Multimedia Systems (1997) 5: 93–104

Multimedia Systems

Embodiments, avatars, clones and agents for multi-user, multi-sensory virtual worlds

Steve Benford[1], John Bowers[2], Lennart E. Fahlén[3], Chris Greenhalgh[1], Dave Snowdon[1]

[1] Department of Computer Science, The University of Nottingham, Nottingham, UK; e-mail: {sdb,cmg,dns}@cs.nott.ac.uk
[2] Department of Psychology, The University of Manchester, Manchester, UK; e-mail: bowers@hera.pych.man.ac.uk
[3] The Swedish Institute of Computer Science, Kista, Stockholm, Sweden; e-mail: lef@sics.se

Abstract. This paper explores the issue of user embodiment within collaborative virtual environments. By user embodiment we mean the provision of users with appropriate body images so as to represent them to others and also to themselves. By collaborative virtual environments we mean multi-user virtual reality systems which explicitly support cooperative work (although we argue that the results of our exploration may also be applied to other kinds of collaborative system). The main part of the paper identifies a list of embodiment design issues grouped by the general themes of personal representation, conveying activity, embodiment in heterogeneous systems, embodiment of agents, and ethical issues. These issues are illustrated with examples from our own DIVE and MASSIVE collaborative virtual environments. The paper also uses this set of issues as an analytical framework for comparing a number of other communication technologies.

Key words: Virtual worlds – Embodiments – Avatars – Clones

1 Introduction

User embodiment concerns the provision of users with *appropriate* body images so as to represent them to others (and also to themselves) in collaborative situations. This paper undertakes a theoretical exploration of this issue based on our experience of constructing and analysing a variety of multi-person virtual worlds or *collaborative virtual environments* (CVEs). Put in more simple terms, this paper addresses the issue of designing people's virtual bodies. We shall refer to such virtual bodies as *embodiments*, although a number of other terms are in general use including *clones* and *avatars* [Stephenson, 92].

The motivation for embodying users within collaborative systems becomes clear when one considers the role of our bodies in everyday (i.e. non-computer-supported) communication. Our bodies provide immediate and continuous information about our presence, activity, attention, availability, mood, status, location, identity, capabilities and many other factors. Our bodies may be explicitly used to communicate, as demonstrated by a number of gestural sign languages, or may provide an important accompaniment to other forms of communication, helping coordinate and manage interaction (e.g. so called "body language").

In our experience, user embodiment becomes an obviously important issue when designing collaborative virtual environments, probably due to their highly graphic nature and the way in which designers are given a free hand in creating objects. However, we believe that many of the issues we raise are equally relevant to cooperative systems in general, where embodiment often seems to be a neglected issue. Indeed, it appears that many collaborative systems still view users as people on the outside looking in and make no provisions for visualizing them *inside* the system. To go a stage further, we argue that without sufficient embodiment, users only become known to one another through their disembodied actions; one might draw an analogy between such users and poltergeists, only visible through paranormal activity. The basic premise of our paper is therefore that:

The inhabitants of collaborative virtual environments (and other kinds of collaborative system) ought to be directly visible to themselves and to others through a process of direct and sufficiently rich embodiment.

The key question then becomes how should users be embodied? In other words, are the body images provided appropriate to supporting collaboration? Furthermore, as opposed to merely discussing the appearance of the virtual body, we also need to focus on its functions, behaviours and its relation to the user's physical body (i.e. how is the body manipulated and controlled?). Thus, an embodiment can be likened to a 'marionette' with active autonomous behaviours together with a series of strings which the user is continuously 'pulling' as smoothly as possible.

Two general issues are worth clarifying at the outset. First, there is the question of whether to design highly realistic humanoid embodiments or more abstract representations? Considerable research effort has already been invested into techniques for modeling the human form. Notable successes include the Jack system [Badler, 93] and the work of the European Humanoid project. However, accurate human model-

Correspondence to: S. Benford

ing is a difficult and resource intensive task. We also suspect that success in this area may only serve to heighten users' expectations. In other words, the more human it looks, the more human it should behave. Finally, virtual embodiments may have to convey information that their real counterparts do not (see the "degree of presence" problem below). Consequently, whilst recognising the importance of realistic human modeling for a variety of applications such as ergonomics testing for vehicles or modeling clothes for the fashion industry, our paper chooses to focus on more abstract representations. Second, there is the question of whether we are designing embodiments for self-representation or for representation to others and whether these are subject to the same requirements? Again, considerable research has already been invested in the area of self-representation, including experimental research into self-embodiment in relation to proprioception and a user's sense of presence in a virtual environment [Slater, 93]. The focus of our paper is on representation to others as part of communication.

Our paper therefore aims to identify a set of design issues which should be considered by the designers of virtual bodies, along with a set of techniques to support them. These are introduced in Sects. 3–7, grouped under the headings of personal representation, conveying activity and focus of attention, heterogeneity, embodying agents, and ethical issues. We argue that designing an appropriate body image will most likely be a case of maintaining a sensible balance between these issues. Furthermore, this balance may be both application and user dependent and will no doubt be constrained by the available computing resources. In the long term, it may be possible to refine our initial list of issues into a 'body builder's work-out'. However, we do not yet have sufficient experience to do this. Instead, we illustrate each issue as it is introduced with examples of implementations or experiences arising from our own DIVE and MASSIVE systems. Section 8 then uses our list as a framework for analysing how a variety of other collaborative virtual environments and more general CSCW systems tackle the issue of user embodiment. However, before progressing to the design issues in detail, Sect. 2 first provides some brief background information on the underlying DIVE and MASSIVE systems.

2 Supporting systems

The authors' experience of user embodiment arises from the construction and use of two general collaborative virtual environments, DIVE at the Swedish Institute of Computer Science, and MASSIVE at the University of Nottingham.

2.1 DIVE

Virtual reality research at the Swedish Institute of Computer Science has concentrated on supporting multi-user virtual environments over local- and wide-area computer networks, and the use of VR as a basis for collaborative work. As part of this work, the DIVE (Distributed Interactive Virtual Environment) system has been developed to enable experimentation and evaluation of research results [Carlsson, 93][Fahlén, 93]. The DIVE system is a tool kit for building distributed VR applications in a heterogeneous network environment. In particular, DIVE allows a number of users and applications to share a virtual environment, where they can interact and communicate in real-time. Audio and video functionality makes it possible to build distributed video-conferencing environments enriched by various services and tools.

The DIVE system has been distributed worldwide and has been used to build a variety of applications. Figure 1 shows a screenshot from a virtual meeting in DIVE.

