MASSIVE: A Collaborative Virtual Environment for Teleconferencing

CHRIS GREENHALGH AND STEVEN BENFORD
The University of Nottingham

We describe a prototype virtual reality teleconferencing system called MASSIVE which has been developed as part of our on-going research into collaborative virtual environments. This system allows multiple users to communicate using arbitrary combinations of audio, graphics, and text media over local and wide area networks. Communication is controlled by a so-called spatial model of interaction so that one user's perception of another user is sensitive to their relative positions and orientations. The key concept in this spatial model is the (quantitative) *awareness* which one object has of another. This is controlled by the observing object's *focus* and the observed object's *nimbus*, which describe regions of interest and projection, respectively. Each object's *aura* defines the total region within which it interacts. This is applied independently in each medium. The system (and the spatial model which it implements) is intended to provide a flexible and natural environment for the spatial mediation of conversation. The model also provides a basis for scaling to relatively large numbers of users. Our design goals include supporting heterogeneity, scalability, spatial mediation, balance of power, and multiple concurrent meetings; MASSIVE meets all of these goals. Our initial experiences show the importance of audio in collaborative VR, and they raise issues about field of view for graphical users, speed of navigation, quality of embodiment, varying perceptions of space, and scalability.

Categories and Subject Descriptors: C.2.4 [**Computer-Communication Networks**]: Distributed Systems; H.1.2 [**Models and Principles**]: User/Machine Systems; H.4.3 [**Information Systems Applications**]: Communications Applications—*computer conferencing and teleconferencıng*; H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems —*artificial realities*; *audio input/output*; H.5.2 [**Information Interfaces and Presentation**]: User Interfaces—*interactıon styles*; *theory and methods*; H.5.3 [**Information Interfaces and Presentation**]: Group and Organizational Interfaces—*synchronous ınteraction*; *theory and models*; I.3.7 [**Computer Graphics**]: Three-Dimensional Graphics and Realism—*virtual reality*

General Terms: Design, Experimentation, Human Factors, Performance, Theory

Additional Key Words and Phrases: CSCW, scalability

This work has been sponsored by the United Kingdom's Engineering and Physical Sciences Research Council (EPSRC) through their Ph.D. studentship program and by the Commission of the European Communities (CEC) through their ESPRIT Basic Research Programme.

Authors' address: Department of Computer Science, The University of Nottingham, Nottingham, NG7 2RD, U.K.; email: {cmg; sdb}@cs.nott.ac.uk.

1. INTRODUCTION

The long-term goal of our research is the development of multiactor, distributed virtual environments to support cooperative work. Early experiences with multiactor VR have already been reported in the literature and include DIVE [Carlsson and Hagsand 1993], Rubber-rocks [Codella et al. 1992], NPSNet [Zyda et al. 1993], and the work of the ATR lab [Takemura and Kishino 1992]. Related ideas can also be found in multiuser recreational environments such as Habitat [Morningstar and Farmer 1991] and various Multiuser Dungeons (MUDS). The specific focus of this article is on the construction of a teleconferencing system in which communication between many 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.

—*Multimedia*: allowing these participants to communicate over different media. In particular, the system should support combinations of aural, visual, and textual communication.

—*Heterogeneity*: allowing users with radically different interface equipment to communicate within a common space. As an extreme case, users of high-end VR systems should be able to manage some kind of interaction 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 toward 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 among them.

—*Wide Area*: operating over wide area networks.

—*Scale*: being capable of scaling to similar numbers of participants as are involved in everyday cooperative activities (e.g., tens or hundreds).

We propose that by meeting all of these goals we will be able to create more flexible, natural, open, and scalable teleconferencing systems than are currently available. We have also been motivated by the observation that human beings have developed powerful social spatial skills in order to manage interaction with one another. In particular, spatial issues appear to be highly significant for controlling turn-taking (e.g., orientation and gaze direction) [Sacks et al. 1974] and other aspects of conversation management (e.g., joining and leaving groups). A shared spatial context also provides peripheral awareness of the presence and activity of others (an important issue in real-world cooperative work [Heath and Luff 1991]).

Our teleconferencing system is called MASSIVE (Model, Architecture, and System for Spatial Interaction in Virtual Environments). MASSIVE is underpinned by a so-called spatial model of interaction which has been developed within the ESPRIT COMIC project. The specific focus of this article is on the application of this spatial model to teleconferencing in the form of the MASSIVE system.

Section 2 of this article summarizes the spatial model of interaction as necessary background. Sections 3 and 4 then describe its implementation in MASSIVE, covering functionality and key implementation techniques respectively. Section 5 presents our initial reflections following testing within an experimental environment that includes both local and wide area networks. Section 6 then discusses performance issues and provides projections of the network bandwidth required to support increased numbers of users. Finally, Section 7 reflects on how MASSIVE meets our stated design goals and outlines issues for future work.

2. THE SPATIAL MODEL OF INTERACTION

Before describing MASSIVE, we first summarize the spatial model of interaction. The origin and details of the model have been described in previous papers [Benford and Fahlén 1993; Benford et al. 1994], and this section therefore only provides a sufficient grounding as is necessary for the remainder of the article.

The spatial model, as its name suggests, uses the properties of space as the basis for mediating interaction. Thus, objects can navigate through space in order to form dynamic subgroups and manage conversations within these subgroups. The spatial model may be divided into two main sections: facilitating scalability and controlling spatial interaction.

