Proceedings of the Fourth European Conference on Computer-Supported Cooperative Work,
September 10–14, Stockholm, Sweden
H Marmolin, Y. Sundblad, and K. Schmidt (Editors)

Virtual Reality Tele-conferencing: Implementation and Experience

Chris Greenhalgh
Department of Computer Science, The University of Nottingham, UK.

Steve Benford
Department of Computer Science, The University of Nottingham, UK.

This paper describes the implementation of and early experiences with a virtual reality tele-conferencing system called MASSIVE. This system includes a full realisation of the spatial model of interaction and its concepts of aura, awareness, focus, nimbus and adapters as was presented at ECSCW'93. This model supports users in interacting over ad-hoc combinations of audio, graphical and textual media through both 3-D and 2-D interfaces. Observations arising from the use of MASSIVE to support laboratory meetings are discussed; these include the need to support richer peripheral awareness, the need to improve the sensitivity of navigation, problems with lack of engagement between users, the need to support varying degrees of presence and problems arising from different perceptions of space between 2-D and 3-D users. Possible solutions to these problems are proposed.

1. Introduction

We describe the implementation of and initial experiences with a virtual reality based tele-conferencing system called MASSIVE in which communication between participants is controlled by movement within a shared virtual space. Specific design goals of this system include:

- MULTIPLE PARTICIPANTS: supporting groups of several participants at different locations in undertaking real time communication with one another.
- MUTLI-MEDIA: allowing these participants to communicate over different media. In particular, the system should support aural, visual and textual communication.
- HETEROGENEITY: allowing users with radically different interface equip-

ment to communicate within a common space. At the extreme, users of high end VR systems should be able to interact with users of, say, VT-100 character based terminals.

- SPATIAL MEDIATION: to support spatially mediated conversation management as opposed to traditional floor control. More specifically, a user's perception of others across different media should be governed by spatial factors such as their relative positions and orientations (e.g. people get louder as you move or turn towards them and vice versa).
- BALANCE OF POWER: there should be a balance of power between speakers and listeners so that (taking the audio medium as an example) speakers can try to influence who can hear them, e.g. by interrupting, and listeners can control who they are hearing.
- VARIED MEETING SCENARIOS: supporting a range of meeting scenarios ranging from face-to-face conversations to lectures and presentations;
- SIMULTANEOUS MEETINGS: allowing many simultaneous meetings to occur with the possibility for users to move between them.
- WIDE AREA: operating over wide area networks.
- SCALE: being capable of scaling to similar numbers of participants as are involved in every-day cooperative activities (e.g. tens or hundreds).

We propose that by meeting all of these goals we will eventually be able create more flexible, natural, open and scalable tele-conferencing systems than are currently available.

At the heart of our system lies the spatial model of interaction which was introduced at ECSCW'93 along with some early concept demonstrators. This paper describes a full implementation of this model and presents some initial observations arising from its use.

2. A brief overview of the spatial model

We begin by very briefly summarising the spatial model of interaction. Full details are given in the ECSCW'93 paper [Benford, 93]. The aim of the model is to support the flexible management of communication in densely populated virtual spaces. The model assumes the existence of some shared spatial frame of reference which is populated by human and other agents which communicate over combinations of different media.

The first component of the model, *aura*, addresses the problem of limiting the number of connections between the occupants of a densely populated space. In its simplest form, an aura is a subspace which scopes the presence of an object in a given medium. A connection between two objects is not made in this medium until the relevant auras collide (e.g. we cannot see each other until our visual auras collide or see each other until aural auras collide).

The concepts of *awareness*, *focus* and *nimbus* control the information passing across a connection once it has been established. An object may have different awareness of each connected object in each medium. Awareness is quantifiable and may range continuously from full, through peripheral to none. Having a low awareness of another object results in little information being received from it and a high awareness results in more detailed information. Thus, awareness provides a way of expressing desired quality of service across different connections. Awareness is medium specific and is interpreted differently for each medium (e.g. it may be mapped onto volume for an audio connection).

Mutual awareness need not be symmetrical and is controlled through focus and nimbus. Nimbus represents the transmitter's control over how information is propagated to other objects while focus represents the receiver's control. Focus and nimbus are typically expressed in terms of the spatial relationship of the objects (i.e. they are spatial fields), although they might also involve other attributes. Thus, the more object B is within object A's focus the more A is aware of B and the more A is within B's nimbus, also the more A is aware of B. More specifically, A's awareness of B in some medium M is a combination of A's focus in M and B's nimbus in M.