2.2 MASSIVE

MASSIVE (Model, Architecture and System for Spatial Interaction in Virtual Environments) is a VR conferencing system which realises the COMIC spatial model of interaction [Greenhalgh, 95]. The main goals of MASSIVE are scale, i.e. supporting as many simultaneous users as possible – and heterogeneity, i.e. supporting interaction between users whose equipment has different capabilities, who employ radically different styles of user interface and who communicate over an ad-hoc mixture of media.

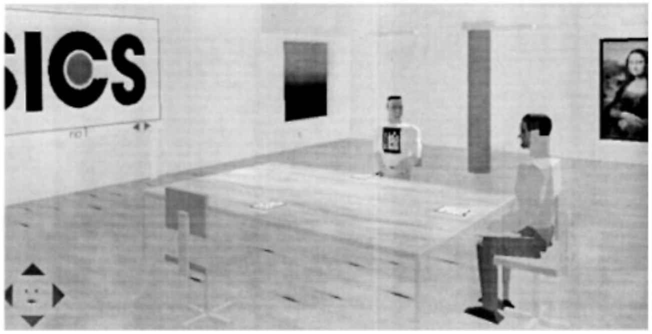
Like DIVE, MASSIVE supports multiple virtual worlds connected via portals. Each world may be inhabited by many concurrent users who can interact over ad-hoc combinations of graphics, audio and text interfaces. The graphics interface renders objects visible in a 3D space and allows users to navigate this space with a full six degrees of freedom. The audio interface allows users to hear objects and supports both real-time conversation and playback of pre-programmed sounds. The text interface provides a MUD (Multi-User Dungeon)-like view of the world via a window or map which looks down onto a 2D plane across which users move. An interesting feature of MASSIVE is inter-working between these different media (e.g. text-only users may interact with graphics users and vice versa). MASSIVE also contains a number of simple reactive objects or agents which react to a user's presence and actions including a text-to-speech converter object and a reactive whiteboard object.

MASSIVE has been used to hold regular meetings of up to nine simultaneous participants over the Internet, including a recent meeting spanning five organisations in the UK, Sweden and Germany. Figure 2 shows a screenshot from the graphics interface of a typical MASSIVE meeting where a number of people are gathered around a virtual conference table. Figure 3 shows a screenshot of the same meeting as seen through the text interface.

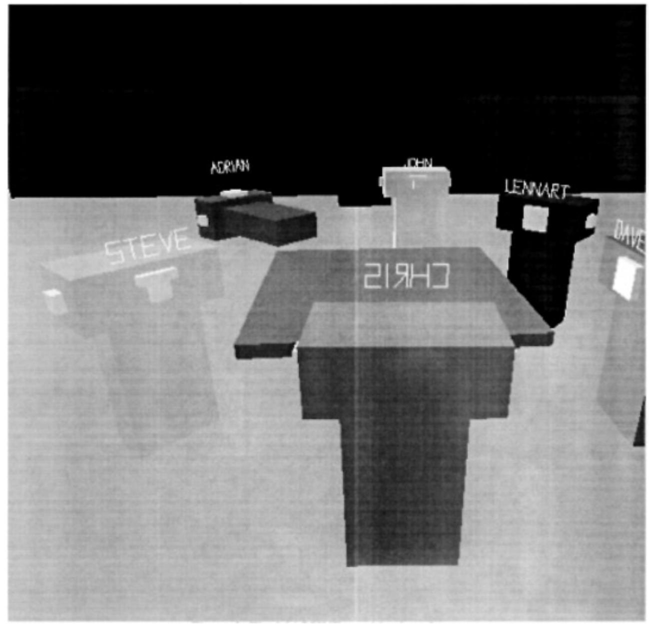
Having briefly described the DIVE and MASSIVE systems which have provided the foundation for developing our ideas on user embodiment, we now turn our attention to specific design issues and techniques, beginning with issues related to basic personal representation.

3 Personal representation

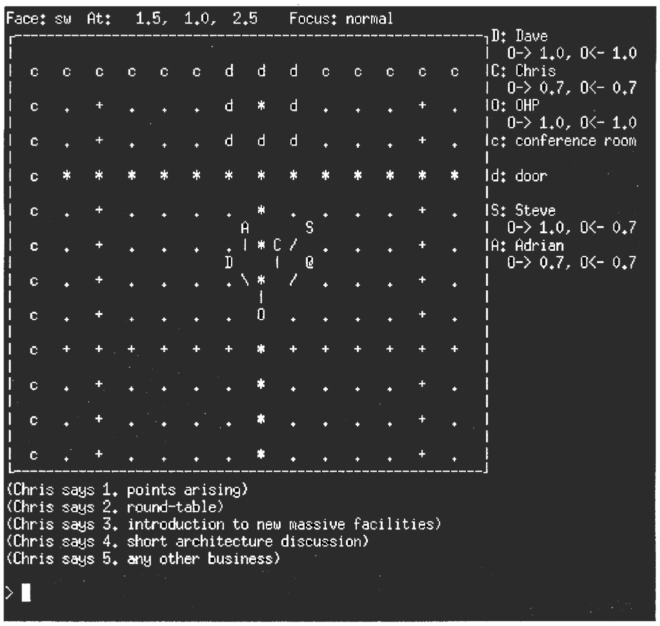
Personal representation concerns an individual's personal appearance.



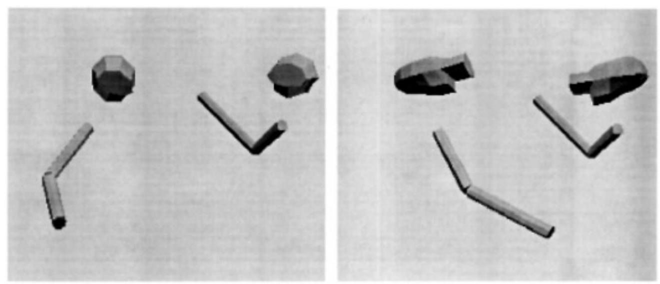
1



2



3



4

Fig. 1. A virtual meeting in DIVE

Fig. 2. The MASSIVE graphics interface

Fig. 3. The MASSIVE text interface

Fig. 4. Example viewpoint and actionpoint minimal embodiment

3.1 Identity

Recognising who someone is from their embodiment is clearly a key issue. In fact, body images might convey identity at several distinct levels of recognition. First, it could be easy to recognise at a glance that the body is representing a human being as opposed to some other kind of object. Second, it might be possible to distinguish between different individuals in an interaction, even if you do not know who they are. Third, once you have learned someone's identity, you might be able to recognise them again (this implies some kind of temporal stability). Underpinning these distinctions is the time span over which a body will be used (e.g. one conversation, a few hours or permanently) and the potential number of inhabitants of the environment (from among how many people does an individual have to be recognised?). Thus, we see the first of many trade-offs, in this case varying the complexity of an embodiment according to the scale and longetivity of the intended interaction.

