction modes such as textual commands, 2D or 3D menus, with the risk to prevent full immersion within the environment, or to threaten the premise of VR interface transparency. In COVEN, we are exploring the alternative solution of spoken natural language interaction (Normand et al., 1997); a prototype system was developed within the context of the Interior Arrangement application scenario, featuring speech-based multimodal interaction capabilities.

Among the different issues underlying natural language multimodal interaction, the central issue of referent computation has been the focus of a large part of our work. The role of referent computation in Natural Language Processing is to match the noun phrases of utterances with the objects of the environment. Referent computation is here impacted by the need to combine spoken utterances and pointer-based designations in a multimodal interaction mode; the specificities of 3D immersion also need to be taken into account, with in particular the open issue of the perception of object salience and contrasts; finally, multi-participant interaction calls for the exploration of multi-participant referential contexts. These two latter points remain largely unexplored in the community at the time of writing, which justifies significant dedicated research effort.

4.3.1 Referring objects within a virtual environment - processing approach

Our approach is an evolution of our past work on multi-modal references for 2D interfaces (Pouteau, 1993). It is framed within the general theoretical background of the structure of human discourse (Grosz, 1977), with the influence of perception however leading to moving away from a denotational view of reference to a view where contrast of the referent to its background is an important element. We are using the notions of referential contexts and axiology developed in (Gaiffe, 1992). We consider that a multimodal interaction builds up a number of different possible *referential contexts*: contexts built up by the discourse (Sidner, 1986), subject to such linguistic phenomena as anaphoric co-references; contexts built up from the activity of the user (task contexts), and from the perception the user gets of his environment, subject to « direct » references.

Our current prototype manages five referential contexts: two anaphoric contexts[4], the topic of the last utterance, and the interaction history; two perception contexts, the so-called salient objects (e.g. new objects in the field of perception, highlighted objects) and the general set of the objects perceptible by a

---

[4] A full multimodal discourse is supported, mixing spoken utterances and direct manipulation actions, and allowing anaphora across interaction modes.

user (visual and audio perception); the remaining context is the general context of the overall environment. Depending on the characteristics of a noun phrase, referent computation takes place within a subset or all of these contexts, searching each context in turn until candidate objects possessing the properties denoted by or inferred from the noun phrase are found. Within each context, referent computation is constrained by the application of contrasting axiologies.

Particular on-going research addresses the handling of plurals based on perception contrasts, typically using the heuristics of spatial cohesion and the sharing of perceptible properties to distinguish groups of objects within a scene.

4.3.2 Spoken interaction in a multi-participant setting

Our approach to multimodal interaction within a collaborative work setting addresses two types of situations: loose co-operation mode, where each participant retains her personal interaction resources, and tight co-operation mode, where interaction with the system is collective to a group of participants (involving in particular collective referential gestures and selections). We defined the concept of ‘interaction agent’ attached to each participant, with agents being shared in collaborative interaction situations. An interaction agent shared by several participants manages a collaborative interaction stream, including collective referential contexts.

The impact of the collective dimension onto the referential contexts is an open question; although the general feeling is that multi-party discourse structure is analogous to single speaker situations, theoretical investigation of the structure of multi-party discourse clearly is insufficient to support any definitive claim as to e.g. the impact of turn-taking on referential coherency, as evoked in (Grosz et al., 1995). In the absence of specific grounding work, theoretical investigations being out of the COVEN scope, our approach is to rely on the hypothesis that a collective referential context will be built up and exploited in a way analogous to simple contexts, and to assess the validity of this assumption through practical experiments. At the time of writing, the main impact of the collective dimension onto our referential contexts consists in the introduction of a group task context allowing to take into account the salience of the objects manipulated by the group in loose co-operation situations. The general issue of collective referential contexts is the subject of further explorations.
Additional details may be found in (COVEN D4.4, 1997).

4.4 Interactive aids for collaborative way finding

As in the real world, the possibility for a virtual city traveller to become 'lost' in a large scale environment is important when considering way finding and understanding the virtual environment. In order for the participant to be able to learn about the environment it must permit him/her to explore it with freedom. This means being able to make mistakes or, in effect, become lost. However, it should also permit the participant the ability to recover from such situations and, to some degree, gain information about the environment. We are interested in techniques which will enhance the information retained from the experience for later application to the real world in a simulation scenario such as the one of the Travel Rehearsal application. In this effort we propose techniques which can increase the behavioural presence (Barfield et al., 1995), (Slater et al., 1995).

We have developed a set of interactive tools to enhance communication and awareness, based around the concept of the 'lost traveller' in VE who encounters another participant (the helper) and proceeds to obtain directions. During such human encounters a number of modes of communication are employed, e.g. verbal, gestures, maps. It is important that any adopted techniques should either be reproducible in the real world or permit information transfer to the real world. The techniques we have adopted are based around the dynamic creation of three dimensional maps for use in way finding. The maps are created using whole body sweeping movements of the participants. They permit knowledge to be communicated between participants as well as allowing the participant to gain a greater overview of the virtual environment by using the World In Miniature metaphor (Stoakley et al., 1995). Our implementation of this metaphor, a Path Planning Map (PPM), has some added features which include the real-time monitoring and representation of all users of the environment, the categorisation of users into subjective groups, and the ability for different subjective groups to collaborate. The PPM can also be used for locating and orienting the user, searching for other users, and for moving directly to a point on the map.