The first component—scalability—is based on the concept of *aura* as originally used in the DIVE system [Carlsson and Hagsand 1993]. Each object in a virtual world has an aura for each medium (visual, audio, text, etc.) in which it can interact. This aura defines the volume of space within which interaction is possible: interaction between two objects only becomes possible when their auras collide or overlap. Exactly what happens when auras collide depends on the details of the system architecture in use but will typically involve establishing a network connection between objects or an exchange of addresses or references. Figure 1 shows a number of objects and their auras for a single medium and shows where connections will exist. As objects move about these connections are destroyed and created as appropriate.

Auras may be defined in many applications by a region of space, i.e., an aura may often be a sort of bounding region surrounding an object and limiting its presence in space. However other definitions are possible; for example, an aura may be a continuously valued function over space or may be

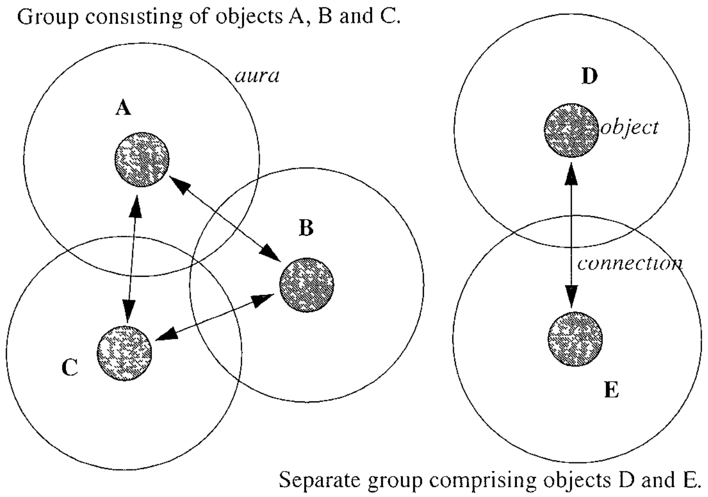

Group consisting of objects A, B and C.
aura
A
B
C
D
object
connection
E
Separate group comprising objects D and E.

Fig. 1. Object interactions based on auras.

defined by a function relating key properties of the objects in the space. The use of aura facilitates scaling to many users by limiting the number of object interactions that must be considered. For example, in the situation shown in Figure 1 only four peer relationships are active out of a possible ten. For sufficiently large virtual worlds, the number of interactions that must be considered will be governed by the extents of the objects' auras and by the population density of the space. So the world can be extended indefinitely at the edges of population, without increasing the apparent (local) complexity for objects closer to the center of the world.

The second component of the spatial model deals with the control of interaction or communication between two objects once communication has been enabled as a result of aura interactions. The main concept involved in controlling such interaction is *awareness*. One object's awareness of another object quantifies the subjective importance or relevance of the other object in a given medium. For example, awareness may be mapped to the volume of an audio channel or the level of detail of a graphical rendering. One object's awareness of another may range from full, through peripheral, to zero awareness. In general, more attention (and more bandwidth and computation) will be devoted to objects with higher awareness values. Awareness between two objects need not be mutually identical.

Mutual levels of awareness are negotiated between objects. Both of the objects in an interaction will wish to affect their levels of mutual awareness—the observing object (or receiver) will wish to focus its attention in particular areas or on particular objects, while the observed object (or transmitter) will wish to control its "visibility" so that objects in some areas are more aware of it than those in other areas (this applies in all media: the notion of "visibility" is applied by analogy to all media, not just the visual

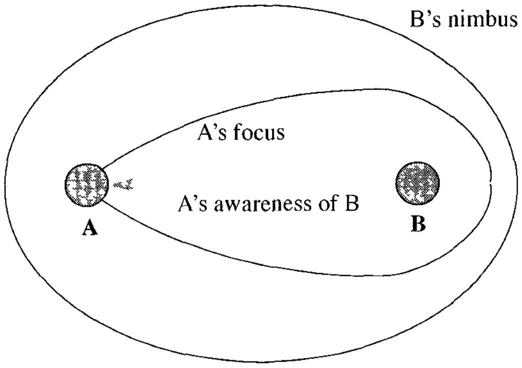


Fig. 2. Negotiating awareness between two objects.

medium). The concept of *focus* describes the observer's allocation of attention, while the concept of *nimbus* describes the observed object's manifestation or observability. So the more an object is within your focus the more aware *you* are of *it*, and the more an object is within your nimbus the more aware *it* is of *you*. The observer's awareness of the observed is then some combination of the observed's nimbus and the observer's focus. Formally, *the level of awareness that an object A has of an object B in medium M is some function of A's focus in M and B's nimbus in M.* Figure 2 shows diagrammatically the relationship between A, A's focus, B, and B's nimbus in the calculation of A's awareness of B in a single medium. In this particular case A will have a high awareness of B in this medium because A is within B's nimbus, *and* B is within A's focus.

As is the case for aura, focus and nimbus may range from simple bounded subspaces through to multivalued spatial fields with arbitrary sizes and shapes. Focus and nimbus are also medium specific. The choice of function to combine focus and nimbus will be application specific. For present purposes, the important point is that the resulting awareness levels between two objects can be used as the basis for managing their interaction. Exactly how this is achieved is again a matter of interpretation for a particular application. One approach is to use awareness levels to control the medium directly (e.g., controlling the volume of an audio channel between two objects). Another is to allow objects to react to the presence of other objects based on specified awareness thresholds (e.g., I might automatically receive text messages from you once a certain threshold has been passed).