In both DIVE and MASSIVE, the simplest form of embodiment is the "blockie", consisting of a few blocks sufficient to convey a general sense of orientation. MASSIVE also encourages the use of colour and name label to distinguish participants in small-scale meetings (see Fig. 2). DIVE has explored a range of more complex embodiments including the use of more humanoid shapes with basic moving limbs and also the use of texture-mapped photographs to provide a more unique form of identification (See Fig. 1). Finally, the MASSIVE text interface represents a user through the first character of their name (see Fig. 3).

3.2 Background

Embodiments might allow other users to obtain more detailed background information about their owners. For example, recent additions to DIVE support embedded URLs within object definitions, so that a user's body may contain a reference to their WWW home page. Querying a body then provides access to this information. This would seem to be a useful extension beyond the capabilities of our real-world bodies and would seem to offer the possibility of smoothing introductions and aiding people with poor memories for names or faces.

3.3 Personalisation

Allowing users to personalise body images is also likely to be important if CVEs are to gain widespread acceptance. Such personalisation allows people to create recognisable body images and may also help them to identify with their own body image. An example of personalisation might be the ability to don virtual garments or jewelry. Clearly, this ability might have a broader social significance by conveying status or associating individuals with some wider social group (i.e. cultural and work dress codes or fashions). However, one can imagine a number of problems with unconstrained personalisation. For example, in his novel Snowcrash, the author Neal Stephenson imagines a scenario where it is necessary to constrain the maximum height of a user's embodiment [Stephenson, 92] due to the importance of physical height in human communication.

Given that most respectable VR systems provide modeling languages for defining the appearance and basic behaviour of objects, the only technical limitation to personalisation would seem to be in the usability of modeling tools. Indeed, both DIVE and MASSIVE allow users to configure their own embodiments. However, perhaps due to the limited availability of such tools, very little personalisation has been observed in practice so far. Given this situation, it is not inconceivable to imagine a worldwide net-based virtual shop where one can buy standard-issue or custom-made embodiments in different formats and degrees of sophistication.

4 Conveying activity

Our bodies provide a powerful medium for conveying information about our ongoing activity.

4.1 Presence

The most basic form of activity is presence and the primary goal of a body image is therefore to convey a sense of someone's presence in a virtual environment. This should be done in an automatic and continuous way, so that other users can tell 'at a glance' who is present. In a visually oriented system (such as most VR systems) this will involve associating each user with one or more graphic objects which are considered to represent them.

4.2 Location

In shared spaces, it may be important for an embodiment to show the location of a user. This may involve conveying both position and orientation within a given spatial frame of reference (i.e. coordinate system). We argue that conveying orientation may be particularly important in collaborative systems due to the significance of orientation to everyday interaction. For example, simple actions such as turning one's back on someone else are loaded with social significance. Consequently, it will often be necessary to provide body images with recognisable front and back regions.

4.3 Viewpoints and actionpoints

Body images might convey a sense of ongoing activity. For example, position and orientation in a data space can indicate which data a given user is currently accessing. Such information can be important in coordinating activity and in encouraging peripheral awareness of the activities of others. We identify two further aspects of conveying activity: representing *viewpoints* and *actionpoints*.

A *viewpoint* represents where in space a person is attending and is closely related to the notion of gaze direction (at least in the visual medium). Understanding the viewpoints of others may be critical to supporting interaction (e.g. in controlling turn-taking in conversation or in providing additional context for interpreting talk, especially when spatial-deictical expressions such as 'over there' or 'here' are uttered). Furthermore, humans have the ability to register the rapidly changing viewpoints of others at a fine level of detail (i.e. tracking the movement of other's eyes even at moderate distances). Previous experimental work in the domain of collaborative 3D design has already shown the importance of conveying users' viewpoints [Shu, 94]. In contrast, an actionpoint represents where in space a person is manipulating. Actionpoints typically correspond to the location of virtual limbs (e.g. a telepointer representing a mouse or the image of a hand representing a data glove).

We propose that a user may possess multiple actionpoints and viewpoints. Notice that we deliberately separate where people are attending from where they are manipulating. Although these are often closely related, there appears to be no reason for insisting that they are strictly synchronized; in the real world, it is quite possible to manipulate a control while attending somewhere else – indeed, this is highly desirable when driving a car! Representing actionpoints involves providing an appropriate image of a limb driven by whatever device a user is employing. Representing viewpoint involves tracking where a user is attending and moving appropriate parts of their embodiment.

Embodiments in both DIVE and MASSIVE are able to convey general body orientation and also basic head movement. Even the extension of simple blockies to support a moving "head" which tracks that of an immersed user provides a very powerful mechanism for conveying broad gaze direction. A similar effect has been realised for desktop users of both MASSIVE and DIVE by providing a separate head controller which can be used to manipulate gaze separately from primary navigation of their body. However, neither DIVE nor MASSIVE support the necessary eye-tracking for high-resolution viewpoint detection (see below for examples of systems that do). The MASSIVE text interface also manages to convey orientation through the use of a simple line extending from the character that represents the user position.

Another interesting aspect of MASSIVE is that, in order to circumvent the problem of limited field of view, and hence peripheral vision, users are able to detach virtual cameras from their embodiment and adopt an “out of body view” (e.g. a perspective over the shoulder view of a bird’s eye view), effectively, separating their actual viewpoint from the embodied representation of their viewpoint. In DIVE there are a number of (user) pre-defined eye positions that the owner of the embodiment can switch between.

DIVE explicitly represents actionpoints in several ways. For desktop users, a line extending from a users body to the point of manipulation in space shows an actionpoint in a highly visible way. For immersed users, hand-held devices such as wands are explicitly represented in graphical form.

In Fig. 4, examples of an actionpoint/viewpoint minimalist virtual body is shown. This embodiment was designed by the Swedish artist Carl Johan Rydell, inspired by the very issues discussed in this article. The design fits very well with the “convention” of present VR systems to track the head (via the HMD) and the tracking of one hand by means of a glove or a wand.