We have integrated into the collaborative tools the idea of body centred interaction (BCI) (Slater & Usoh, 1993) for application to a worlds-within-worlds or depth of presence concept (Slater et al., 1994). In the context of map exploration the BCI paradigm is used in dynamically sweeping out a road in the 3D space and in interactive shrinking and growing of the lost traveller. That is, after interactively creating a map outlining the route to the point of interest, the lost traveller may explore this map 'internally' by shrinking down to the scale of the map and navigating it from within.
In this context we have also developed several methods for the process of interactive map drawing. These range from a 2D based approach to a 3D interactive space. Whilst creating the map collaboration may also occur in several ways:-

- the helper may construct a map and the lost traveller annotates it.
- the lost traveller constructs a map of the destination (e.g. tall building with gas tower nearby) and the helper constructs the path to the location.
- the environment contains *n* number of helpers and *m* number collaborate in the map construction.

The system is gesture based allowing the recognition of:

- a horizontal sweeping action to represent a road which is dynamically created as the helper moves his hand. This road is created with sidewalks and street lighting;
- a "cross" gesture to denote a pedestrian crossing with a set of traffic lights;
- a circular gesture to denote a roundabout;
- a zigzag gesture to denote a zebra crossing.

These artefacts are placed within the 3D map at the position the gesture took place (see Figure 8). Certain gestures may only be active depending on the state of the scene. A pedestrian crossing gesture, for instance, will be recognised if made on the virtual road.
We have undertaken a number of collaborative trials with two participants using a 2D and a 3D interactive space system. The collaboration was made between an SGI High Impact running a desktop VE, and an SGI Onyx running an immersive scenario. The participants were able to collaborate and interactively construct a set of roads and populate them with traffic based artefacts. We are presently investigating the depth of presence method and its relationship to knowledge acquisition in learning a large scale virtual environment.

5. Scalability enabling techniques

The scalability of collaborative virtual environments provides a major focus for COVEN. The project is addressing two related scalability issues: the scaleable rendering of large virtual environments, and spatial scoping techniques in relation to network architectures. The former considers how very-large virtual environment models can be rendered using a combination of enhanced view volume culling and occlusion culling techniques. The latter considers support for large numbers of simultaneous participants through the scoping of each individual's knowledge of the virtual environment and then the mapping of these "scopes of interaction" onto an appropriate underlying network architecture. We now consider each of these two areas in turn.

5.1 Scalable rendering of virtual environments

One of the demonstration scenarios of the project is for Virtual Travel Rehearsal between London's Heathrow airport and the centre of London (see section 2.1.4). In order to facilitate this, a large model of central London, and the underground train network is being constructed - Figure 9 is showing a sub-part of this model. Even though this model is in its early stages, at the time of writing it cannot be rendered in real-time on an Infinite Reality Onyx system. An essential part of the work of the project therefore is the exploitation of techniques for real-time rendering of large and complex environments. This is especially complex, since the model includes not only highly detailed interior scenes (for example, offices) but also, of course, outdoor scenes - hence the traditional culling techniques cannot be straightforwardly employed.

Two complementary techniques are being developed. One is based on view volume culling, and the second on occlusion culling. The view volume culling technique (Slater & Chrysanthou, 1997) is based on a probabilistic caching scheme. Objects are represented by smallest bounding ellipsoids. These ellipsoids may be used to compute the maximal distance of objects from the boundary clipping planes. Objects can therefore be partitioned into three sets: those completely inside the view volume, those on the boundary, and those completely outside. Those on the inside can be rendered without clipping. Those on the boundary must be clipped. Those on the outside can be further partitioned into distinct sets depending on their distance from the view volume. Associated with each set is a sampling probability, inversely proportional to the distance from the view volume. At each frame these sets are sampled - so that only a very small fraction of all 'outside' objects need to be examined at all in each frame. Preliminary results for this method are encouraging, not only in execution speed comparison to bounding box approaches, but also in terms of the small error that is introduced.

Level of detail techniques are also being developed in order to support all the COVEN scenarios. Two distinct themes are being explored: dynamic level of detail for general complex models and specific work on dynamic level of detail for animated human avatars. The former is a preliminary study on dynamically deconstructing and reconstructing meshes from vertex sets stored in an efficient search structure. <u>The latter was evoked in section 4.1.1 when presenting our techniques for modelling and animating virtual humans.</u>

5.2 Scoping techniques; scalable network architectures

The COVEN research builds upon a number of techniques which have recently been proposed to support the scalability of CVEs, including various forms of spatial scoping. Several researchers have considered how to introduce some form of spatial partitioning into shared virtual worlds (typically involving notions of zones or regions) and then map this onto a underlying network architecture in an efficient way. For example, NPSNET (Macedonia et al., 1995) tiles the world with hexagonal cells, each with its own multicast group, so that observers need consider only near-by cells. The Spline system (Barrus & Anderson, 1996), composes the world from arbitrarily shaped regions or "locales" which localise interaction and which may be "stitched" together by arbitrary 3D transformations. A similar "zoning" approach is adopted by Broll in his multi-user extensions to VRML (Broll, 1997). In a slightly different vein, RING (Funkhouser, 1996) scopes interaction and communication according to potential visibility in densely occluded environments (e.g. within buildings) while localising interactions at servers responsible for specific regions of the world. COVEN research aims to extend this work through: a greater focus on the audio medium as a major bottleneck to scalability; support for richer and more varied forms of social interaction; and defining scopes that are more dynamic in terms of their extent, shape, location and inter-relationships.

