User Embodiment in Collaborative Virtual Environments

Steve Benford
Department of Computer Science
The University of Nottingham, Nottingham, UK
Tel: 44-602-514203
E-mail: sdb@cs.nott.ac.uk

Chris Greenhalgh
Department of Computer Science
The University of Nottingham, Nottingham, UK
Tel: 44-602-514225
E-mail: cmg@cs.nott.ac.uk

John Bowers
Department of Psychology
The University of Manchester, Manchester, UK
Tel: 44-61-275-2599
E-mail: bowers@hera.pych.man.ac.uk

Dave Snowdon
Department of Computer Science
The University of Nottingham, Nottingham, UK
Tel: 44-602-514225
E-mail: dns@cs.nott.ac.uk

Lennart E. Fahlén
The Swedish Institute of Computer Science
Stockholm, Sweden
Tel: 46-8-752-1539
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 co-operative 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 including: presence, location, identity, activity, availability, history of activity, viewpoint, actionpoint, gesture, facial expression, voluntary versus involuntary expression, degree of presence, reflecting capabilities, physical properties, active bodies, time and change, manipulating your view of others, representation across multiple media, autonomous and distributed body parts, truthfulness and efficiency. Following this, we show how these issues are reflected in our own DIVE and MASSIVE prototype systems and also show how they can be used to analyse several other existing collaborative systems.

KEYWORDS: virtual reality, CSCW, embodiment

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 presents an early theoretical exploration of this issue based on our experience of constructing and analysing a variety of collaborative virtual environments: multi-user virtual reality systems which support co-operative work.

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 number of gestural sign languages or may provide an important accompaniment to other forms of communication, helping co-ordinate 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, the sense of user immersion, 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 co-operative systems in general, where embodiment often seems to be a neglected issue (it appears that many collaborative systems still view users as people on the outside looking in). 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 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.

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 listed in section two and constitute a diverse, and occasionally conflicting, set of requirements. Designing an appropriate body image will most likely be a case of maintaining a sensible balance between them. 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, in section three we describe how the issues are currently reflected in two of our own collaborative virtual environments, DIVE and MASSIVE, giving examples of the bodies we have constructed so far. Section four then uses our list as a framework for analysing how a variety of other collaborative virtual environments and more general CSCW systems tackle user embodiment.

DESIGN ISSUES AND TECHNIQUES
In this section we identify a list of design issues for user embodiments as well as possible techniques for dealing with them. As indicated above, we approach these issues from the perspective of collaborative virtual environments, although we encourage the reader to consider their application to other kinds of collaborative system. We begin with the fundamental issues of presence, location and identity.

Presence
The primary goal of a body image is 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 graphics objects which are considered to represent them.

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. co-ordinate 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.

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 don't 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). Fourth, you might be able to find out who someone is from their body image. 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?).

Allowing users to personalise body images is also likely to be important if collaborative virtual environments 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 in turn. An example of personalisation might be the ability to don virtual garments or jewellery. 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).

Activity, viewpoints and actionpoints
Body images might convey a sense of on-going 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 co-ordinating activity and in encouraging peripheral awareness of the activities of others. We identify two further aspects of conveying activity: representing user's viewpoints and representing their 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 three dimensional design has already shown the importance of conveying users' viewpoints [7]. 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 synchronised; 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. Later on we shall see systems that show general body position, head position or even eye position depending on the power of the tracking facilities in use.

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 (e.g., the mind behind the body may have popped out of the office for a few seconds). 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. This mind/body separation could 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.

As a concrete example of this issue, we cite some of our early experiences with the DIVE system (see below). One of the interesting aspects of DIVE is that a user process that exits unexpectedly often leaves behind a 'corpse' (an empty graphics embodiment). A long DIVE session may produce several such corpses (particularly when developing and testing new applications), which can cause confusion. As a result, two informal conventions have been established among DIVE users. First, 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. Second, bodies that turn out to be corpses are 'buried' (i.e. moved) below the ground plane. It would be useful to have some more graceful mechanisms for dealing with this problem!

Gesture and facial expression

Gesture is an important part of conversation and ranges from almost sub-conscious 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 [2]). Another approach involves capturing expression information from the human face using an array of sensors on the skin, modelling 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 [6] and also the work of Thalmann [8] and Quéau [9]).