4.4 Availability and degree of presence

Related to the idea of conveying activity is the idea of showing availability for interaction. The aim here is to convey some sense of how busy and/or interruptable a person is. This might be achieved implicitly by displaying sufficient information about a person’s current activity or explicitly through the use of some indicator on their body. This leads us to the further issue of degree of presence. Virtual reality can introduce a strong separation between mind and body. In other words, the presence of a virtual body strongly suggests the presence of the user when this may not, in fact, be the case. This is particularly likely to happen with ‘desktop’ (i.e. screen-based VR) where there is only a minimal connection between the physical user and their virtual body. Indeed, we have recorded several examples of this phenomena in our use of DIVE and MASSIVE. For example,

- user’s being distracted by events in the real world;
- a user process crashing and leaving behind a “corpse” entry in the world database, and
- users’ maintaining a partial presence in several locations by logging in more than once.

Such mind/body separations can cause a number of problems such as the social embarrassment and wasted effort involved in one person talking to an empty body for any significant amount of time. As a result, it may be important to explicitly show the degree of actual presence in a virtual body. For example, the system might track a user’s idle time and employ mechanisms such as increasing translucence or closing eyes to suggest decreasing presence. This has been implemented within DIVE by researchers at Lancaster University in the UK, who have used changing colour to indicate idle time. It might also be possible to put one’s body into a suspended state, indicating partial presence to others and perhaps recording ongoing conversation to be replayed when subsequently woken up. This has been realised in a simple manner in MASSIVE through the provision of a “sleep” gesture. We have also seen that other less formal conventions have been established among DIVE users. For example, on meeting a stationary embodiment, one grabs it and gives it a shake (DIVE allows you to pick other people up). An angry reaction tells you that the embodiment is occupied.

4.5 Gesture and facial expression

Gesture is an important part of conversation and ranges from almost subconscious accompaniment to speech to complete and well-formed sign languages for the deaf. Support for gesture implies that we need to consider what kinds of ‘limbs’ are present. Facial expression also plays a key role in human interaction as the most powerful external representation of emotion, either conscious or sub-conscious. Facial expression seems strongly related to gesture. However, the granularity of detail involved is much finer and the technical problems inherent in its capture and representation correspondingly more difficult. A crude, but possibly effective approach, might be to texture map video onto an appropriate facial surface of a body image (e.g. the “Talking Heads” at the Media Lab [Brand, 87]). Another approach involves capturing expression information from the human face, using an array of sensors on the skin, modeling it and reproducing it on the body image (e.g. the work of ATR where they explicitly track the movement of a user’s face and combine it with models of facial muscles and skin [Ohya, 93] and also the work of Thalmann [Thalmann, 93] and Quéau [Quéau, 94]).

This discussion of gesture and facial expression relates to a further issue, that of voluntary versus involuntary expression. Real bodies provide us with the ability to consciously express ourselves as a supplement or alternative to other forms of communication. Virtual bodies can support this by providing an appropriate set of limbs and ‘strings’ with which to manipulate them. The more flexible the limbs; the richer the gestural language. However, we suspect that users may find ways of gesturing with even very simple limbs. On the other hand, involuntary expression (i.e. that over which users have little control) is also important (looks of shock, anger, fear etc.). However, support for this is technically much harder as it requires automatic capture of sufficiently rich data about the user. This is the real problem we are up against with the facial expression issue – how to capture involuntary expressions.

In terms of our current systems, DIVE supports the display of real-time video streams within 3D environments which may be used to integrate more traditional video-conferencing users. MASSIVE, on the other hand, provides a selection of simple pre-programmed gestures such as sleeping and blushing.

4.6 Relationships with objects and information

In applications which involve access to large volumes of information such as the sharing of information in populated information terrains [Benford, 95] or the emergence of VR browsers for the World Wide Web (e.g. Silicon Graphic's Webspace), it may be important to explicitly represent various relationships between people and information objects. For example, queries to an embodiment might be used to uncover related items of information (e.g. "show me all documents authored by this person"). Conversely, queries to information may be used to uncover related people (e.g. "show me the current editor of this document"). For example, in fields such as software engineering it may be important to understand who is currently working on a particular system component. Thus, we can imagine extending software visualisations to include visualisation of relationships such as "is responsible for" or "is editing" between information objects and user embodiments.

One application of DIVE, called VR-VIBE, supports the process of collaborative information retrieval from online bibliographies [Benford, 95]. VR-VIBE allows multiple embodied users to see each other exploring and manipulating a shared document visualisation. In addition, users are able to explore various relationships between people and documents either by querying embodiments or document icons. Current relationships include *authorship*, *reading*, *editing*, *has read* and *has edited*. For example, having selected the "authorship" relationship, selection of another user's body then has the effect of highlighting all documents that they have authored, and selection of a document will show its authors (either as a marker embodiment if they are not currently present or by highlighting their actual embodiment if they are).

4.7 History of activity

Embodiments might support historical awareness of past presence and activity. In other words, conveying who has been present in the past and what they have done. Clearly, we are extending the meaning of 'body' beyond its normal use here. One example might be carving out trails and pathways through virtual space in much the same way as they are worn into the physical world. Another example is given by the above relationships between people and information such as "has read" and "has edited", which convey information about the past activities of other users.

As an example, we have extended the standard body actor in Division's dVS system to provide support for showing the history of activity of a user in a shared environment. In dVS the body actor creates an embodiment representing the user and controls the input and output devices associated with that user. In our extended body actor, history of activity is shown by leaving trails representing the path that the user has taken through the environment. These trails consist of a series of icons, placed at regular intervals and oriented to show the direction of travel. The user can configure both the geometry and colour of these icons so that it is possible to identify which user left a given trail. In order for the environment not to become cluttered with 'stale' trails, the user is also able to specify how frequently their embodiment will leave a trail icon and the length of time the icons in the trail will persist before they are removed.

5 Heterogeneity

Our next set of embodiment issues relate to the general issue of heterogeneity in potentially densely populated spaces. Heterogeneity arises from the observation that in scalable systems one cannot assume that everyone uses the same equipment, or even equipment with the same general capabilities. For example, in a heterogeneous system, immersed users may interact with desktop users; users of PCs may interact with users of graphics super-computers; and, in MASSIVE, text-only users may interact with graphics users.

5.1 Manipulating one's view of other people

In heterogeneous systems, it will be important for an observer to be able to control their view of other people's bodies. For example, as the user of a sophisticated graphics computer, I may have the processing power to generate a highly complex and fully textured embodiment. However, this is of little benefit to an observer who does not have a machine with hardware texturing support. Indeed, the complexity of my body would be counter-productive as the observer would be forced to expend valuable computing resources on rendering my body when it could better be used to render other objects. In the worst case, my body might act as a black hole, such that, once looked at, it could not easily be turned away from! As a result, the observer should be able to exert some influence over how other people appear to them, perhaps selecting from among a set of possible bodies the one that most suits their needs and capabilities. In short, we propose that it is important for the both the owner and the observers of an embodiment to control how it appears.