Our approach has been to extend the so-called spatial model of interaction as defined in the previous COMIC European basic research project (Benford et al., 1995), <u>as well as to investigate complementary lower-level spatial structuring approaches.</u> The following sections provide a brief overview of each in turn.

5.2.1 <u>Extending the spatial model of interaction</u>

The original spatial model provided a set of generic mechanisms by which the users of a CVE could negotiate mutual interaction in a range of different media. These included: *aura*, a sub-space limiting personal presence; *focus*, a spatial expression of a user's interests in others; and *nimbus*, an expression of a user's desire to project their information at others. The original spatial model was influential in shaping several CVE systems including DIVE, MASSIVE and to some extent Sony's Community Place (Lea, Honda, & Matsuda, 1997). It was also subsequently re-formulated for non-VR co-operative systems (Rodden, 1996). COVEN has contributed to the extension of the spatial model in two ways. First, the introduction of "third party objects" allows for the representation of the effects of context, such boundaries and other spatial structures, on mutual awareness and provides a range of scoping techniques. Second, the mapping of the extended model onto a multicast based network architecture assists with network scalability. Both of these developments are described in detail in Greenhalgh & Benford (1997).

Scoping with third party objects

A third party object in the spatial model is an independent object which can affect the mutual awareness between other objects. In general, third party objects can have two kinds of effects: they can directly adapt existing awareness relationships (e.g., attenuating awareness as might a spatial boundary); and can provide a secondary source view of other objects (e.g., providing an aggregate representation of a group of objects). These effects can be applied independently in each available medium of interaction and are most powerful when combined. For example, a third party object supporting a crowd of participants might replace awareness of individual crowd members with a single crowd representation at a low level of detail, but might allow its members to be perceived directly as individuals at a higher level of detail (e.g., on entering the crowd). Third party objects are themselves “first class” objects in a virtual world and can therefore be embodied, interacted with, mobile, dynamic in extent, varying in shape and can apply their effects recursively to one another. Consequently, they can be used to create a variety of different scoping effects including: nested structures of irregularly shaped and sized bounded spaces such as rooms within buildings; tiled zones within open terrain; representations of mobile crowds of participants; data districts which provide aggregate representation of clusters within an information space; and even vehicles which can move a group of participants through a virtual space within their own private region of communication.

A multicast network architecture

The structure of a virtual environment, as defined by a dynamically shifting hierarchy of third party objects, is mapped on to an underlying multicast network architecture. Each third party object can be associated with several multicast groups. One set of multicast groups is used to transmit state updates (e.g., movements) and continuous media information (e.g., audio and video) from artefacts associated with the third party to observers. An application developer can choose whether to split media across different multicast groups or not depending on local constraints (e.g., balancing the need for individual treatment of media versus limitations on the number of possible multicast groups and the management overheads associated with them). A further multicast group provides a back-channel so that an observer (e.g. a newly arrived object) can send a state-snapshot request to all of the transmitting artefacts. The membership of these different multicast groups is determined as follows. An artefact (i.e., a source of information within the world) transmits its state updates to the multicast groups associated with its smallest enclosing third party object. Observing objects can then “page in” potentially many multicast groups in order to receive state updates by overlapping their personal auras (areas of interest) with the relevant third parties. The extent of these auras can be defined on an individual and moment-to-moment basis.

Each third party object also acts as an artefact in relation to its smallest enclosing third party. Consequently, a hierarchy of multicast groups is established corresponding to the nested structure of the virtual world as defined by third party objects. In particular, a third party can compute an aggregate representation of its own multicast traffic and can transmit this to its superior in the multicast hierarchy. Finally, given that third party objects can be mobile and dynamically resizeable within the virtual world, the hierarchy of multicast groups dynamically evolves to reflect changes in the world structure. For example, a crowd of people or a group vehicle entering a bounded room or zone will automatically cause the multicast groups associated with the former to become subordinate to those associated with the later within the multicast hierarchy.

5.2.2 A space-scale structure for perception-based filtering

A different structuring approach is the one of lower-level techniques that do not require particular modelling pre-processing on the world database. COVEN investigates techniques relying on hierarchical spatial grids to filter information depending on observers' perception in large outdoor distributed virtual environments. A space-scale structure is used to dynamically reference sources of information (typically visual and sound sources) depending on their position and scale. This structure is traversed from observers' positions in the world in order to retrieve sources that closest match their perception needs, as defined by a specific rejection function. The current source rejection function is its projected area along an axis toward the observer on a perpendicular plane. This structure is fully described in Farcet & Torguet (1998).