Finally, aura, focus and nimbus, and hence awareness, might be altered by various adapter objects. Adapters might represent communication tools such podia (aura and nimbus amplifiers), or boundaries (e.g. windows which attenuate audio awareness but not visual awareness), or other kinds of object; they provide a degree of extensibility to the model.

3. MASSIVE functionality

Our system is called MASSIVE (Model, Architecture and System for Spatial Interaction in Virtual Environments!). This section provides a user's view of MASSIVE's functionality.

Within any given instantiation of the system the MASSIVE universe is structured as a set of virtual worlds connected via portals. Each world defines a disjoint and infinitely large virtual space which may be inhabited by many concurrent users. Portals allow users to jump from one world to another.

Users can interact with one another over combinations of graphics, audio and text media. The graphics interface renders objects visible in a 3-D space and allows users to navigate this space with a full six degrees of freedom. The audio interface allows users to hear objects and supports both real-time conversation and playback of pre-programmed sounds. The text interface provides a MUD-like view of the world via a window (or map) which looks down onto an infinite 2-D plane across which user moves (similar in style to the UNIX games Rogue and Nethack). Text users are embodied using a few text characters and may interact by typing text messages to one another or by "emoting" (e.g. smile, grimace etc.).

A key feature of MASSIVE is that these three kinds of interface may be arbitrarily combined according to the capabilities of a user's terminal equipment. Thus, at one extreme, the user of a sophisticated graphics workstation may simultaneously run the graphics, audio and text clients, the latter being slaved to the graphics client in order to provide a map facility and to allow interaction with non-audio users. At the other extreme, the user of a dumb terminal (e.g. a VT-100) may run the text client alone. It is also possible to combine the text and audio clients without the graphics client and so on.

In order to allow interaction between these different clients a text user may export a graphics body into the graphics medium even though they cannot see it themselves. Similarly, a graphics user may export a text body into the text medium. In other words, text users can be embodied in the graphics medium and graphics users can be embodied in the text medium. MASSIVE uses a dynamic brokering mechanism (described below) to determine whether objects have any media in common whenever they meet in space (i.e. on aura collision). The net effect is that users of radically different equipment may interact, albeit in a limited way, within a common virtual world; for example, text users may appear as slow-speaking, slow moving flatlanders to graphics users.

All media (i.e. graphics, text and audio) are driven by the spatial model. Specifically:

- audio awareness levels are mapped onto volume; this means that audio interaction is sensitive to both the distance between and the relative orientations of the objects involved. This is observable in general conversation and also forms the basis of the "audio gallery" where users wander round a selection of audio-exhibits which play audio samples.
- graphics awareness levels are compared against threshold values to select one from a number of alternative object appearances according to the observer's location and orientation. This is typically used to display an object in more detail as awareness of it increases, although arbitrary changes are possible.
- the display of text messages is governed by levels of awareness as shown in table I, below; this lists awareness levels (values between 0 and 1) and the effects they have on the display of text messages.

Awareness	Level	Example Text Display
0.0-0.2	none	
0.2-0.4	presence	Chris at 0,0
0.4-0.6	events	"Chris says something"
0.6-0.8	peripheral	"(Chris says hi!)"
0.8-1.0	full	"Chris says hi!"

Table I: example levels of awareness for the text medium

Aura, focus and nimbus are attached to the user's current position and are therefore manipulated by moving about. Thus, turning towards another person may bring them more into ones focus or nimbus. In addition, users may explicitly manipulate awareness by choosing between three general settings for focus and nimbus:

- normal - provides conical focus and nimbus regions projecting out from the user which allows for full awareness of a few objects and peripheral awareness of other objects;
- narrow - a smaller aura and a thinner cone for focus and nimbus which enables private conversation (maximum awareness only occurs when two users are directly face to face, and there is little peripheral awareness);
- wide - a spherical region intended for general all round awareness (this nullifies the directional effects of focus and nimbus).

Users may also dynamically alter both the range and conical angle of focus and nimbus (aura is automatically updated when this happens). Thus, it is possible to arbitrarily widen and narrow focus and nimbus and to telescope them in and out to any desired range.

Four adapter objects have also been implemented:-

- A podium which extends the auras and nimbi of its users to cover a wider area, allowing them to address a crowd of other users;
- A conference table which replaces its users' normal auras, foci and nimbi with a new ones which span the table.
- A text to speech translator which converts messages in the text medium to synthesised speech in the audio medium (implemented using a public domain text to speech package).
- A text to graphics translator which displays messages received through the text medium on a "board" object in the graphics medium.