The notion of spatial focus as a way of directing attention and hence filtering information is intuitively familiar from our everyday experience. The notion of nimbus requires a little more explanation. In general terms, a nimbus is a subspace in which an object makes some aspect of itself available to others. This could be its presence, identity, activity, or some combination of these. Nimbus allows an object to try to influence other objects (i.e., to project themselves or their activity in order be heard or seen). The need for a concept such as nimbus, to complement that of focus, becomes clear once we remember the notable phenomenon uncovered by Heath and Luff [1991] whereby

London Underground control room operators often deliberately projected their activity so as to encourage awareness among others. Thus, nimbus is the necessary converse of focus and is required to achieve a power balance in interaction.

Aura, focus, and nimbus, and hence awareness, may be manipulated in three general ways:

—Through fundamental spatial actions such as movement and orientation. Thus, as I move or turn, my aura, focus, and nimbus would typically follow me.

—Explicitly, through a few key parameters. For example, I might deliberately switch between wide or narrow settings for focus and nimbus.

—Through various adapter objects which modify focus and nimbus in some way.

In essence, an adapter is an object which, when activated, amplifies or attenuates aura, focus, or nimbus. There might be many types of adapter including:

—*Communication Tools*: for example, a user might step onto a "podium." In terms of the spatial model, a podium adapter object would then amplify their audio aura and nimbus. As a second example, the user might sit at a virtual "table." Behind the scenes, an adapter object would fold their auras, foci, and nimbi for several media into a common space with the other people already seated at the table, thus allowing a semiprivate discussion within a shared space.

—*Translators*: objects which translate between different media. For example, converting speech to text and vice versa.

—*Boundaries*: objects which divide space into different regions, constrain movement, and influence the flow of awareness [Bowers 1992]. For example, a "virtual door" might conditionally obstruct movement and attenuate awareness (the condition for movement being the possession of a key) whereas a "virtual window" might obstruct traversal and attenuate audio awareness but not influence visual awareness.

In summary, under the spatial model, objects move across space until their auras collide, which enables communication between them. This communication is subsequently managed according to mutual levels of awareness which are negotiated through the use of focus and nimbus. These are influenced in turn by adapter objects. It is important to remember that aura, focus, and nimbus operate on a per-medium basis and may range from simple containment spaces to arbitrary spatial fields. Having introduced the spatial model of interaction we now turn to the main subject of our article—its realization within the MASSIVE teleconferencing system.

3. MASSIVE FUNCTIONALITY: A USER'S VIEW

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 3D space and allows users to navigate this space with a full six degrees of freedom. The audio interface allow users to hear objects and supports both real-time conversion and playback of preprogrammed sounds. The text interface provides a MUD-like view of the world via a window (or map) which looks down onto an infinite 2D plane across which a user moves. 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 interfaces 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 nonaudio 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. One effect of this heterogeneity is to allow us to populate MASSIVE with large numbers of users at relatively low cost.

All media (i.e., graphics, text, and audio) are driven by the spatial model. This means that, first, interaction in a given medium is not possible until aura collision occurs in that medium. Thus, an object cannot be seen until graphics auras collide and cannot be heard until audio auras collide. Second, the information transferred across each of the three media is directly controlled through awareness values which are computed from focus and nimbus. 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 forms the basis of the "audio gallery" where users wander around a selection of audio exhibits (objects 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

Table I. Text Medium Awareness Levels and Their Effects

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!"

object in more detail as awareness of it increases, although arbitrary changes are possible; and

—the display of text messages is governed by mutual levels of awareness as shown in Table I. This lists awareness levels (values between 0 and 1) and the effects they have on the display of text messages.

Aura, focus, and nimbus are attached to the user's current position and are therefore manipulated by moving about. 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 allow for full awareness of a few objects (within that cone) and peripheral awareness of other objects.

—*Narrow*: a smaller aura and a thinner cone for focus and nimbus which enable private conversation (maximum awareness only occurs when two users are directly face-to-face, when there is little peripheral awareness).

—*Wide*: a spherical region intended for general all-around awareness (this nullifies the directional effects of focus and nimbus).

Four adapter objects are also provided:

—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 new ones which span the table.

—A text-to-speech translator which converts messages in the text medium to synthesized speech in the audio medium (implemented using a public domain text-to-speech package).

—A text-to-graphics translator which takes messages in the text medium and displays them on a large screen 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 automatically begin translating their text messages and retransmitting them in the audio medium, enabling nearby audio users to hear them. Consequently, many users can use them simultaneously and can jostle them around to negotiate access.

A user's embodiment determines how they appear to other users. Each user may specify his or her 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 nonimmersive (and hence monoscopic) graphics user has a single eye; and a text user has the letter "T" embossed on his or her 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 labeled with the name of the user they represent in order to aid identification. 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 the person is currently facing.