A mixed client-server and multicast network architecture exploiting the space-scale structure filtering mechanisms is currently being investigated. The network infrastructure will be composed of servers communicating through a fast multicast backbone, combined with clients communicating with servers through low bandwidth links. The role of servers is to maintain their copy of the space-scale structure and establish connectivity with other servers - in terms of multicast groups joining and leaving -, based on their respective clients' interests (i.e. perception needs of client-managed observers).

5.2.3 Future work

In summary, COVEN is developing a flexible approach to scoping interaction within collaborative virtual environments. This approach aims to combine scalability, achieved through the mapping of different scopes of interaction onto an underlying multicast network architecture, with a richness and flexibility of social interaction, achieved through a general purpose notion of boundaries and groupings and their effects on mutual awareness.

The COVEN spatial model approach has been implemented in the MASSIVE-2 system and, as reported in Greenhalgh et. al. (1997), several demonstrations of third party objects have already been created including crowds of participants, data districts within a WWW visualisation and spatial boundaries with varying effects on visibility and audibility. Current work is considering how this approach can be realised within the more general COVEN platform, and how and to which extent it can be connected to the space-scale structure approach which is now being further refined.

Finally, our objective is also to fully assess how this work can be related to other COVEN research such as the work on crowd simulation (cf. Section 4) and on scaleable rendering as described above.

6. Experimental work

The COVEN networking trials are based on the enhanced version of the commercially available dVS platform, and on the DIVE research platform, and involve multiple sites distributed over the UK, The Netherlands, France, Sweden, Greece and Switzerland. The trials are being carried out in three distinct phases, based on different versions of the demonstrator applications, making an extensive use of ISDN in all phases and introducing higher capacity networking such as ATM in phases 2 and 3. There are two broad objectives to the experiments, network assessment (see Section 6.1) and usage evaluation (see Section 6.2) of the applications. As the first phase draws to a close we report the experiences and results we have gained, and overview what we hope to achieve in subsequent trial phases.

6.1 General network infrastructure

Our experimental phase at the time of writing this, trialling the applications developed under dVS, is focusing on a subset of five COVEN sites over the UK and The Netherlands. These are UCL, Division Ltd., Nottingham, TNO-FEL and KPN Research, shown by dark markers in Figure 10. The network infrastructure is ISDN based, with UCL acting as a hub in a star topology. Each site connects to UCL with a bonded pair of ISDN channels giving 128Kb of available bandwidth. Audio is provided using rat, a Robust Audio Tool developed under the MICE project at UCL, which uses at most one ISDN channel's worth of bandwidth. A text communication channel is provided within the applications themselves, with as a complementary tool a modified version of ytalk, a multi-party version of the standard UNIX talk command.

There have been over 20 trials so far in this first phase (i.e., 20 distinct virtual meetings between the five sites), and by its conclusion we expect this number to approach 30. Initially, there were a number of trials using the standard dVISE tutorial environment when the network infrastructure was being put in place. After that, more than half the trials have used the Virtual Conferencing and Business Game applications. The citizen application has been used less frequently so far, with 5 trials taking place - Figure 8 provides a screenshot of one of these trials. We have also given live networked demonstrations at the "21st Century: the Communication Age" event in the Brussels European Parliament on the 18th of June 1997.

6.2 Assessing and modelling network performance

The network assessment aims to measure the characteristics of the network traffic generated by the applications and to combine those with network emulation techniques in order to i) understand how existing network technologies such as ISDN and ATM might best be deployed in order to support these kinds of applications and ii) understand the general requirements of future networking protocols to support CVEs.

The trials plan consists of a configuration phase followed by three trial phases. The configuration phase established initial connectivity with the baseline COVEN platform, essentially the standard dVS/dVISE environment. This has been followed by the first phase where the initial COVEN applications have been tested in a networked setting. This has involved the five partners listed above. Phase two of the trials will see us testing the more developed on-line applications, improved using the feedback gained from the first trial phase, and with additional functionality. This will take place in year three of the project. The final trial phase will see us evaluate the scalability extensions which will have been incorporated into the COVEN platform. This will take place in the final year of the project. These two later phases will see both dVS and DIVE in use to demonstrate different aspects of the COVEN platform. It is envisaged that most of the partners will take part in these latter trial phases.

Our approach towards understanding the network traffic characteristics of CVEs involves a three step methodology which was initially used in evaluating the MASSIVE-1 system at Nottingham (Greenhalgh, 1997) and which is now being refined within COVEN. The three steps are:

1. Developing a 'user profile' of typical behaviours during an application session which provides summary information concerning how often people move, talk, interact with objects and so forth and which tells us to what extent these actions are correlated between users (e.g. do they tend to move as a group or randomly as individuals?). Such a profile can be constructed through statistical analysis of system logs gathered over a series of experimental trials.
2. Combining this user profile with information about specific message sizes for a given application and/or platform in order to predict traffic characteristics such as expected mean and peak traffic distributions across different network links. The result of this step is a mathematical model of expected network traffic.
3. Validating this model from step 2 against actual measured network traffic from the experiments as delivered by tools such as TCP Dump and others.

Additional details as well as preliminary results may be found in (COVEN D3.2, 1997). At present we are gathering the raw data (i.e., event logs from trials) which is required for step 1. This will be combined and contrasted with existing data from the previous MASSIVE-1 trials and will lead to our first model of network traffic characteristics for the COVEN platform (steps 2 and 3 above) by the end of 1997.