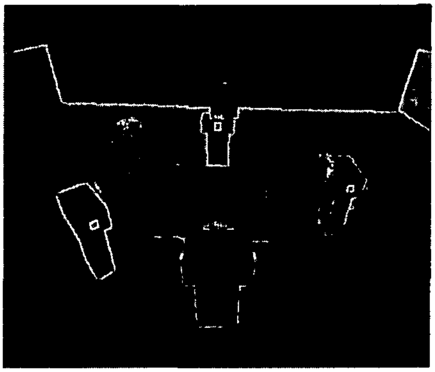
These adapters are themselves driven by the spatial model so that they only become active when a user gets sufficiently close to them. For example, a text interface user approaching the text-to-speech adapter will cause the adapter to activate and to automatically begin translating their text messages and re-transmitting them in the audio medium, enabling nearby audio users to hear them. Consequently, many users can use them simultaneously and can jostle around them to negotiate access.

A user's embodiment determines how they appear to other users. Each user may specify their own graphics embodiment in a personal configuration file using a simple geometry description format. In addition, we provide some default graphics embodiments intended to convey the communication capabilities of the users they represent (which is an important issue in a heterogeneous environment). For example, an audio user has ears, a non-immersive (and hence monoscopic) user has a single eye and a text user has the letter "T" embossed on their head. The aim of such embodiments is to provide other users with the necessary basic communication cues to decide how to address them. The basic shape of graphics embodiments is also intended to convey orientation in a simple and efficient manner. Graphics embodiments may be labelled with the name of the user they represent in order to aid identification. Finally, users may emote to one another by switching between different bodies by using key strokes.

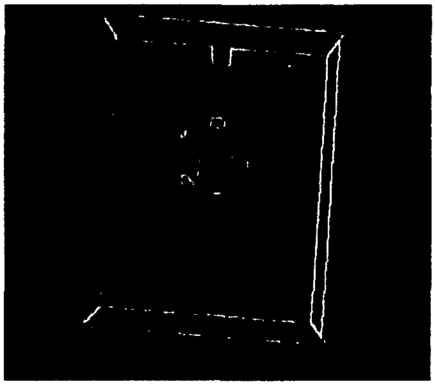
Text embodiments consist of a single character (the first letter of the person's chosen name) along with a short line which indicates the direction in which this person is currently facing.

Users may define any number of worlds containing simple graphics scenery and other objects. These worlds may be interconnected in any configuration via portals. An important aspect of MASSIVE is that in multi-site use across wide area networks each site may define its own local worlds; portals can then be used to allow users to move between sites in a transparent manner. Thus, each MASSIVE site can define its own conferencing environment as well as connecting to the broader "universe" of MASSIVE worlds across wide area networks.

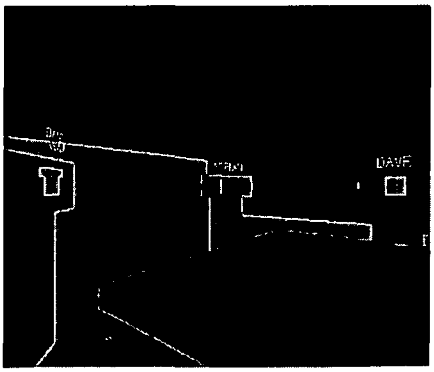
We complete this overview of MASSIVE's functionality with two sets of screen-shots.



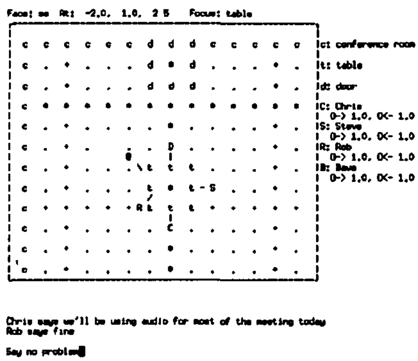
1 (a) perspective view



1 (b) bird's eye view



1 (c) eye-level view

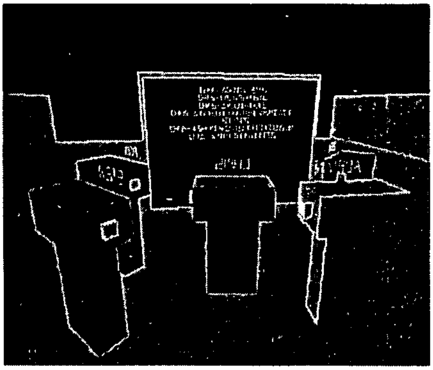


1 (d) text user's view

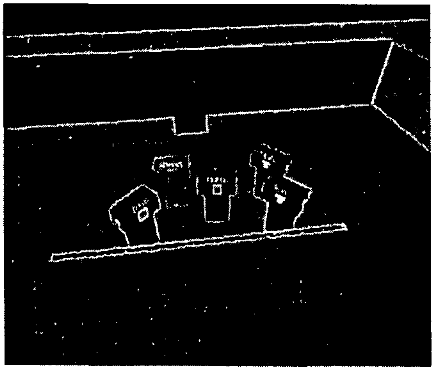
Figure 1: Five meeting participants around the "table" adapter