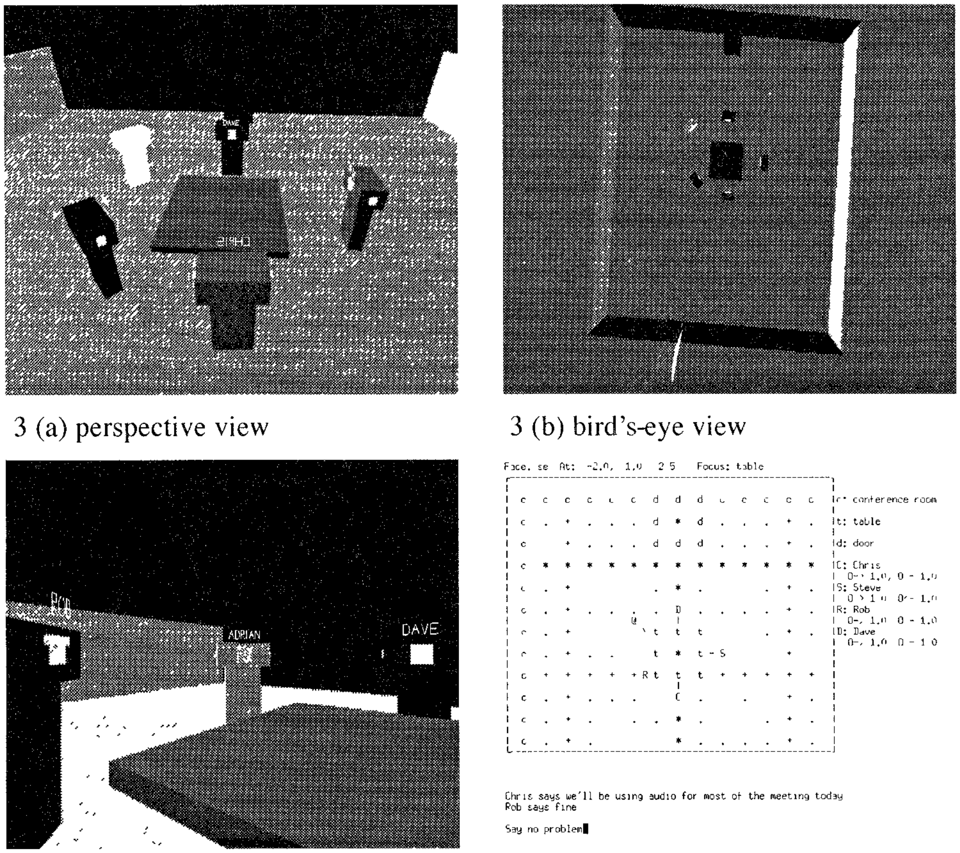
Users may define any number of worlds containing simple graphics scenery and other objects. These worlds may be interconnected in any configuration via portals. An important aspect of MASSIVE is that in multisite use across wide area networks each site may define its own local worlds; portals can then 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 connect to the broader "universe" of MASSIVE worlds. (We complete this overview of MASSIVE's functionality with two sets of screen shots.)

Figure 3 shows a meeting in progress involving five participants who are using the conference table adapter. Figure 3(a) provides a perspective view of the scene and Figure 3(b) a bird's-eye view (obtaining different views has been made possible by the recent viewpoint extensions described in Section 5.2). Figure 3(c) shows the default eye-level view that participants normally experience from inside their bodies. Finally, Figure 3(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 ("O−> " denotes our awareness of them, while "O < −" denotes their awareness of us). The area at the bottom of the image shows the on-going text conversation.

Figure 4 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 4(d), the message is automatically displayed on the board for graphics users to see, as in Figures 4(a)–(c). Remember, graphics users can also run supplementary text clients as well.

4. MASSIVE IMPLEMENTATION

This section describes briefly 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.



3 (a) perspective view

3 (b) bird's-eye view

```
Face. se  At: -2.0,  1.0   2 5    Focus: table
 -----------------------------------------------
|                                               |
|  c  c  c  c  c  c  d  d  d  c  c  c  c  c   |c: conference room
|                                               |
|  c  .  +  .  .  .  d  *  d  .  .  .  +  .   |t: table
|                                               |
|  c     +  .  .  .  d  d  d  .  .  .  +  .   |d: door
|                                               |
|  c  *  *  *  *  *  *  *  *  *  *  *  *  *   |C: Chris
|                                               |  0-> 1.0, 0 - 1.0
|  c  .  +           .  *  .           .  +  .   |S: Steve
|                                               |  0 > 1 0  0<- 1.0
|  c  .  +  .  .  .  .  D  .  .  .  .  +  .   |R: Rob
|                  @     |                      |  0-, 1.0  0 - 1.0
|  c  .  +             \ t  t  t        .  +  .   |D: Dave
|                                               |  0-, 1.0  0 - 1 0
|  c  .  +  .  .        t  *  t - S        +      |
|                                               |
|  c  +  +  +  +  + R t  t  t  +  +  +  +  +   |
|                        |                      |
|  c  .  +  .  .  .     C  .     .     +  .   |
|                                               |
|  c  .  +  .     .  .  *  .        .  +  .   |
|                                               |
|  c  .  +  .           *  .  .  .  .  +  .   |
 -----------------------------------------------

Chris says we'll be using audio for most of the meeting today
Rob says fine

Say no problem█
```