6.3 Assessing usability aspects

The first of COVEN's three phases of usability evaluation, concentrated on the initial demonstrator applications as described in Section 2, namely the Virtual Conferencing, Business Game and Virtual Travel Agency scenarios. There were three main objectives of this work: to provide support to refine the design of the applications, to further develop an understanding of CVE technology concepts and to explore and refine the actual usability techniques available for CVE evaluation.

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 to take into account the concomitant 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. From this framework three main threads of work were derived:

- 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, especially auxiliary case-controlled experiments focused on the evaluation of the factors to the concept of *presence*.
- Identification of a CVE specific evaluation methodology and development of CVE specific evaluation techniques.

The results of these separate threads identified a significant number of issues at system, interaction and application levels, part of which are general findings applicable to all CVE systems and applications, not just those developed by COVEN.

6.3.1 Framework for Evaluation

In the absence of an existing dedicated CVE evaluation methodology (cf. Durlach & Mavor, 1995), our first task was the design of a framework for the COVEN usability evaluation (COVEN D3.3, 1997). Given the prototypical nature of the COVEN applications, and our desire to understand the essential components of CVE technology, we had to strike a balance between the concerns of *usability engineering* and *scientific enquiry* frameworks. In the context of CVEs these have not been developed, though certainly both viewpoints are represented in recent work on single-user VE technology (for example see Rosenblum et al., 1996; and Welsh et al., 1996).

A number of aspects of CVE technology need to be taken into account because they constrain our options when choosing evaluation method (Tromp, 1997). Since the participants in the CVE have an embodiment in the environment, we are concerned with both their actions in the application space and outside as they interact with the user interface. In particular we are not only interested in traditional HCI aspects such as efficiency and representation of user interface, but also in the participant's perception of space, sense of presence, awareness of other participants, and collaboration requirements.

6.3.2 Usability Inspection

We identified the *heuristic evaluation* and *cognitive walkthrough* methods (Nielsen, 1994) as likely to provide valuable feedback on the applications, though some potential problems were identified in the specific questions and heuristics used, requiring adjustments and caution. In particular, we had to allow for the fact that CVE participants have a lot of freedom of navigation and manipulation and that interaction dialogues are not easy to constrain. As another example, Nielsen's heuristic of a 'minimalist design' did not appear as fully relevant in the context of VEs since the extra detail might help to enhance the participant's perception of space and their sense of presence.

The inspection was carried out by four independent inspectors using task breakdown of three scenarios for each application. The issues were collated and separated into three classes:

1. System problems including lack of functionality, performance and display quality.
2. Interaction problems that concern the actions of navigating, and picking and selection of objects.

3. Application-level problems concerning the actual actions and meaning of objects within the environment, including metaphor issues.

A typical system problem was that of the simulation slowing or stopping when new scene components are loaded when changing zone. Such a problem pervades all applications built upon the system. Interaction problem reflected the difficulties encountered with the particular desktop controls employed in the default platform configuration. As an example it was found to be impossible to carry an object through the environment since navigation and picking modes could not be activated simultaneously. The application issues were broad in nature, from problems with objects whose operation was not obvious such as the teleporter tool enabling virtual flight to Rhodes in the Virtual Travel Agency application, to wider topics such as how best to represent group functions to the group members.

6.3.3 Network Trials

The usability aspect of the network trials focused on exploring technology concepts rather than fine grained evaluation of each application. We were interested in how people used the various communication media, and to what extent they found that they were able to collaborate successfully. Data was gathered based upon questionnaires posed after the completion of each of the network trials held up until the end of September 1997 (see section 6.1 for details of the network set-up and tools used).

A number of issues became apparent very quickly. Notably there was a problem with passing through doors between zones because of a number of factors which led to the formation of a “door policy” which had to be rigidly followed in order for everyone to pass through successfully. There were also a number of issues raised regarding the integration of communication media and lack of feedback about who was speaking or interacting with objects. Participants used all the communication media but found it awkward to have to switch attention between the different windows. They were also disturbed by the fact that they might hear someone talking or see them interacting with an object, while that person’s avatar (standard dVS avatar) would not animate to reflect this action. These issues are being directly addressed in the next versions of the demonstrators and platforms, in particular with the integration of the advanced virtual human techniques presented in section 4.

6.3.4 Auxiliary Trials

Within the framework we proposed, the role of the auxiliary trials was to investigate very specific aspects of VE technology and to propose refined methodologies for the 2nd and 3rd stages of usability evaluations. One major experiment was undertaken in this vein: a study of the effect of body movement on presence (Slater et al., 1997a; Slater et al., 1997b). This was a continuation of an ongoing thread of research on the factors affecting presence and methods for measuring subjective and objective presence (Slater and Wilbur, 1997).