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.

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. An example might be carving out trails and pathways through virtual space in much the same way as they are worn into the physical world.

Manipulating one's view of other people

In heterogeneous systems where users might employ equipment with radically different capabilities (see MASSIVE below), it will be important for the 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 resource on rendering my body when it could better be used to render other objects. 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 have some control over 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 [10]. The ability for people to adopt subjective world views (e.g., seeing different representations of an embodiment) represents a significant challenge to current VR architectures.

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.).

Autonomous and distributed body parts

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 (e.g. logging on more than once) or though some kind of computer agent acting on their behalf (e.g. issuing a database query while browsing an information visualisation).

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. More specifically, we suspect that approaches which attempt to reproduce the human physical form in as full detail as possible may in fact be wasteful and that more abstract approaches which reflect the above issues in simple ways may be more appropriate (unless it turns out that users cannot relate to abstract bodies). 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 prioritising the above issues in any given communication scenario. In fact, the real challenge with embodiment will be to prioritise the issues listed in this section according to specific user and application needs and then to find ways of supporting them within a limited computing resource.

Truthfulness

This 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 or 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?) One 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.

In summary, 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. The following section now describes how some of these issues have been dealt with in our own DIVE and MASSIVE prototype collaborative virtual environments.

EMBODIMENT IN DIVE AND MASSIVE

The authors have been involved in the construction of two general collaborative virtual environments, DIVE at the Swedish Institute of Computer Science, and MASSIVE at the University of Nottingham. This section considers how the above design issues are reflected in these systems.

Embodiment in 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 [5]. 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.

A variety of embodiments have been implemented within the DIVE system. The simplest are the 'blockies' which are composed from a few basic graphics objects. The general shape of blockies is sufficient to convey presence, location and orientation (the most common example being a letter 'T' shape). In terms of identity, simple static cartoon-like facial features suggest that a blockie represents a human and the ability for people to personalise their own body images supports some differentiation between individuals (DIVE provides a general geometry description language with which users may specify their own body shapes if they wish). A more advanced DIVE body for immersive use texture maps a static photograph onto the face of the body, thus providing greater support for identifying users in larger scale communication scenarios. This body also provides a graphic representation of the user's arm which tracks their hand position in the physical world via a 3-D mouse.

The display of a solid white line extending from a DIVE body to the point of manipulation in space represents actionpoint in a simple and powerful way and enables other users to see what actions a user is engaged in (e.g., selecting objects). In various DIVE data visualisation applications, each user may also be associated with a different colour which is used to show which data they are accessing (selected objects change to this colour), thereby providing limited peripheral awareness of their activity.

Immersive blockies also support a moving head which tracks the position of the user's head in the real world via their head-mounted display (i.e. a six degrees of freedom sensor attached to the top of the user's head). This is very effective at conveying viewpoint, general activity and degree of presence. Finally, video conferencing participants can be represented in DIVE through a video window.

Figure 1 shows a DIVE conference scenario involving a range of embodiments. From left to right we see: an immersed user with humanoid body, textured face and tracked head and arm; a simple non-immersive blockie sporting a humorous propeller hat; a video conferencing participant; and a second immersive user. The scene also shows some DIVE collaboration support tools: a functioning whiteboard which can also be used to create documents and a conference table for document distribution.

Embodiment in 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 [1]. The main goals of MASSIVE are scale (i.e. supporting as many simultaneous users as possible) and heterogeneity (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).

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 3-D space and allows users to navigate this space with a full six degrees of freedom. The audio interface allows users to hear objects and supports both real-time conversation and playback of pre-programmed sounds. The text interface provides a MUD (Multi-User Dungeon)-like view of the world via a window (or map) which looks down onto a 2-D plane across which users move. Text users are embodied using a few text characters and may interact by typing messages to one another or by 'emoting' (e.g. smile, grimace, etc.).