This requirement poses a serious problem for most of today's multi-user VR systems – that of subjective variability. Current systems are highly objective in their world view. In other words, all observers see the same world (albeit from different perspectives). A notable exception in this regard is the VEOS system [Bricken, 94]. The ability for people to adopt subjective world views (e.g. seeing different representations of an embodiment) represents a challenge to current VR architectures.

5.2 Representation across multiple media

Up to now we have spoken mainly in terms of visual body images. However, body images will be required in all available communication media including audio and text. For example, audio body images might centre around voice tone and quality, be it that of the real person or be it artificial. Text body images (as used in multi-user dungeons) might involve text names and descriptions or (in a collaborative authoring application) a text body's 'limbs' might be represented by familiar word processing tools and icons (cursor, scissors etc.). Thus, in the MASSIVE system, we see examples of both graphics and text-only embodiments.

5.3 Cross-medium embodiment and mutuality

Systems supporting radically different modes of interaction (e.g. text and graphics or immersive and desktop) need to consider issues of cross-medium embodiment and mutuality. Cross-medium embodiment refers to the ability to project an embodiment into a medium even if that medium cannot be perceived by its owner. Mutuality then refers to the question of whether such cross-medium embodiment (or indeed, embodiment in the same medium) should be symmetrical. In other words, can users be invisible to one another?

As an example, MASSIVE supports cross-medium embodiment between the text and graphics media. Thus, a text-only user sitting at, say, a VT-100 terminal may still project a graphics body for graphics users to see. Conversely graphics users may still appear in the text medium. In MASSIVE, the choice of whether to export such additional embodiments is local to each object and so it is also possible for users to be invisible to one another.

5.4 Capabilities

In a heterogeneous system, one may not always be aware of the capabilities of another user's terminal equipment. For example, on meeting another person in MASSIVE how is one to know whether to speak to them or type text messages? Embodiments might provide clues as to the communication, and other, capabilities of their owners. For example, the presence of ears on an embodiment might suggest the capability to process audio signals. Conversely, creating bodies with redundant features might be misleading (e.g. is it wise to create ears if the owner cannot hear)?

MASSIVE embodiments have been deliberately designed to convey communication capabilities. Thus, ears and a moving mouth suggests audio capability; a single eye suggests mono-scopic graphics capability; and a letter 'T' on the forehead suggests text-only capability (see Fig. 2).

5.5 Efficiency

There will always be a limit to available computing and communications resources. As a result, embodiments should be as efficient as possible, by conveying the above information in simple ways. Furthermore, we need to support 'graceful degradation', so that users with less powerful hardware or simpler interfaces can obtain sufficiently useful information without being overloaded. This suggests prioritizing the above issues in any given communication scenario. In fact, the real challenge with embodiment will be to prioritize the issues listed in this section according to specific user and application needs and then to find ways of supporting them within limited computing and bandwidth resources.

5.6 Embodiments as flocks

We have discussed virtual bodies as if they are localised within some small region of space. We may also need to consider cases where people are in several places at a time, either through multiple direct presence or through some kind of computer agent (see below) acting on their behalf. An example could be a scout sent out to map the virtual terrain. Furthermore, it could be useful to be able to attach "capabilities" to ones embodiment in a distributed and rather loose way. This then also implies that it might be possible to borrow or steal capabilities from other users!

Finally, let us point to an interesting distributed embodiment designed by the SF writer Vernor Vinge. In one of his books [Vinge, 92], he describes an alien life form, which is in the form of a "pack" of wolf- or dog-like creatures all of which play different roles. The pack's power and importance is completely defined by the different capabilities the different parts (the members) contribute. The being has one virtual mind that is constituted by the "sub-minds" being in varying degrees of telepathic contact.

6 Embodying agents

In this section, we turn our attention to the embodiment of non-human agents in virtual environments. Such agents might range from relatively simple tools to complex semi-intelligent processes.

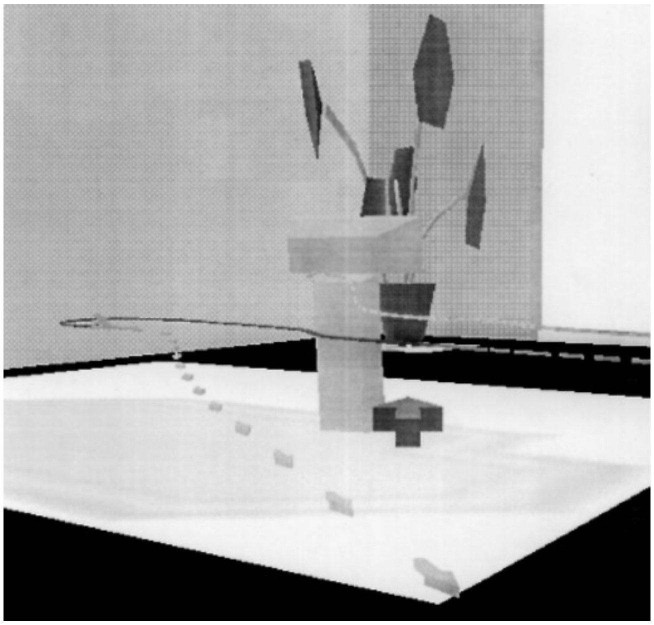
6.1 Engagement of attention

It is important to convey whether an agent's attention has been engaged. This is particularly true of agents which receive commands through broadcast media which are also used for general conversation (e.g. agents which are controlled by speech, gesture or text in a collaborative setting). Collaborative settings also give rise to the problem of multiple users trying to interact with the same agent. In this case, it may be important to show which users have managed to engage the agent at any moment in time.

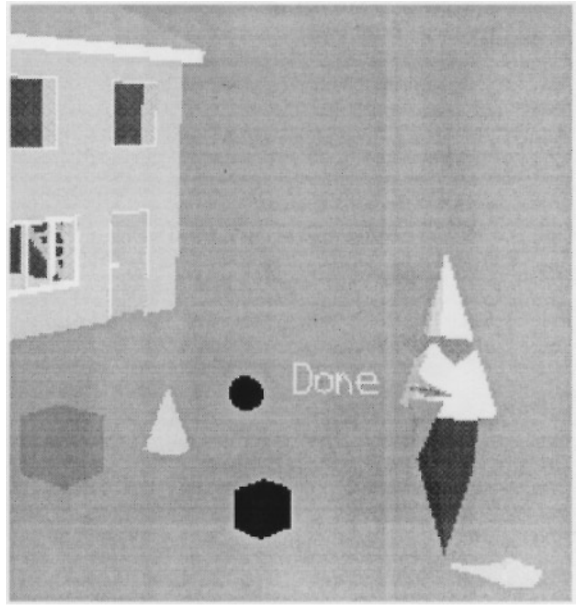
As an example, in Fig. 6, we present the implementation of a simple personal assistant agent in DIVE, controlled through a natural language interface (speech or text) [Karlgren, 95]. The agent takes natural language commands from its owner and attempts to interpret these through the application of various contextual devices (e.g. where is the current visual focus of the user and what objects has the user recently dealt with?). The agent is embodied as a small character attached to the user's visor. Engagement of the agent's attention is shown by the agent turning to the user and nodding and also giving some auditory feedback.