The new study showed that body movement and presence were associated, so that the greater the appropriate whole body movement, such as bending and stretching, carried out by a participant, the greater the degree of reported and behavioural presence. Additionally, it was found that the behavioural measure of presence used for this experiment was correlated with subjective presence elicited by questionnaire. The main contribution of this work to the COVEN evaluation methodology is the ‘virtual ante-room’ technique. This involves the participant entering a virtual copy of the laboratory in which the experiment took place, thus providing a gentle transition from the real world of the laboratory, through to the virtual version of the laboratory, and then out through a virtual door to the area where the main task was to be carried out. An additional benefit of this procedure is that perceptual discrepancies between the real laboratory and the virtual laboratory can be used as the foundation for a behavioural measure of presence.

6.3.5 Future Work

Overall the inspection and network trials were successful in that they generated issues that cover all the development going on within COVEN. In particular system and integration issues are being addressed by work on the COVEN platform (see Section 3), interface issues are being addressed in the work on collaborative interaction techniques (see Section 4) and application issues will be addressed in the next versions of the demonstrators (see Section 2).

In addition the next iteration of the evaluation will build upon our experiences with adaptation of existing HCI methods for CVEs as learned from this iteration and also upon our results from the auxiliary trial about methods for assessing presence.

7. Summary; perspectives

This paper has provided a review of the main activities in the COVEN project. COVEN is an ambitious initiative gathering an inter-disciplinary team of about forty persons at twelve academic and industrial sites across Europe, in an effort toward a comprehensive exploration of the issues in the development of Collaborative Virtual Environments. It is our belief that initiatives of this type are fundamental to enable the take off of multi-user networked virtual reality technology, and to eventually reap the promises of this technology for supporting cooperative work and new forms of social life.

We have described the status and results achieved after two years of work. Several application scenarios are being developed, both for professional and general public usage. These are driving the development of the COVEN Platform, based on extensions to the dVS and DIVE systems. This work on supporting techniques is complemented by two streams of research at interaction and system levels. Four types of collaborative interaction techniques are adressed: realistic embodiment, subjective views, speech-based multimodal interaction, and collaborative way finding. A particular research on scalability enabling techniques is investigating the scalable rendering of virtual environments as well as scoping techniques for scalable network architectures. Finally, our experimental work is a core component of the project, involving the assessment and modelling of network performance along with usability evaluations.

Our results at mid-term of the project may be summarised as: a sum of collective experience gathered in designing, implementing and evaluating the applications; preliminary scientific and technical results on collaborative interaction and scalability support, as well as network performance assessment; initial findings and methodological results on usability assessment; consolidated platform technological components.

Future work in the second part of the project will further develop, consolidate and integrate these achievements. Indeed a significant part of our effort should be dedicated to transversal analyses so as to further develop the connections between the different streams of work, and fully reap the benefits of our multi-threaded approach. Finally, a general objective will be to consolidate, generalise and disseminate the lessons learned from our practical experience.

8. Acknowledgements

We would like to acknowledge a number of persons who have contributed or are actively contributing to the COVEN project in addition to the authors. Vartkes Goetcherian from Arax Ltd; Andrew Fusniak, Steve Ghee and Chris Davies from Division Ltd; Judith Dijkhuis from KPN Research; Rob Aspin and Andy Colebourne from Lancaster; Anneli Avatare, Lennart Fahlèn, Pär Hansson, Marten Stenius, Olov Stahl and Anders Wallberg from SICS, as well as Olof Hagsand, who initially wrote parts of the DIVE description; Jean-Benoît Bonne, Alexis Dollé and David Roussel from Thomson-CSF LCR; Kurt Donkers, Adri Mourits and Hans Jense from TNO-FEL.

9. References

Avatare, A., Frécon, E., Hagsand, O., Jää-Aro, K., Simsarian, K., Ståhl, O. (1995). "The DIVE files description for DIVE 3.2", Reference Document, http://www.sics.se/dive/manual/dive_file_format/.

Barfield W, Sheridan T., Zeltzer D., Slater M. (1995). Presence and Performance Within Virtual Environments, in W. Barfield and T. Furness (eds) *Virtual Environments and Advanced Interface Design*, Oxford University Press.

Barrus, J. and Anderson, D. (1996). Locales and Beacons. *Proc. 1996 IEEE Virtual Reality Annual International Symposium (VRAIS'96),* San Jose, 1996.

Benford, S. D., Bowers, J., Fahlen, L. E., Greenhalgh, C. M., Mariani, J. and Rodden, T. R., (1995). Networked Virtual Reality and Co-operative Work. *Presence: Teleoperators and Virtual Environments*, 4, (4), pp 364-386, MIT Press, ISSN1054-7460, Fall 1995.

Benford, S. D., Bowers, J. M., Fahlen, L. E., Greenhalgh, C. M., and Snowdon, D. N. (1997). Embodiments, Avatars, Clones and Agents for Multi-user, Multi-sensory Virtual Worlds. *Multimedia Systems*, Springer-Verlag, 5 (2), pp. 93-104.

Boulic, R., Magnenat-Thalmann, N. M., and Thalmann, D. (1990). A Global Human Walking Model With Real Time Kinematics Personification. *The Visual Computer,* Vol. 6, December 1990, pp. 334-358.

Boulic, R., Huang, Z., Magnenat-Thalmann, N., and Thalmann, D. (1994). Goal Oriented Design and Correction of Articulated Figure Motion with the TRACK system. *Computers and Graphics,* Pergamon Press, Vol. 18, No 4, pp. 443-452.