3 (c) eye-level view

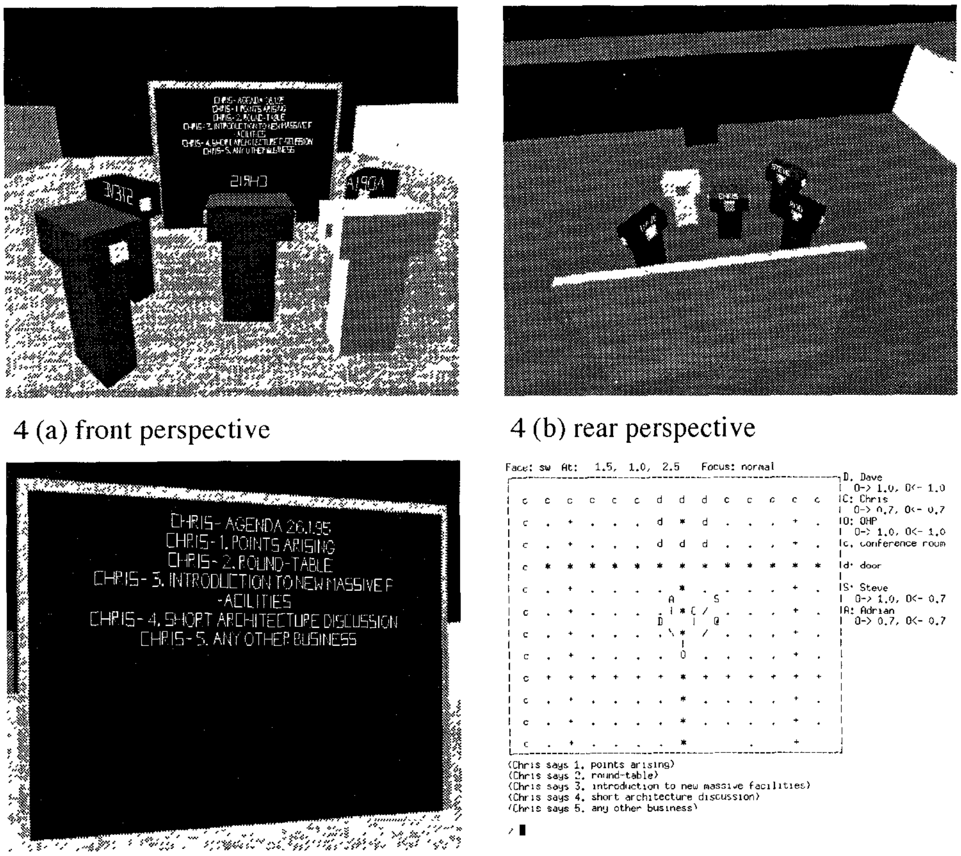
3 (d) text user's view

Fig. 3. Five meeting participants around the "table" adapter.

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 at least one compatible medium. Second, these objects must become sufficiently proximate in order for their auras to collide. These two preconditions 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 and Sventek [1992]) 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 toward each other, and begin to interact. This process is summarized in Figure 5.

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


CHRIS- AGENDA 26.1.95
CHRIS- 1. POINTS ARISING
CHRIS- 2. ROUND-TABLE
CHRIS- 3. INTRODUCTION TO NEW MASSIVE F
-ACILITIES
CHRIS- 4. SHORT ARCHITECTURE DISCUSSION
CHRIS- 5. ANY OTHER BUSINESS

4 (a) front perspective

4 (b) rear perspective

4 (c) close up of the board

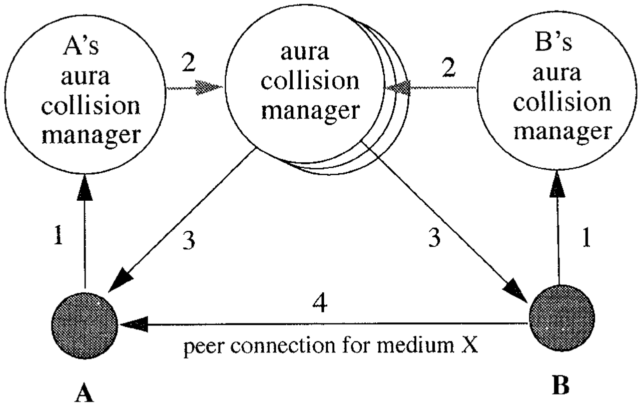
```
Face: sw  At:  1.5, 1.0, 2.5   Focus: normal
 ---------------------------------------------D. Dave
|                                            |  O-> 1.0, O<- 1.0
| c  c  c  c  c  c  d  d  d  c  c  c  c  c   |C: Chris
|                                            |  O-> 0.7, O<- 0.7
| c  .  +  .  .  .  d  *  d  .  .  .  +  .   |O: OHP
|                                            |  O-> 1.0, O<- 1.0
| c  .  +  .  .  .  d  d  d  .  .  .  +  .   |c, conference room
|                                            |
| c  *  *  *  *  *  *  *  *  *  *  *  *  *   |d: door
|                                            |
| c  .  +  .  .  .  .  *  .  .  .  .  +  .   |S: Steve
|                      A     S               |  O-> 1.0, O<- 0.7
| c  .  +  .  .  .  .  | * C /  .  .  .  +  .|A: Adrian
|                   D     |   @              |  O-> 0.7, O<- 0.7
| c  .  +  .  .  .  .  \ *   /  .  .  .  +  .|
|                         |                  |
| c  .  +  .  .  .  .     O  .  .  .  .  +  .|
|                                            |
| c  +  +  +  +  +  +  *  +  +  +  +  +  +   |
|                                            |
| c  .  +  .  .  .  .  *  .  .  .  .  +  .   |
|                                            |
| c  .  +  .  .  .  .  *  .  .  .  .  +  .   |
|                                            |
| c  .  +  .  .  .  .  *        .     +      |
 --------------------------------------------
(Chris says 1. points arising)
(Chris says 2. round-table)
(Chris says 3. introduction to new massive facilities)
(Chris says 4. short architecture discussion)
(Chris says 5. any other business)

/ █
```

4 (d) text user's view

Fig. 4. Five meeting participants using the "board" text-to-graphics adapter.

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 locally configured partial 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 for exchanging information.



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.

Fig. 5. Objects involved in spatial trading.

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

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."