Figure 1 shows a meeting in progress involving five participants who are using the conference table adapter. Figure 1 (a) provides a perspective view of the scene and figure 1 (b) a birds-eye view (obtaining different views has been made possible by the recent viewpoint extensions described in section 5.2 below). Figure 1 (c) shows the default eye-level view that participants normally experience from inside their bodies. Finally, figure 1 (d) shows a text user's view of the same scene. Note the use of simple characters to represent the conference table, walls, door and users in the text view (see the key at the right of the image). Also note the display of mutual awareness levels for users of whom we are currently aware ("0->" denotes our awareness of them, while "0<-" denotes their awareness of us). The area at the bottom of the image shows the on-going text conversation.

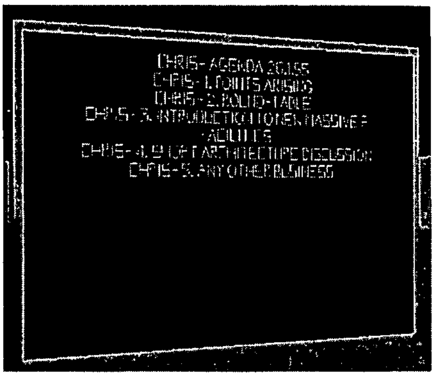
Figure 2 shows the same five participants using the text-to-graphics board adapter. When a user stands sufficiently close to the board and enters a message in a text client as in figure 2 (d), their message is automatically displayed on the board for graphics users to see as in figures 2 (a) to (c). Remember, graphics users can also run supplementary text clients as well.



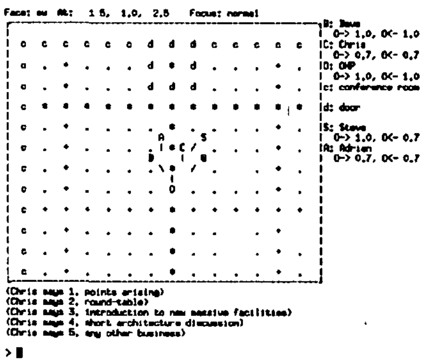
2 (a) front perspective



2 (b) rear perspective



2 (c) close up of the board



2 (d) text user's view

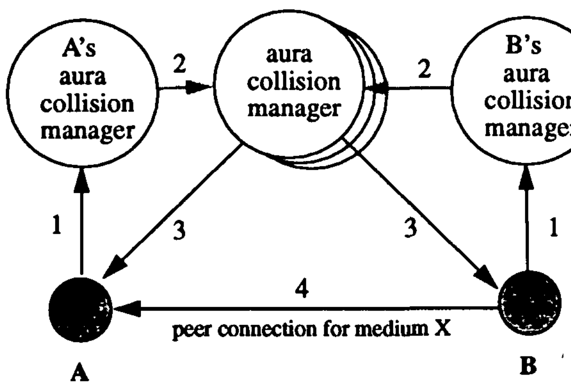
Figure 2: Five meeting participants using the "board" text to graphics adapter

4. MASSIVE Implementation

This section briefly describes some of the implementation techniques that have been introduced in order to provide the functionality described in the last section. In particular, we discuss the implementation of aura, focus, nimbus and adapters.

4.1. Auras and spatial trading

Interaction between objects only becomes possible if two conditions are met. First, it must be established that the objects involved support some at least one compatible medium. Second, these objects must become sufficiently proximate in order for their auras to collide. These two pre-conditions are reflected in the concept of *spatial trading*. Spatial trading combines the virtual reality technique of collision detection with the distributed systems concept of trading (e.g. [Van der Linden, 92]), or request brokering as it is sometimes called. To explain how spatial trading operates we follow the sequence of events which occurs when two objects enter a MASSIVE virtual world, move towards each other and begin to interact. This process is summarised in figure 3.



1. object declares its world and media to the local aura collision manager.
2. local aura manager passes the object on to the appropriate aura collision manager for that world.
3. aura collision manager detects collisions and passes out mutual interface references to peer objects.
4. peer objects exchange information via media controlled by awareness.