The graphics, text and audio interfaces may be arbitrarily combined according to the capabilities of a user's terminal equipment. Furthermore, users may export an embodiment into a medium that they cannot receive themselves (thus, a text user can be made visible in the graphics medium and vice versa). The net effect is that users of radically different equipment may interact, albeit in a limited way, within a common virtual world (e.g.

text users may appear as slow-speaking, slow moving flatlanders to graphics users). For example, at one extreme, the user of a sophisticated graphics workstation may simultaneously run the graphics, audio and text clients (the latter providing a map facility and allowing interaction with non-audio users). At the other, 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 and so on. One effect of this heterogeneity is to allow us to populate MASSIVE with large numbers of users at relatively low cost.

MASSIVE graphics embodiments are based on DIVE blockies (although, as with DIVE, users can specify their own geometry via a simple modelling language). Blockies are also automatically labelled with the name of their owner so as to aid identification. In the text interface, users are embodied by a single character (typically the first letter of their chosen name) which shows position and may help identify users in a limited way. An additional line (single character) points in the direction the user is currently facing. Thus, using only two characters, the MASSIVE text interface attempts to convey presence, location, orientation and identity.

Given MASSIVE's inherent heterogeneity, its embodiments need to convey users' capabilities to one another. For example, considering the graphics interface, an audio capable user has ears; a desk-top graphics user (monoscopic) has a single eye; an immersed stereo user would have two eyes and a text user ('textie') has the letter 'T' embossed on their head. Thus, on meeting another user, it should be possible to quickly work out how they perceive you and through which media you can communicate with them (e.g., should you use the audio channel or send text messages?).

Figure 2 shows an example of the graphics interface showing a conference involving five users (we are one of them). We see two non-immersed, audio capable users facing each other across the conference table (ears and a single eye) and a text-only user facing diagonally towards us. We can also see that another non-audio capable user has their back to us.

4. EMBODIMENT IN OTHER SYSTEMS

Next, we briefly analyse the embodiments provided by four further 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 co-operative work. In order to save space, we only discuss those issues that are actually supported by the chosen examples.
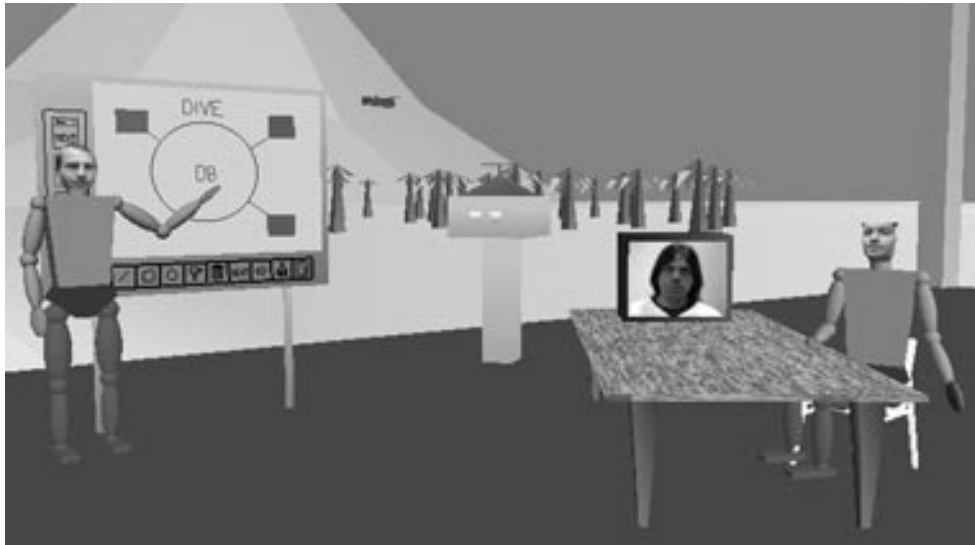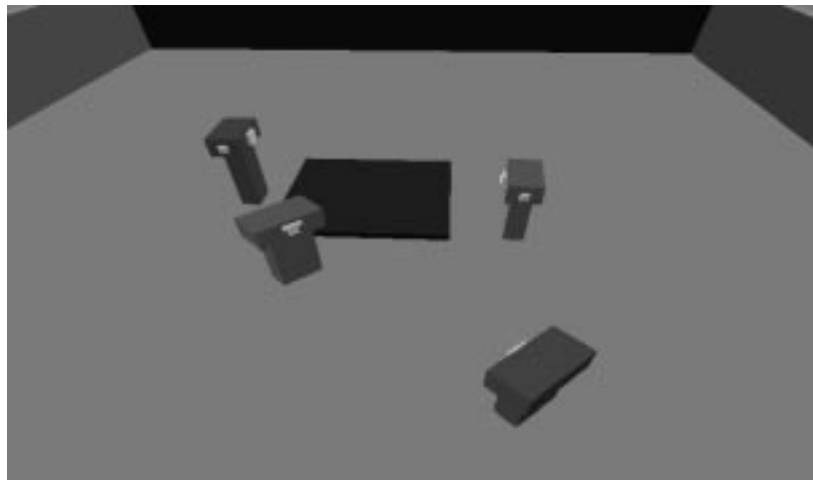
Figure 1: Various embodiments attend a DIVE conference