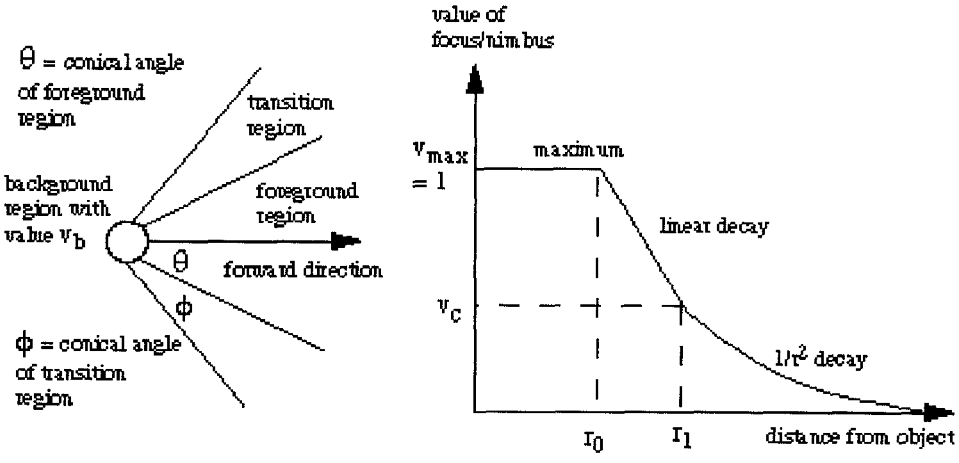


Fig. 6. Focus and nimbus function.

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 of foci and nimbi can be obtained. These parameters can be used to control the behavior 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 realized by simply changing the values of a few key parameters while still using the same basic function code (see below). Figure 6 summarizes 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 summarizes 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 realize 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 their 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

Table II. Parameters Affecting Focus and Nimbus

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

threshold level the adapter is triggered. This mechanism enables several people to use an adapter simultaneously and allows adapters to exhibit their own spatial properties (e.g., implementing a highly directional microphone).

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 moves subsequently away from an adapter 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 REFLECTIONS

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 three-site meeting between The University of Nottingham, Lancaster University, and Queen Mary and Westfield College, London, over the U.K.'s SuperJANET Wide Area Network.

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 three 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 write down quickly their own reflections after the meeting's close.

The three-site distributed meeting also involved six participants: three at Nottingham, one at Lancaster, and two at Queen Mary and Westfield College and lasted for an hour and a half. Five of the participants were audio/graphical users, and one was an audio/text user. This time, there were three experienced users and three novices, and each person was in a physically isolated space. There was no predetermined structure to the meeting other than to see whether three-site wide-area operation was even possible. Once again the proceedings were videotaped, 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 teleconferencing 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 straightforward 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. Clearly the participants 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 to provide 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 one's 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 head-mounted display, a Virtual Research EyeGen 3, has a field of view of width 40–50 degrees (although this was not 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 bird's-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 bird's-eye views). In addition, given MASSIVE's flexible distributed architecture, it is easy to attach these additional viewpoints dynamically 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 toward 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 bird's-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 backward 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 analyzing and reproducing key aspects of facial expressions such as eye-tracking and mouth movement as in the work of Ohya et al. [1993] and Thalmann [1993]. 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 [1987].

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 apparently completely 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 switch their bodies explicitly 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 interworking between 3D graphics users and 2D text users. Although they are mutually visible within a common space, their perception of that space seems quite different. In particular, the "texties" (text users) 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 associate more easily 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. NETWORK PERFORMANCE

Eventually, MASSIVE aims to support large numbers of users interacting in large and complicated spaces. This goal has been the driving force behind the introduction of the aura concept. Clearly, the current implementation of MASSIVE has not been tested with large numbers of users. This section therefore provides a *projection* of how performance might scale (focusing on network traffic requirements); it is based on measurements of user behavior from initial trials with small user groups and an understanding of MASSIVE's network protocols. The following gives preliminary network performance data for MASSIVE v.1.2.

For a single user running visual, audio, and textual interfaces there are four main sources of network traffic:

—coordinating the multiple user clients;

—keeping the aura collision manager up-to-date (i.e., notification of movement);

—interacting with other users and objects (during aura collision); and

—digital audio.

We will not consider the requirements of digital audio here: this subject is being dealt with more directly in the MBone community and elsewhere [Macedonia and Brutzman 1994]. Of the other three sources, the first two are independent of the other users in the world. The third, interaction with other users, is directly proportional to the number of other users in aura collision at any moment in time. We assume for this analysis that:

—The contribution of passive objects (e.g., scenery) is much less than that of users. MASSIVE's design goal is supporting interaction between users, and simple worlds have proved quite adequate for this task. As a rule of thumb, a MASSIVE passive object is broadly equivalent in terms of bandwidth to 0.2 normal users.

—Users move 25% of the time and when moving do so at 6Hz. These figures are based on data gathered from networked tests of MASSIVE with 6 users, the 6Hz update rate being on a Sun 10/ZX and seeming to be adequate for desktop use. In these tests, monitoring tools were used to log users' activities, from which an initial profile of typical behavior was constructed.

—Users move between worlds or groups of users at a rate such that they change the peers with whom they interact approximately once a minute.

—All users are interacting (on average) with a constant number of other users. Each user interacts with M others. For example, this could be as a result of users forming disjoint groups of $(M + 1)$ users or being spread evenly through space with an appropriate aura size relative to the density of users.

—All users are running visual, audio, and textual interfaces on comparable hardware.

Table III gives network demands for a single user given the above assumptions and includes all protocol overheads including acknowledgments and resends. The two principal events considered are a user moving (KB per

Table III. Network Traffic for a Single User

	Movement KB/step	KB/second	New peer KB/peer	KB/second	Total KB/second
Per user	1.2	1.8	2.1	< 0.1	1.8
Per peer per user	2.1	3.2	13.2	0.2	3.4

movement in column two and KB per second at 1.5Hz overall movement rate in column three) and a user gaining a new peer (i.e., starting to interact with another user, KB per occurrence in column four, and KB per second at a rate of once a minute in column five). The final column shows total bandwidth required in KB per second. The first row of figures shows how much network traffic is generated for a single user alone in the world (i.e., coordinating clients and informing the aura manager) while the second row shows how much additional traffic is generated per user for each peer with which a user typically interacts.