Figure 3: Objects involved in spatial trading

On entering a world, an object contacts the local spatial trader, called the *aura collision manager*, and declares the world which it wishes to join and the media which it supports. The address of this aura collision manager is the only information that an object requires in order to enter any local or linked world. An aura collision manager is responsible for detecting aura collisions for each declared medium in one or more worlds. Each aura manager has a partial locally-configured list of other aura managers and the worlds which they manage. Thus objects may be passed from one aura manager to another when they change worlds. A second object subsequently entering the world will go through the same procedure of declaring its world and media to its local aura collision manager and being passed on to the appropriate aura collision manager for that world.

Each aura collision manager monitors the auras of all objects known to it. Upon detecting an aura collision (within any given world and medium) the aura collision manager passes out mutual addresses to the objects involved, enabling them to establish a peer connection.

Notice how MASSIVE's implementation of spatial trading meets the goals of heterogeneity and scaleability. Heterogeneity is realised through the aura collision manager effectively registering all media and worlds currently active. This enables MASSIVE to dynamically cope with hitherto unseen media. Scaleability is supported by distributing the responsibility for detecting aura collisions between multiple aura collision managers, thereby avoiding excessive centralisation.

4.2. Focus and nimbus

Once connected through spatial trading the calculation of mutual awareness levels is the responsibility of the peer objects themselves. This is achieved through a simple peer protocol which allows any pair of objects to exchange information describing their positions and orientations and values of focus and nimbus. The communication protocol for each medium (e.g. graphics, audio or text) is derived by extending this basic peer protocol to handle additional medium-specific information (e.g. transmission of audio data in the case of the audio medium).

In the current implementation objects are described by a point location in space; focus and nimbus are described by mathematical functions which yield an awareness value in the range 0 (minimum) to 1 (maximum). Our current awareness function, which is used to combine focus and nimbus to give overall awareness, is multiplicative. I.e. focus and nimbus values are simply multiplied together to give awareness. This gives equal control to the observer and the observed, and is "subtractive" in nature - i.e. either party can force zero (no) awareness, but neither party can force awareness against the other's "wishes."

Our current focus and nimbus function has been designed to be general purpose so that, by changing the values of a few key parameters, a wide range foci and nimbi can be obtained. These parameters can be used to control the behaviour of focus and nimbus with respect to both the relative positions and orientations of objects. Thus, our three focus and nimbus settings and different adapters can all be realised by simply changing the values of a few key parameters while still using the same basic function code (see below). Figure 4 summarises our general focus/nimbus function using a polar coordinate model.

The left of the diagram shows how focus and nimbus are divided into three conical regions: a foreground region in which they take a maximum value; a background region in which they take some minimum value; and a transition region in which they change linearly from the foreground to the background value. The right of the diagram shows how the values of focus and nimbus depend on distance from an object and are again divided into three regions: they take the maximum value up to an initial radius; they then decay linearly to a cut off value at a second radius; beyond this, they tail off according to an inverse square law. Table II summarises the parameters which can therefore be used to control focus and nimbus.

4.3. Adapters

There are two issues to be dealt with when implementing adapter objects: how to trigger the use of an adapter and how to realise its effect on aura, focus, nimbus and awareness. Both of these issues are addressed through the introduction of a separate adapter medium. Adapters exist in their own medium, complete with its own aura, focus and nimbus. Any object wishing to use an adapter must therefore support this medium so that as the object moves about it will connect to adapters as a result of aura collisions in the adapter medium. When an object's awareness of an adapter crosses some threshold level the adapter is triggered. This mechanism enables several people to use an adapter simultaneously and also allows adapters to exhibit their own spatial properties (e.g. implementing a highly directional microphone).

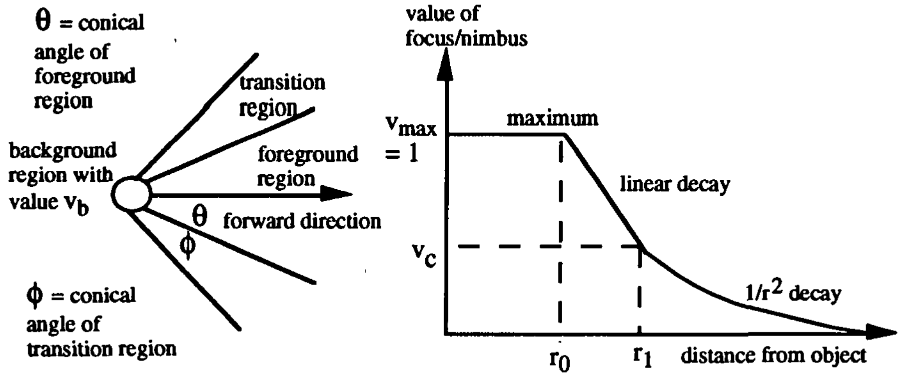


Figure 4: focus and nimbus function used by MASSIVE

name	meaning
θ	conical angle of foreground region
ϕ	conical angle of transition region
v_b	focus/nimbus value of background region
r_0	radius of extent of maximum value
r_1	radius of extent of linear transition
v_c	cut off value for linear transition