MASSIVE also contains some simple reactive objects which may be commanded through the text medium. An interesting feature of these objects is that their treatment of text input depends upon whether their attention has been engaged or not, which, in turn, depends on the users proximity and orientation to the agent. One example (deliberately chosen to demonstrate the notion of spatial engagement of attention) is a text recorder device. Text utterances directly targeted at the recorder are interpreted as commands (e.g. "play" and "record"), whereas other more peripherally overheard utterances are recorded or ignored depending on the current state of the recorder. The recorder shows that its attention has been engaged by a particular user by changing its appearance. Figure 7 shows an example of the MASSIVE recorder when its attention has been engaged.

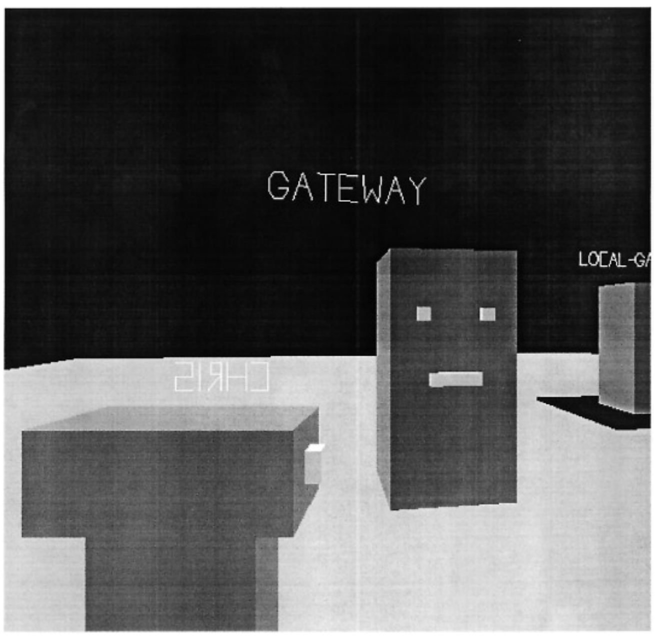
Another key aspect of embodying agents is representing their internal state or processing activity. For example, the DIVE agent above visually displays the choices of candidate objects it is considering as a list of text items above its head.



5



6



7

Fig. 5. Trails in dVS
Fig. 6. Talkative agent in DIVE
Fig. 7. The MASSIVE text recorder

6.2 Anthropomorphism

A more general issue for agent embodiment is that of anthropomorphism – attributing agents with a human appearance and behaviour. There have been several recent attempts at developing anthropomorphic agents including Fujitsu's *Charlottes* and Microsoft's *Bob*. A particularly interesting example is Takeuchi's work on situated facial displays which has constructed a computer-generated face driven by a video-image-processing system which attempts to engage the gaze of humans who are playing a card game with it or who are just passing by the computer console [Takeuchi, 95]. For a general overview of anthropomorphism in user interface design, see [Don, 92].

6.3 Uninhabited bodies as personal agents

As a final note in this section, we propose that the distinction between agents and human embodiments may not always be as clear as it at first seems. We have already discussed the problem of variable degree of presence in Sect. 4 above and proposed solutions such as the ability to put temporarily uninhabited embodiments into a sleeping position. One can also imagine that such a sleeping embodiment might act as a personal agent, capable of performing rudimentary communication functions on behalf of its absent owner. For example, it might acts a recorder for ongoing conversation or might be able to take messages. It might even be capable of locating its owner on the network and alerting them to the fact that someone else is trying to communicate with them.

7 Ethics and truthfulness

Our final issue relates to nearly all of those raised above. It concerns the degree of truth of a body image. In essence, should a body image represent a person as they are in the physical world or should it be created entirely at the whim or fancy of its owner? We should understand the consequences of both alternatives, or indeed of anything in between. Examples include: truth about identity (can people pretend to be other people?), truth about facial expression (imagine a world full of perfect poker players), and truth about capabilities (this body has ears on, can they hear me?) On the one hand, lying can be dangerous. On the other, constraining people to the brutal physical truth may be too limiting or boring. The solution may be to specify a *gradient* of body attributes that are increasingly difficult to modify. Those that are easy require relatively little resource. Those that are not require more. For example, changing virtual garments might be easy, whereas changing size or face of voice might be difficult. Truthfulness may also be situation dependent (i.e. different degrees may be required for different worlds, applications, contexts etc.). For example, simulation-type VR applications may require a very high level of truthfulness.

A particularly interesting aspect of truthfulness is the requirement (or not) for mutuality of embodiment as mentioned above. To draw another example from the realm of fiction, the crux of the recent book and film Disclosure [Crichton 93] hinges on the non-mutual embodiment of two different users of a database; one using a 3D information visualisation and the other using a traditional 2D interface.

To summarize so far, we have proposed a list of design issues that need to be considered by the designers of virtual bodies along with some possible techniques for addressing them. This list is summarized by Table 1. The following section now uses this list of issues as an analytical framework for understanding some other communication technologies.

8 Embodiment in other systems

Next, we briefly analyse the embodiments provided by four existing technologies, matching them up to the issues identified previously. The four technologies are: *dVS*, the commercial VR system from DIVISION; ATR's *Collaborative Workspace*; the multi-user VR game, *Doom*; and the general use of video as a communication medium. These specific examples have been chosen because of their diversity and because they highlight some interesting aspects of embodiment. Given more space, a wide range of other applications might also have been considered. Indeed, our intention is that designers of future collaborative applications could perform a similar exercise to the following and so gauge the likely effectiveness and limitations of their proposed body images for cooperative work. In order to save space, we only discuss those issues that are actually supported by the chosen examples.