Boulic, R., Capin, T., Huang, Z., Moccozet, L., Molet, T., Kalra, P., Lintermann, B., Magnenat-Thalmann, N., Pandzic, I., Saar, K., Schmitt, A., Shen, J., and Thalmann, D. (1995). The HUMANOID Environment for Interactive Animation of Multiple Deformable Human Characters, *Proc. Eurographics '95*, Maastricht, August 1995, pp. 337-348.

Broll, W., (1997). Distributed Virtual Reality for Everyone - a framework for networked VR on the Internet, in: *Proc. Of the IEEE Virtual Reality Annual International Symposium 1997 (VRAIS'97),* March 1-5, Albuquerque, NM.

Carlsson, C. and Hagsand, O. (1993a). "DIVE - A Multi User Virtual Reality System", IEEE VRAIS, Sept.

Carlsson, C. and Hagsand, O. (1993b). "DIVE - A Platform for Multi-User Virtual Environments", Computers and Graphics 17(6).

Colebourne, A. (199XX) *AC3D - 3D Graphics Modeller - HOME PAGE*. http: /www.comp.lancs.ac.uk/computing/users/andy/ac3d.html.

COVEN D2.6. (1997). *Guidelines for building CVE applications*. Public Report number D2.6. Van Liempd, G. Ed. August 1997.

COVEN D3.2. (1997). *Network requirements of baseline platform*. Public Report number D3.2. Benford, S. Ed. January 1997.

COVEN D3.3. (1997). *Usage evaluation of the initial applications*. Public Report number D3.3. Steed, A. Ed. October 1997.

COVEN D4.4. (1997). *Interaction techniques - integration into applications*. Public Report number D4.4. Normand, V. Ed. To be delivered in December 1997.

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

Farcet, N. and Torguet, P. (1998). Space-scale structure for information rejection in large-scale distributed virtual environments. *Submitted to IEEE Virtual Reality Annual International Symposium, 1998*.

Floyd, S. et al. (1995). Reliable Multicast Framework for Light-Weight Sessions and Application Level Framing. *Proc. SIGCOMM 95*, Sept. 1995, ACM Press, New York, pp. 242-256.

Frécon, E. and Hagsand, 0. (1995a). "The DIVE/TCL Behaviour Interface", Reference Document, http://www.sics.se/dive/manual/tcl-behaviour.tcl.

Frécon, E. and Hagsand, O. (1995b). "The DIVE Client Interface", Reference Document, http://www.sics.se/dive/manual/dci.tcl.

Funkhouser, Thomas A. (1996). Network Topologies for Scalable Multi-User Virtual Environments. In *Proc. 1996 IEEE Virtual Reality Annual International Symposium (VRAIS'96),* San Jose, CA, April, 1996.

Gaiffe, B. (1992). *Référence et dialogue homme-machine: vers un modèle adapté au multi-modal*. Ph.D. Thesis in Computer Science. Université Henri Poincaré, Nancy.

Greenhalgh, C., (1997). Analysing movement and world transitions in virtual reality tele-conferencing. In *Proc. ECSCW'97*, pp 313-328, Lancaster, UK, September 1997, Kluwer.

Greenhalgh, C. M. and Benford S. D., (1997). Supporting Rich And Dynamic Communication In Large Scale Collaborative Virtual Environments, accepted for publication in Presence (in press).

Grosz, B. (1977). *The representation and use of focus in dialogue understanding*. Technical Report 151, Artificial Intelligence Center, SRI International.

Grosz, B., Joshi, A. K., and Weinstein, S. (1995). *Centering: a framework for modelling the local coherence of discourse*. Report 95-01, The Institute for Research in Cognitive Science, University of Pennsylvania. January.

Hagsand, O. (1996). "Interactive Multi-User VEs in the DIVE System", *IEEE Multimedia Magazine*, Vol. 3, Number 1.

Kumar, Vinay. (1995). *MBone: Interactive Multimedia On The Internet*. Macmillan Publishing, November 1995, ISBN: 1-56205-397-3.

Lea, R., Honda, Y. and Matsuda, K. (1997). Virtual Society: Collaboration in 3D Spaces on the Internet, Computer Supported Cooperative Work: *The Journal of Collaborative Computing*, 6, pp. 227-250, Kluwer Academic Publishers, The Netherlands.

Macedonia, Michael R., Zyda, Michael J., Pratt, David R., Brutzman, Donald P., and Barham, Paul T. (1995): "Exploiting Reality with Multicast Groups: A Network Architecture for Large-scale Virtual Environments", in *Proc. 1995 IEEE Virtual Reality Annual International Symposium (VRAIS'95),* 11-15 March, 1995, RTP, North Carolina.

Musse, S. R., and Thalmann, D. (1997). A Model of Human Crowd Behavior: Groups Inter-Relationship and Collision Detection Analysis. In *Proceedings of the Eurographics Workshop* in Budapest, Hungary - Sept. 1997, Springer Wien NewYork.

Nielsen, J , Mack, R.L., (1994). *Usability inspection methods*. John Wiley and Sons, New York, NY.