Table II: parameters affecting focus and nimbus

When triggered by an object, an adapter passes a new set of focus/nimbus parameters back to the object via the adapter medium. These new parameters replace the object's current aura, focus and nimbus parameters. Thus, an adapter may extend the range of focus or nimbus, may change their shape (i.e. conical angle) or may alter the way in which they fade to the background level. When an object subsequently moves away from an adaptor so that it is no longer triggered the object restores its original focus, nimbus and aura parameters.

Having discussed some key aspects of MASSIVE's implementation, we now turn our attention to some initial reflections arising from the implementation and early piloting activities.

5. Initial experiences

In this section we present some initial reflections on MASSIVE arising from recent experience. In particular, we reflect on two recent events: a laboratory meeting over the local area network in our own laboratory and a five site meeting between The University of Nottingham (UK), Lancaster University (UK), GMD (Germany), The Swedish Institute of Computer Science - SICS (Sweden) and the Royal Institute of Technology - KTH (Sweden), spanning three countries.

The laboratory meeting involved six participants connected over a single segment of Ethernet and lasted for half an hour. The hardware configuration was two SGI Indigo2s, a SUN 10 ZX and two SGI Indys, so that each participant was capable of using the audio, graphical and textual media. All but two of the participants were in physically separate rooms and even these two had their backs to each other and were using headphones. The six participants included the developer of MASSIVE, four users who had previously been involved in demonstrations and one novice user. The task was to conduct our weekly laboratory meeting, involving a round table presentation from each person followed by a loosely chaired free discussion. The view of one of the participants was captured on video and participants were asked to quickly write down their own reflections after the meeting's close.

The five site distributed meeting involved eight participants, three at Nottingham, one at Lancaster, two at SICS, one at KTH and one at GMD, and lasted for an hour and a half. All of the participants were audio/graphical. Each person was in a physically isolated space. Once again, the proceedings were videoed and participants were encouraged to write down their own observations.

The following informal observations constitute a rough and ready summary of what happened. Their main purpose is to identify some of the immediate and major issues that should be addressed in order to progress virtual reality tele-conferencing to a more useful state. Where appropriate we propose possible solutions.

5.1. It works!

First it must be stated that, technically at least, it works. It was straight-forward to install the software at each site onto standard machines via FTP and we would be confident of doing this at other sites. Configuring the wide area meeting took a little time and there were some minor teething problems, but nothing serious. It was not necessary to book time on networks or schedule conference calls. The meeting was open to as many participants as wanted to join at each site and people could come and go as they pleased.

Second, it was fun. The participants clearly enjoyed themselves and there were several light hearted moments (particularly involving the text-to-speech translator).

Third, the machines and the network coped, albeit under strain at times. Sometimes the graphics slowed down and the audio broke up (problems of packet based audio), but most of the time people could communicate.

The experience uncovered some interesting issues.

5.2. Limited peripheral awareness

A key goal of MASSIVE is providing the ability to separate what is immediate from what is peripheral. However, in the graphics medium, the current field of view seems to be too limited to provide a powerful sense of periphery at the edges of ones field of vision (although periphery in terms of distance *is* experienced). The screen based view has a default field of view of 64 degrees and although this can be widened (a parameter can be set in the graphics client code) larger fields of view introduce serious perspective distortion. Our current headmounted display, a Virtual Research EyeGen 3, has a field of view of width 40-50 degrees (although this wasn't used in the trial meetings). It is possible to buy headmounts with wider fields, but usually at the cost of lower resolution. Thus, in neither the screen based nor the immersive modes can we achieve anywhere near our real-world field of view of about 150 degrees width. The clearest indication of this problem was the difficulty experienced by participants in the, usually simple, act of forming a circle at the start of the laboratory meeting.

Our immediate solution to this problem has been to provide users with a choice of new 'camera angles' from which to view the world, coupled with the ability to zoom in and out on each of them. In addition to the normal 'in-body' view, users may now adopt a perspective view over the shoulder, a birds eye view, a front on view looking at themselves and side views. They may also adopt multiple simultaneous viewpoints which track one another (e.g. simultaneous in-body and birds eye views). In addition, given MASSIVE's flexible distributed architecture, it is easy to dynamically attach these additional viewpoints to other people, not just to oneself. Thus, one might view the world through someone else's eyes. In turn, this poses the question of how and when to configure different combinations of viewpoints. One approach might be to extend adapter objects towards being more general configuration management tools. For example, in addition to adapting my aura, focus and nimbus, the conference table adapter described above might automatically provide me with an additional birds eye view of the table while I am seated at it.