Table 1. Issues relevant to the embodiment of users

Personal representation	Identity
	Background information
	Personalisation
Conveying activity	Presence
	Location
	Viewpoints and actionpoints
	Availability and degree of presence
	Gesture and facial expression
	Relationships with objects and information
	History of activity
Scalability and heterogeneity	Manipulating views of other
	Embodiment in multiple media
	Cross medium embodiment
	Mutuality
	Capabilities
	Efficiency
Embodying agents	Engagement of attention
	Anthropomorphism
	Uninhabited bodies as personal agents
Ethics	Truthfulness

8.1 dVS

dVS (version 2) from DIVISION Ltd., has been chosen as a typical example of current commercially available VR systems [Grimsdale, 91]. dVS supports multi-user VR applications running on both DIVISION's own hardware and on Silicon Graphics machines. Users may operate in either immersive or desktop modes. The default embodiment in dVS is a telepointer, although the authors have seen examples involving a disembodied head and a single limb. dVS addresses the following design issues:

- Presence and location – users are directly represented and the use of head- and hand-tracking support some notion of general location and orientation, although the lack of a body linking the two make this difficult to discern.
- Viewpoint and actionpoints – supported through head- and hand-tracking.
- Gesture – supported through the tracked hand only (though the representation of the hand as a pointer severely limits this ability).

8.2 Collaborative workspace

The ATR lab has been exploring the use of VR to support cooperative work for some years [Takemura, 92]. The main thrust of their research has been on supporting two-party teleconferencing and, in particular, on automatically capturing and reproducing facial expressions. Their collaborative workspace prototype achieves this by attaching a video camera to a head-mounted frame which also supports a position tracker. The use of small reflective discs attached to the user's face allows automatic analysis of their facial movements from the video image. This is then used to animate a texture-mapped model of the user's face. Collaborative workspace addresses the following issues:

- Presence – users are directly represented as humanoid-looking forms (as realistic as possible).
- Location – as far as we know, the user occupies a relatively fixed overall position (e.g. seated at a table).
- Identity – the aim is to make the user look as much like themselves as possible using a human head model onto which a photographic image of the user is textured and then animated.
- Viewpoint – the user's head position is tracked and represented, as are the positions of their eyes. Thus, this system is one of the very few to convey gaze direction at a very detailed level.
- Actionpoint – the user wears a single data glove and the position of one hand is therefore tracked.
- Gesture – supported through the tracked hand.
- Facial expression – this appears to be the primary focus of this work and a reasonably sophisticated range of facial expressions are possible through the use of tracked mouth, eyebrows and eyes. Both voluntary and involuntary expression are supported.
- Degree of presence – this is not really a problem due to the use of head-, eye- and hand-tracking.
- Efficiency – does not appear to be a key requirement of the project given the super-computers used.

Complimentary, and equally impressive, work on the capture and reproduction of facial expressions has been reported by Thalmann [Thalmann, 93]. In this case, the user is not constrained to wearing a head-mounted camera or any facial 'jewelry' or special make-up. The advantage of this is clearly a lack of intrusiveness. However, the disadvantage appears to be the inability to combine facial expressions with head-tracking.

8.3 Doom

Doom is a multi-user VR game for networked PCs. Doom has been chosen as a representative VR entertainment application intended for mass use and also because it supports many embodiment issues within very limited computing resources. Doom allows up to four users to navigate through a maze of corridors and rooms killing everything that they meet using a variety of weapons. The multi-user version can either be played in death-match mode (i.e. scoring points for killing each other) or, most interestingly, in cooperative mode (i.e. scoring points for killing other things together). Although this may seem far removed from a useful cooperative system, Doom contains several features worth noting. First, the graphics in Doom realise navigable texture-mapped environments on a 486 PC platform. In order to achieve this level of graphics performance, the designers of Doom have placed some constraints on their virtual worlds such as restricting them to use only perpendicular surfaces. Indeed, this is what makes the issue of embodiment in Doom particularly interesting; efficiency is of very great importance. Doom addresses the following design issues:

- Presence – users are directly represented as humanoids.
- Location – each user has a location and a limited number of orientations. Doom portrays users using flat 2D textures which are always perpendicular to the observer. Swapping between several such textures showing the user from different angles (North, South, East and West) conveys an approximate orientation.
- Identity – other users (player characters in gaming terminology) are distinguished from computer-generated monsters (non-player characters). Each user also wears a different colour tunic.
- Activity and availability – the activity of firing weapons is clearly shown.
- Viewpoint – only supported through rough orientation.
- Actionpoint – the impact point of weapons is shown, as is the trace of projectiles for some weapons.
- Facial expression – this is not visible in other people. However, the user does see a separate self-image which shows how healthy they are.
- Degree of presence – there is no mistaking a corpse.
- Time and change – not supported except for the user's self-image where improvements in health are portrayed.
- Truthfulness – people cannot alter their body images.
- Efficiency – this is where Doom excels; the whole system is an exercise in achieving maximum possible functionality with extremely limited resources.

8.4 Video

The use of video in collaborative applications is becoming increasingly widespread and makes an interesting contrast to the above VR based examples. As opposed to considering any specific video conferencing system, we focus on the nature of embodiment within video as a general medium.

- Presence – the presence of the person in front of the camera is clearly represented. However, in situations where there are one-way connections (e.g. media space "glances" or surveillance cameras), the presence of the person behind the camera may not be.
- Location – the physical location of a user may be shown to some degree. However, there is no real sense of a common location (i.e. you cannot place many people in relation to each other). The same is true of orientation. Other than knowing whether they are facing the camera or not, you cannot tell where someone is looking. First, if they are looking off camera, what are they looking at? Second, in groups of more than two people, who are they looking at if they peer into the camera?
- Identity – is conveyed nearly as well as in the real world (subject to picture resolution problems). Personalisation requires altering your physical self.
- Activity and availability – It may be possible to tell whether someone is busy or not but not what they are doing. Several researchers have investigated techniques for displaying availability to make a video connection (e.g. metaphors such as "doors").
- Viewpoint – not really supported, although you might be fooled otherwise (the orientation issue from above).
- Gesture – supported as in the real world subject to field-of-view constraints.
- Facial expression – obviously supported (both voluntary and involuntary).
- Truthfulness – generally enforces the brutal truth as there is little chance to break away from the real person's

appearance. Some more advanced systems may allow some manipulation of video images.

9 Summary

The premise of this paper has been that user embodiment is a key issue for CVEs (and indeed, for other kinds of collaborative system). Given this assumption, we have identified an initial list of issues as being relevant to the embodiment of users.

We have also shown how these issues are currently reflected by our own DIVE and MASSIVE CVEs as well as several others.

We suspect that the importance of any given design issue will be both application- and user-specific and that the art of virtual body building will involve identifying the important issues in each case and supporting them within the available computing resource. However, at the present time, our list remains only an initial framework for the discussion and exploration of embodiment. In our future work we aim to realise a larger number of these issues within our own DIVE and MASSIVE systems, gaining deeper insights into their relative importance and possible implementation. In the longer term, we would hope to refine our list into a complete 'body builder's work-out', supporting the choice and analysis of the most appropriate designs for the available equipment, application, users, scale and longevity of intended collaborative applications.