Normand, V., Pernel, D. and Bacconnet, B. Speech-based Multimodal Interaction in Virtual Environments: Research at the Thomson-CSF Corporate Research Laboratories. To appear in *PRESENCE: Teleoperators and Virtual Environment*, December 1997.

Ousterhout, J. (1994). "Tcl and the Tk toolkit", Addison-Wesley Publishing Co.

Pouteau, X. (1993). Applying multimodality to command and control applications: the Melodia prototype. *Proceedings of the ERCIM (European Research Consortium for Informatics and Mathematics) workshop on Multi-modal Human-Computer Interaction,* Nancy, France, 2-4 November.

Rodden, T. R. (1996): "Populating the Application: A Model of Awareness for Cooperative Applications", *Proc. CSCW'96*, pp 67-76, November 1996, Boston, US, ACM Press.

Rosenblum, L., Durbin, J., Obeysekare, U., Silbert, L., Tate, D., Templeman, J., Agrawal, J., Fasulo, D., Meyer, T., Newton, G., Shaley, A. (1996). Shipboard VR: From Damage Control to Design, *IEEE Computer Graphics and Applications,* November 1996.

Sidner, C. (1986). Focusing in the comprehension of definite anaphora. In *Readings in Natural Language Processing*, pp. 363-394, Morgan-Kauffman.

Slater M., Usoh M. (1993). Body Centred Interaction in Immersive Virtual Environments, *Virtual Reality and Artificial Life*, University of Geneva, 25-26 November. Also in "Artificial Life and Virtual Reality", Eds. N. Thalmann and D. Thalmann, John Wiley & Sons, England, 1994, pp.125-147.Slater, M., Steed, A., McCarthy J., F. Maringelli, F., (1997a). The Influence of Body Movement on Presence in Virtual Environments, submitted for publication.

Slater M., Usoh M., Steed A. (1994). Depth of Presence in Virtual Environments, *Presence: Teleoperators and Virtual Environments*, MIT Press, 3(2), pp130-144.

Slater M., Usoh M., Steed A. (1995). Taking Steps: The Influence of a Walking Technique on Presence in Virtual Reality, In *ACM Transactions on Computer-Human Interface,* Special Issue on Virtual Reality Software and Technology, 2(3):201-219.

Slater, M., Steed, A., McCarthy J., F. Maringelli, F., (1997b) The Virtual Ante-Room: Assessing Presence through Expectation and Surprise, submitted for publication.

Slater, M., Chrysanthou, Y., (1997). View Volume Culling Using a Probabilistic Caching Scheme, *ACM Virtual Reality Software and Technology VRST'97,* Lausanne, Switzerland 1997.

Slater, M., Wilbur, S., (1997). A Framework for Immersive Virtual Environments (FIVE): Speculations of the Role of Presence in Virtual Environments, *Presence: Teleoperators and Virtual Environment.* 6(6), MIT Press.

Smith, G. and Rodden, T. (1995). "SOL : a toolkit for cooperative interfaces", *International Journal of Human Computer Studies (IJHCS).* No 42. Academic Press Limited. Pages 207-234.

Smith, G. (1996) "Cooperative Virtual Environments: lessons from 2D multi user interfaces". *Proc. Conf. Computer Supported Collaborative Work 1996 (CSCW '96).* Boston, Massachusetts, USA. November 16 20. Pages 390-398.

Smith, G. and Mariani, J. (1997). "Using Subjective Views to Enhance 3D Applications". *Proc. Conf. Virtual Reality Software and Technology 1997 (VRST '97).* ACM Press. Lausanne, Switzerland. Pages 130-146.

Stoakley, R., Conway, M., & Pausch, R. (1995). Virtual Reality on a WIM: Interactive Worlds in Miniature, *ACM CHI'95 Conference on Human Factors in Computing Systems*.

Thalmann, D., Shen, J., and Chauvineau, E. (1996). Fast Realistic Human Body Deformations for Animation and VR Applications. *Computer Graphics International'96*, Pohang, Korea, June, 1996, pp. 166-173.

Tromp, J.G., (1997). Methodology of Distributed CVE Evaluations, Poster Presentation at UK VR SIG’97, in: *Proceedings of UK VR SIG 1997*, Bristol.

Welsh, R.B., Blackman, T.T., Liu, A., Mellers, B.B., Stark, L.W., (1996). The Effects of Pictorial Realism, Delay of Visual Feedback, and Observer Interactivity on the Subjective Sense of Presence. *Presence: Teleoperators and Virtual Environments*, 5(3), MIT Press.

List of Figures

Figure 1: The COVEN consortium.

Figure 2: Screenshot of the Business Game in use, including the humanoid avatars described in section 4.1.

Figure 3: Screenshot of the Virtual Travel Agency in use in a network trial, with the standard dVS avatars.

Figure 4: The overall COVEN platform.

Figure 5: Scaling virtual humans.

Figure 6: An autonomous crowd.

Figure 7: Two ‘subjective views’ of the Interior Arrangement application.

Figure 8: Collaborative way finding - after creating the dynamic map, the lost traveller has the opportunity to inhabit it.

Figure 9: Model of the UCL area, part of the very large scene used for the Travel Rehearsal application.

Figure 10: Location and trials status of the COVEN Partners