5.3.Navigation difficulties

There were numerous examples of people experiencing problems moving about, one of the most common being a tendency to fall backwards into portals through which one has just emerged. There was an obvious difference between novice and more experienced users which suggests a significant learning curve, but even experienced users still encountered problems. At a finer level of detail, current interaction techniques for moving one's virtual head and body appear too unwieldy to support rapid movement. This is particularly true when using a mouse to drive the screen based interface. When combined with a limited field of view this hampers the ability to use gaze direction or even body position to negotiate turn taking in conversation (see below). The use of magnetic tracking devices attached to the user's head may speed up interaction, but current devices still suffer from noticeable lag. The solution seems to lie in the development and use of better tracking devices and more "exotic" controls for screen based systems.

5.4. Lack of engagement

There were a number of breakdowns in the conversation, including several cases of participants being unsure as to whether they had been heard. Although there were some examples of back channels, these were generally few and far between. There might be several causes for this, including the lack of consistent audio quality and hence lack of confidence in being heard as well as considerable variability of microphone sensitivity. However, we suspect that there may be more general problems with engaging other users. In particular, even though the current graphics medium allows one to tell at a glance who is present in the current conversational group and who is approaching and leaving, lack of fine detail such as precise gaze direction make it hard to tell who is directly attending at any moment in time. Lack of visual feedback as to when people are speaking may be another factor here.

Immediate steps might involve improving the quality and reliability of the audio channel as well as the consistency of microphones and other audio hardware. Longer term work might involve analysing and reproducing key aspects of facial expressions such as eye-tracking and mouth movement as in the work of [Ohya, 93] and [Thalmann,93]. A small step has already been taken in this direction with the addition of a simple graphics mouth to the default MASSIVE embodiment which appears when the user speaks. Alternatively, one could consider texture mapping real time video onto embodiments as in the "Talking Heads" of [Brand, 87].

5.5. Degree of presence

Several times during the meetings, it became clear that the inhabitants of various embodiments had become involved in external activities and were not fully present. The most extreme case involved one user blatantly ignoring another even though they were being directly addressed. The problem here seems to involve conveying the degree of presence of different participants. This relates to the above problem of engagement and might be at least partially addressed through the same mechanisms (i.e. reproduction of dynamic user information such as facial expressions). However, one might also allow users to explicitly switch their body between different degrees of presence. In such cases, uninhabited bodies might act as markers or contact points for alerting their owners and inviting them to communicate (i.e. one would "prod" a body in order to grab the attention of its owner). Using the spatial model, one could construct a body which alerted its user only when directly addressed and which otherwise monitored background conversation (perhaps recording it).

5.6. Different perceptions of space

A more surprising observation concerns inter-working between 3-D graphics users and 2-D text users. Although they are mutually visible within a common space, their perception of that space seems quite different. In particular, the "texties" seem to lack any notion of personal space and tend to stand directly in front of others or even walk straight through them. In contrast, graphics users tend to maintain a reasonable distance from others. The problem may be that the graphics field of view is much more limited than the textual one (which is 360 degrees) so that the graphics users are forced to stand back in order to obtain a decent view. On the other hand, it may be that the graphics view is sufficiently rich for people to more easily associate the embodiments they see with other people and so feel compelled to behave in a socially polite manner in contrast to the text users. Either way, there appear to be some deeper issues involved when users with radically different interfaces interact in a common space.

6. Conclusions

This paper has described a prototype virtual reality based tele-conferencing system called MASSIVE. We begin our conclusions by considering how the implementation of MASSIVE meets the general design goals listed in the introduction.

- MULTIPLE PARTICIPANTS: the system demonstrably supports groups of at least six concurrent users.
- MUTLI-MEDIA: communication is possible in audio, visual and textual media.
- HETEROGENEITY: these three media can be arbitrarily combined accord-

ing to a user's terminal equipment and requirements. Furthermore, users may be embodied in media which they cannot display themselves (thus, text and graphics users can communicate). The concept of spatial trading has been introduced whereby the communication capabilities of users are dynamically matched whenever they become sufficiently proximate. Finally, text to speech and text to graphics translator adapter objects have been provided to further enhance cross-medium communication.