Hence, from the final column of the table, the total network traffic, T, in KB/second, for a total of N users each in aura collision with (i.e., interacting with) M other users is

$$T = N(3.4M + 1.8).$$

Clearly, for any given number of active peers, M (e.g., for constant group size), the total bandwidth is proportional to the total number of users. On the other hand, where all users are in a single group, $M = N - 1$ and

$$T = 3.4N^2 - 1.6N.$$

This is the limiting case where aura is effectively absent. Figure 7 shows how total bandwidth varies as a function of number of users for a range of group sizes (M). This illustrates the power of aura in converting what would otherwise be $O(N^2)$ bandwidth requirements to $O(MN)$. Note that the axes in the figures are all logarithmic.

Based on this analysis, Figure 8 shows possible group sizes versus number of users for different common network bandwidths. For example, the graph suggests that an Ethernet (fully loaded at 10Mb/second) could support 20 mutually aware users or about 80 users in groups of 5, while a 150Mb/second ATM connection is equivalent to just over 70 mutually aware users or about 1200 users in groups of 5.

7. CONCLUSIONS

This article has described a prototype virtual reality-based teleconferencing system called MASSIVE. We begin our conclusions by considering how MASSIVE meets the design goals listed in the introduction.

Multiple Participants. The system demonstrably supports groups of at least six concurrent users.

Multimedia. Communication is possible in audio, visual, and textual media.

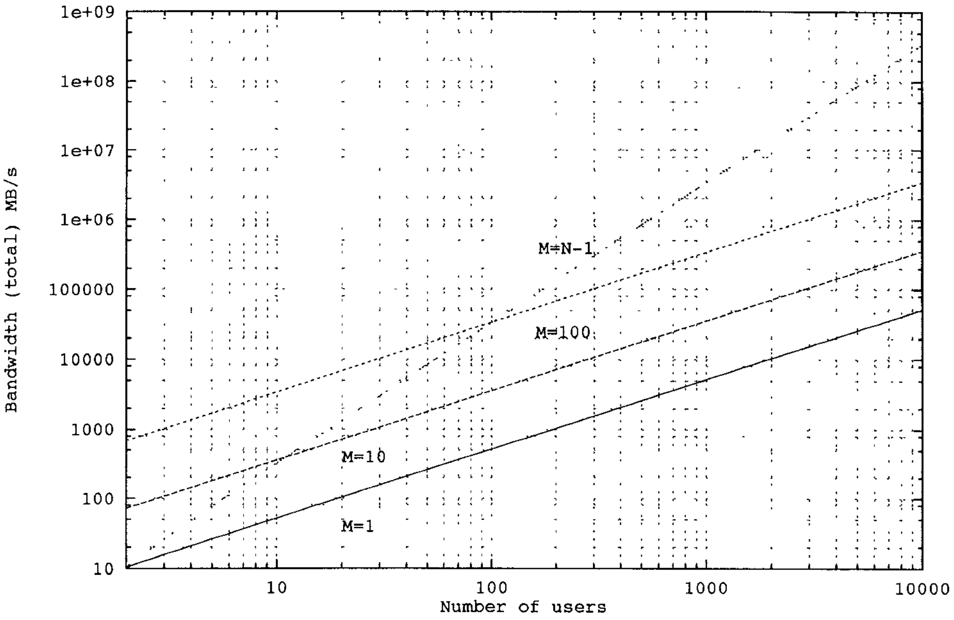


Fig. 7. Total bandwidth versus number of users for different group sizes.

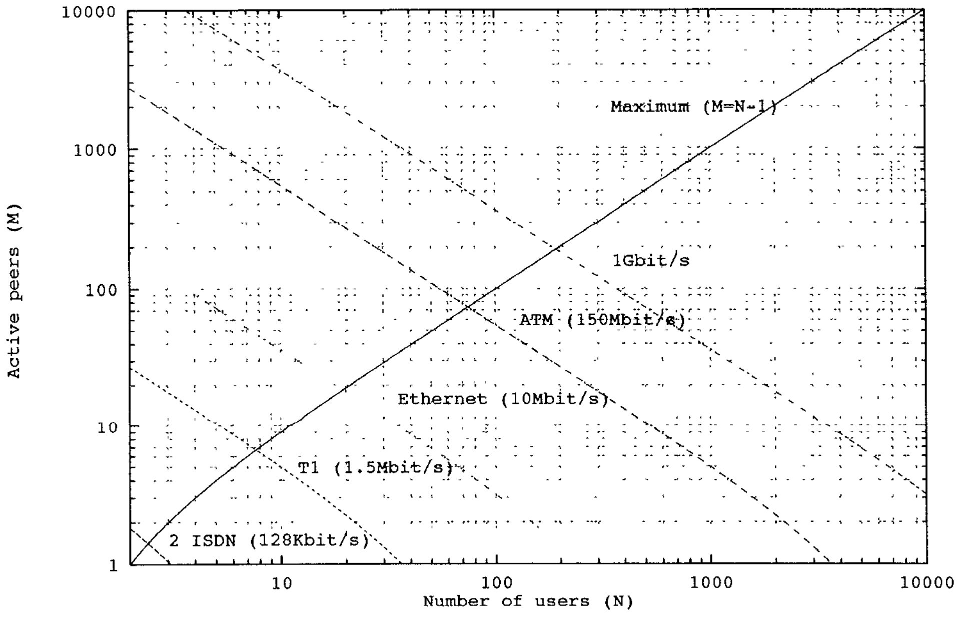


Fig. 8. Group size versus number of users for common bandwidths.