Acknowledgement. This work has been sponsored by the CEC through the COMIC ESPRIT Basic Research Action and by the UK's Engineering and Physical Science Research Council (EPSRC) through the Virtuosi project and its Ph.D. studentship programme.

References

[Badler, 93] Badler NI, Phillips CB, Webber BL (1993) Simulating humans – computer graphics animation and control, Oxford University Press, Oxford

[Benford, 95] Benford S, Bowers J, Fahlén LE, Greenhalgh C, Mariani J, Rodden T (1987) Networked virtual reality and co-operative work, Presence – Teleoperators Virtual Environ 4:364–386

[Brand, 87] Brand S (1987) The medialab – inventing the future at MIT. Viking Penguin

[Bricken, 94] Bricken W, Coco G (1994) The VEOS project, Presence – Teleoperators Virtual Environ 3:111–129

[Carlsson, 93] Carlsson C, Hagsand O (1993) DIVE – A platform for multi-user virtual environments. Comput Graphics 17

[Crichton, 93] Crichton M (1993) Disclosure. Ballantine Books

[Don, 92] Don A, Brennan S, Laurel B, Schneiderman B (1992) Anthropomorphism: from Eliza to Terminator 2. In: Proc. CHI'92 Human Factors in Computing Systems, Monterey, May 3–7, 1992, ACM Press, pp 67–70

[Fahlén, 93] Fahlén LE, Brown CG, Stahl O, Carlsson C, (1993) A space based model for user interaction in shared synthetic environments. In: Proc. InterCHI'93, ACM Press, pp 43–48

[Greenhalgh, 95] Greenhalgh C, Benford S (1995) MASSIVE: A virtual reality system for tele-conferencing. ACM Trans Comp Human Interfaces (TOCHI) 2:239–261

[Grimsdale, 91] Grimsdale C (1993) Supervision – a parallel architecture for virtual reality. In: Earnshaw RA, Gigante MA, Jones H (eds.) Virtual Reality Systems. Academic Press, New York, pp 29–37

[Karlgren, 95] Karlgren J, Bretan I, Frost N, Jonsson L (1995) Interaction models, reference, and interactivity for speech interfaces to virtual environments. In: Proceedings of Second Eurographics Workshop on Virtual Environments – Realism and Real Time

[Ohya, 93] Ohya J, Kitamura Y, Takemura H, Kishino F, Terashima N (1993) Real-time reproduction of 3D human images in virtual space teleconferencing. In: Proc. VRAIS'93, IEEE, Seattle Washington September, 1993, pp 408–414

[Quéau, 94] Quéau P (1994) Real-time facial analysis and image rendering for televirtuality applications. In: Loeffler CE, Søby M, Ødegård O (eds.) Notes from Virtual Reality Oslo '94 – Networks and Applications

[Shu, 94] Shu L, Flowers W (1994) Teledesign: groupware user experiments in three-dimensional computer-aided design, Collaborative Comput 1:1–14

[Slater, 93] Slater M, Usoh M. (1993) Representations systems, perceptual position and presence in immersive virtual environments. Presence 2:221–233

[Stephenson, 92] Stephenson N (1992) Snowcrash, ROC Publishing

[Takemura, 92] Takemura H, Kishino K. (1992) Cooperative work environment using virtual workspace. In: Proc. CSCW'92, Toronto, Nov 1992, ACM Press

[Takeuchi, 95] Takeuchi A (1995) Situated facial displays: towards social interaction. In: Proc. ACM Conference on Computer Human Interaction (CHI'95), Demver, Colo., May 1995, ACM Press

[Thalmann, 93] Thalmann D (1993), Using virtual reality techniques in the animation process. In: Earnshaw RA, Gigante MA, Jones H (eds.) Virtual Reality Systems, Academic Press, New York, pp 143–159

[Vinge, 92] Vinge V (1992) A fire upon the deep. Millennium

STEVE BENFORD is a senior lecturer in computer science at the University of Nottingham. His research interests span virtual reality, distributed systems and cooperative work. Previously he has worked in areas such as directory services and group communication. His current interests centre around the idea of collaborative virtual environments; multi-user, distributed virtual reality systems that provide explicit support for cooperative work. As part of his teaching activities, he has also been involved in the development of a system called Ceilidh pronounced (kay-ley) for automatically marking and administrating students programming work.

DR. JOHN BOWERS works at the Department of Psychology, University of Manchester. He has been a visiting researcher at Lancaster University, the Swedish Institute for Computer Science and Royal Institute of Technology, Stockholm. He researches work and human communication using qualitative methods (ethnography and interaction analysis), in particular focusing on the design and usage of computing technology (especially advanced co-operative systems and virtual reality). His overall aim is to use empirical study to reformulate foundational issues in cognitive science, computer science, and organizational theory. Dr. Bowers was a founding editor of the journal 'Computer-Supported Cooperative Work'. He also writes computer music and has had compositions performed on several concerts in Europe, most recently as part of 'Membrane', a series of multimedia performances combining dance, music and interactive virtual reality.

LENNART E. FAHLEN. Manager of the Distributed Collaborative Environments Group, DCE, at the Swedish Institute of Computer Science, SICS. The DCE group's focus is on research and development of tools and techniques for distributed virtual environments and it is the main developer and maintainer of the DIVE virtual reality system.

CHRIS GREENHALGH is a lecturer in the Department of Computer Science at the University of Nottingham, UK. He is also completing a PhD on large-scale collaborative virtual environments for which he built the MASSIVE virtual reality teleconferencing system. This has been used to hold international research meetings and is used on an ongoing basis for meetings between sites across the UK. He is currently developing a second version of MASSIVE to support 100+ simultaneous users. His main interests at present are virtual reality, distributed systems and computer-supported cooperative work.

DAVE SNOWDOWN. In 1991, whilst completing his M.Sc on an object-oriented system for representing and typesetting music, David Snowdown became interested in virtual reality and co-founded the "Advanced Interfaces Group" at the University of Manchester. At this time he also started

research into the design of multi-user VR systems with the goal of obtaining an PhD. As part of his PhD he designed and implemented a distributed, object-oriented, multi-user, VR system called AVIARY. In January 1994, David Snowdown moved to Nottigham University to work on the DTI/EPSRC-funded VirtuOsi project. In addition to his work on the design of VR systems, he also started investigating the visualization of abstract information and the issues concerning user embodiment in virtual environments.