Figure 2: Users show their capabilities at a MASSIVE conference

dVS

dVS, from DIVISION Ltd, has been chosen as a typical example of current commercially available VR systems [3]. dVS supports multi-user virtual reality 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 hard to discern at times.
- Viewpoint and actionpoints - supported through head and hand tracking.

- Gesture - supported through the tracked hand only (and the representation of the hand as a pointer severely limits this ability).

Collaborative Workspace

The ATR lab have been exploring the use of virtual reality to support co-operative work for some years [6]. 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 disks attached to the users 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 [8]. In this case, the user is not constrained to wearing a head-mounted camera or any facial 'jewellery' 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.

Doom
Doom is a multi-user virtual reality 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 to kill each other) or, most interestingly, in co-operative mode (i.e. scoring points for killing other things together). Although this may seem far removed from a useful co-operative system, Doom contains several features worth noting. First, the graphics in Doom realise navigable texture mapped environments on a 486 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 2-D 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 users 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.

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 can't 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.

SUMMARY

The premise of this paper has been that user embodiment is a key issue for collaborative virtual environments (and indeed, for other kinds of collaborative system). Given this assumption, we have identified the following initial list of issues as being relevant to the embodiment of users: presence, location, identity, activity, availability, history of activity, viewpoint, actionpoint, gesture, facial expression, voluntary versus involuntary expression, degree of presence, capabilities, physical properties, manipulating one's view of others, multiple media, distributed bodies, truthfulness and efficiency. We have also shown how these issues are currently reflected in our own DIVE and MASSIVE collaborative virtual environments 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.

REFERENCES

1. Benford, S., Bowers, J., Fahlén, L. E., and Greenhalgh, C., Managing Mutual Awareness in Collaborative Virtual Environments, *Proc. Virtual Reality Systems and Technology (VRST) '94*, August, 1994, Singapore.

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

3. Grimsdale, C., Supervision - A Parallel Architecture for Virtual Reality, in *Virtual Reality Systems*, Earnshaw, R.A., Gigante, M.A and Jones, H. (eds), Academic Press, 1993, ISBN 0-12-227748-1.

4. Haruo Takemure and Fumio Kishino, *Cooperative Work Environment Using Virtual Workspace*, In Proc. CSCW'92, Toronto, Nov 1992, ACM Press.

5. Lennart E. Fahlén, Charles Grant Brown, Olov Stahl, Christer Carlsson, *A Space Based Model for User Interaction in Shared Synthetic Environments*, in Proc.InterCHI'93, ACM Press, 1993.

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

7. Shu, L., and Flowers, W., Teledesign: groupware user experiments in three-dimensional computer-aided design, Collaborative Computing, 1(1), Chapman & Hall, 94.

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

9. Quéau, Philippe, Real Time Facial Analysis and Image Rendering for Televirtuality Applications, in Notes from Virtual Reality Oslo '94 — Networks and Applications, eds. Loeffler, Carl E and Søby, Morten and Ødegård, Ola, August 1994.

10. Bricken, William AND Coco, Geoffrey, The VEOS Project, in Presence — Teleoperators and Virtual Environments, Vol. 3, No 2, MIT Press 1994, pp 111--129.

Add mention of jack stuff and reference

Add ref to HCI paper for viewpoints and actionpoints

Suspended bodies -

  suspended - change appearance, monitor conversation, record, wake up if spoken to

  wake up - change apearnace, play back conversation if required

Extended trails in dVS

  identity

fade from view
Label with 'authenticated' tab