- SPATIAL MEDIATION: the implementation of the spatial model of interaction means that users' perceptions of one another in any given medium are sensitive to their relative positions and orientations; this is done with the intention of replacing traditional conference floor-control with a more autonomous and natural form of mediation.
- BALANCE OF POWER: conversation is influenced through movement, and everyone is free to move as they want at any time. Furthermore, support for both focus and nimbus means that the transmitter and receiver can both influence how any given utterance is eventually perceived. Adapter objects such as the podium alter this power balance without destroying it.
- VARIED MEETING SCENARIOS: in its most basic mode the system supports face-to-face conversation. However, the use of narrow focus and nimbus settings and also the conference table allow for more private discussions within a shared space. Similarly, the podium supports presentations and lectures to larger groups. So different worlds can be configured to support different meeting styles and sizes by including different adapters and scenery.
- SIMULTANEOUS MEETINGS: these are supported at several levels of granularity. First, different meetings may be held at the same time but in different worlds. Second, several meetings may be held in the same world at the same time, separated by simple partitions or just by distance. If these meetings are far apart they will be completely oblivious to one another; if they are close some mutual awareness may spill over (e.g. participants in one meeting may be able to see that the other meeting is happening without being able to hear what is being said). Participants are free to move between meetings at any time.
- WIDE AREA: Operation over wide area networks has been successfully demonstrated. This is encouraged by allowing each site to construct and master their own worlds locally, and then to connect them to remote ones via portals (similar to the way information is published on the World Wide Web).
- SCALE: The implementation of aura and spatial trading enhances the scaleability of system by removing the necessity for an object to maintain connections to all others objects all of the time. Providing multiple worlds also aids scaling. Finally, in a more pragmatic sense, the heterogeneous nature of MASSIVE encourages greater participation in worlds by allowing as many users as possible to participate using a wide range of technologies.

We are pleased to report that, from a technical perspective, the system works and has been used to hold multi-site meetings over wide area networks. However, several key issues have been identified that require further consideration including providing richer peripheral awareness, supporting easier and more rapid navigation, resolving problems with engagement, conveying varying degrees of presence and reconciling differences in perception between 2-D and 3-D users. These issues provide an agenda for future research along with the general problem of redesigning MASSIVE to support far greater numbers of users than at present.

To conclude, the MASSIVE system represents an early attempt to develop a collaborative virtual environment for tele-conferencing. We argue that, in spite of a number of challenges that have arisen, MASSIVE demonstrates the potential of such environments to go beyond our current tele-conferencing and shared space environments towards more flexible, natural and scalable future systems.

Acknowledgements

This work has been sponsored by the UK's Engineering and Physical Sciences Research Council (EPSRC) through their PhD studentship programme and by the Commission of the European Communities (CEC) through their ESPRIT Basic Research Programme. We would also like to thank Lennart Fahlén and John Bowers for their work on the spatial model. We would also like to thank Adrian Bullock, Rob Ingram, Dave Snowdon and Ok Ki Lee of the University of Nottingham; Andy Colbourne, Gareth Smith and Tom Rodden of Lancaster University; Lannart Fahlén of SICS; Kai-Mikael J-Aro of KTH; John Bowers of the University of Manchester; Mel Slater and Rob Kooper of QMW; and Dave England of GMD for volunteering to be cybernauts (we did consider using dogs, monkeys and rats at first, but they were too smart to volunteer ;-)).

References

[Benford, 93] Benford, S. and Fahlén,L., A Spatial Model of Interaction for Large Virtual Environments, In Proc. ECSCW'93 - Third European Conference on Computer Supported Cooperative Work, Milano, September, 1993, Kluwer Academic Publishers.

[Brand, 87] Brand, S., The Medialab - Inventing the future at MIT, Viking Penguin, 1987, ISBN 0-670-81442-3, p. 91-93.

[Carlsson, 93] Carlsson, C., and Hagsand, O., DIVE - A Platform for Multi-User Virtual Environments, Computer & Graphics, Vol 17, No. 6, 1993, pp. 663-669.

[Ohya, 93] Ohya, J., Kitamura, Y., Takemura, H., Kishino, F., Terashima, N., Real-time Reproduction of 3D Human Images in Virtual Space Teleconferencing, in Proc.VRAIS'93, IEEE, Seattle Washington September, 1993, pp. 408-414.

[Thalmann, 93] Thalmann, D., Using Virtual Reality Techniques in the Animation Process, in Virtual Reality Systems, Earnshaw, R.A., Gigante, M.A and Jones, H. (eds), Academic Press, 1993, ISBN 0-12-227748-1.

[Van der Linden, 92] Van Der Linden, R.J., and Sventek, J.S., The ANSA Trading Service, in IEEE Distributed Processing Technical Committee Newsletter, Vol 14, No 1, (Special Issue on Naming Facilities in Internet Environments and Distributed Systems), pp 28-34, June 1992.