Heterogeneity. These three media can be arbitrarily combined according 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 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 sites to construct and master their own worlds locally and then to connect them to remove 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 scalability of the system by removing the necessity for an object to maintain connections to all other 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 multisite 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 2D and 3D users. These issues provide an agenda for future research. In addition, a number of other interesting future developments are apparent:

—*Supporting the Medium of Video with the Spatial Model*: Mutual awareness between different users in a virtual space might be used to configure the quality of service of video transmissions. Thus, high awareness would result in a high-quality image (e.g., large size, high resolution, and fast frame rate) whereas low awareness would result in a scaling down of these parameters. In this way, the spatial model might be used to manage multiple video streams in a flexible and dynamic way.

—*Constructing More-Interesting Spaces*: Our current meeting worlds tend to mimic real-world structures. We might construct more-abstract spaces based on the visualization of data of interest to a number of users. For example, stockbrokers might meet within a visualization of financial data, and users of the World Wide Web might meet within some visualization of the Web.

—*Speech and Gestural Control of Objects*: We have already seen how the spatial model in MASSIVE is used to trigger various adapter objects automatically. We might extend this to include the control of speech and gestural commands. For example, an object might only react to commands above a certain awareness threshold. The advantages of such an approach might be that speech and gestural command of objects could be naturally combined with conversation with other users and that, by setting focus and nimbus appropriately, multiple objects could be simultaneously controlled by multiple users.

Although no actual performance data for a large number of users is currently available, this article has provided a projection of network traffic for different numbers of users divided into different-sized aura groups. This projection is based on measurements of user behavior in smaller-scale trials combined with an understanding of MASSIVE's network protocols.

To conclude, the MASSIVE system represents an early attempt to develop a collaborative virtual environment for teleconferencing. 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 teleconferencing and shared-space environments toward more-flexible, natural, and scalable future systems.

ACKNOWLEDGMENTS

We thank Lennart Fahlén and John Bowers for their work on the spatial model.

REFERENCES

BENFORD, S. AND FAHLÉN, L. 1993. A spatial model of interaction for large virtual environments. In *Proceedings of ECSCW '93—3rd European Conference on Computer Supported Cooperative Work* (Milan, Italy, Sept.). Kluwer Academic, Dordrecht, Germany.

BENFORD, S., BOWERS, J., FAHLÉN, L., GREENHALGH, C., MARIANI, M., AND RODDEN, T. 1995. Networked virtual reality and co-operative work. *Presence*. To be published.

BOWERS, J. M. 1992. Modelling awareness and interaction in virtual spaces. In *Proceedings of the 5th Multi-G Workshop* (Kista-Stockholm, Sweden, Dec.).

BRAND, S. 1987. *The Medialab—Inventing the Future at MIT*. Penguin, New York, 91–93.

CARLSSON, C. AND HAGSAND, O. 1993. DIVE—A platform for multi-user virtual environments. *Comput. Graph. 17*, 6, 663–669.

CODELLA, C., LAWRENCE, R., KOVED, J., LEWIS, J. B., LING, D. T., LIPSCOMB, J. S., RABENHORST, F. A., WANG, C. P., NORTON, N., SWEENEY, P., AND TURK, G. 1992. Interactive simulation in a multi-person virtual world. In *Proceedings of CHI '92*. ACM Press, New York.

HEATH, C. AND LUFF, P. 1991. Collaborative activity and technological design: Task coordination in London Underground Control Rooms. In *Proceedings of ECSCW91* (Sept. 25–27), L. Bannon, M. Robinson, and K. Schmidt, Eds. Kluwer, Dordrecht, Germany.

MACEDONIA, M. R. AND BRUTZMAN, D. P. 1994. MBone provides audio and video across the Internet. *IEEE Comput. 27*, 4 (Apr.), 30–36.

MORNINGSTAR, C. AND FARMER, F. R. 1991. The lessons of Lucasfilm's Habitat. In *Cyberspace: First Steps*, M. Benedikt, Ed. The MIT Press, Cambridge, Mass., 273–302.

OHYA, J., KITAMURA, Y., TAKEMURA, H., KISHINO, F., AND TERASHIMA, N. 1993. Real-time reproduction of 3D human images in virtual space teleconferencing. In *Proceedings of VRAIS '93* (Seattle, Wash., Sept.). IEEE, New York, 408–414.

SACKS, H., SCHEGLOFF, E., AND JEFFERSON, G. 1974. A simplest systematics for the organisation of turn-taking in conversation. *Language 50*, 696–735.

TAKEMURA, H. AND KISHINO, F. 1992. Cooperative work environment using virtual workspace. In *Proceedings of CSCW '92* (Toronto, Canada, Nov.). ACM Press, New York.

THALMANN, D. 1993. Using virtual reality techniques in the animation process. In *Virtual Reality Systems*, R. A. Earnshaw, M. A. Gigante, and H. Jones, Eds. Academic Press, New York.

VAN DER LINDEN, R. J. AND SVENTEK, J. S. 1992. The ANSA trading service. *IEEE Distrib. Process. Tech. Commun. Newslett. 14*, 1, 28–34.

ZYDA, M. J., PRATT, D. R., FALBY, J. S., LOMBARDO, C., AND KELLEHER, K. M. 1993. The software required for computer generation of virtual environments. *Presence 2*, 2 (Spring), 130–140.

Received November 1994; revised February 1995; accepted February 1995