# Supporting Awareness and Interaction through Collaborative Virtual Interfaces

*Mike Fraser, Steve Benford*

Communications Research Group
School of Computer Science & I.T.
University of Nottingham
University Park
Nottingham NG7 2RD, U.K.
+44 115 951 4225(/4203)
{mcf, sdb}@cs.nott.ac.uk

*Jon Hindmarsh, Christian Heath*

Work, Interaction and Technology Research Group
The Management Centre
Franklin-Wilkins Building
King's College London
London SE1 8WA, U.K.
+44 20 7848 4194(/4496)
{jon.hindmarsh, christian.heath}@kcl.ac.uk

**ABSTRACT**

This paper explores interfaces to virtual environments supporting multiple users. An interface to an environment allowing interaction with virtual artefacts is constructed, drawing on previous proposals for 'desktop' virtual environments. These include the use of Peripheral Lenses to support peripheral awareness in collaboration; and extending the ways in which users' actions are represented for each other. Through a qualitative analysis of a design task, the effect of the proposals is outlined. Observations indicate that, whilst these designs go some way to re-constructing physical co-presence in terms of awareness and interaction through the environment, some issues remain. Notably, peripheral distortion in supporting awareness may cause problematic interactions with and through the virtual world; and extended representations of actions may still allow problems in re-assembling the composition of others' actions. We discuss the potential for: designing representations for distorted peripheral perception; and explicitly displaying the course of action in object-focused interaction.

**KEYWORDS**: Collaborative Virtual Environments, User Presentation, Peripheral Lenses, Action Representation

**INTRODUCTION**

Whilst there has been wide-ranging research in the utility of interfaces to virtual environments in single-user applications, less attention has been paid to their design in scenarios involving multiple users – so called Collaborative Virtual Environments (CVEs). Performing collaborative tasks through distributed, synchronous computing technologies renders the interface to that technology critical, in some ways more than for conventional 'stand-alone' applications. In addition to performing actions and receiving feedback through the interface, users require information about other users' actions and activities [2].

Applications supporting multiple participants, such as groupware systems, have often used the notions of strict or relaxed WYSIWIS [15] to enable multiple users to perceive a common interface with others. These approaches are intended to effectively harmonise the representation of action between all users. However, adopting this policy for a CVE interface may direct away from some of the anticipated strengths of CVEs – namely, the ability to convey a sense of individual perspective on the environment. CVEs would appear to better support the kinds of 'quasi-corporeal' interaction, which exact an independent view on the shared application. Interaction policies such as this have been implemented in other types of collaborative system, notably the Kansas system, which describes that "What You See Is What I Think You See" [14].

However, our own work has shown that participants find this concept difficult to sustain in interacting through CVEs – re-assembling what another can perceive may be hampered by the interface onto the virtual environment [12]. Indeed, problems occur not only with 'seeing what the other is seeing', but more generally with *seeing what another is doing*, whether that activity be speaking, moving or gesturing. This earlier work consisted of an experimental task in which the layout of furniture in a virtual room was designed collaboratively. Our initial goal was to inspect the support CVEs offered with


*UIST '99. Asheville, NC*

respect to 'object-focused interaction', which is often critical to everyday work [10]. The experiment was built on studies of the ways in which people collaborate with artefacts and features of the world in both real environments and through video-based technologies [11, 7]. Observational analysis of video data from the trials revealed a number of issues related to collaboration 'through' our system. Some more detail is provided later, but in summary our findings were:

- The view of an object under discussion was often 'fragmented' or separated from the image of the others' embodiment. As a result users found it difficult to make sense of talk and activity without firstly seeing (and thus seeking) the other's embodiment in relation to relevant objects. These problems developed primarily due to the limited horizontal field of view and the lack of rapid gaze direction provided by the CVE;
- Participants compensated for this 'fragmentation' of the space by using talk to make explicit actions and visual conduct that are recurrently implicit in co-present interaction. In particular, referencing became a topic in and of itself, rather than being subsumed within more general actions and activities;
- Participants faced problems assessing and monitoring the perspectives, orientations and activities of the other(s) even when the other's embodiment was in view. For instance, they had particular trouble assessing what should be available on the other's screen even when they could see the other's avatar and its position and general orientation in the world.

Design possibilities for virtual environment interfaces were outlined, including modifying and extending the ways in which actions may be represented as a resource for display and perception in CVEs. Therefore, it seemed prudent to pursue, implement and evaluate interfaces and representations to attend to these problems.

## INTERFACE DESIGN

In this section, we identify and describe an interface based upon previous proposals for 'desktop' virtual environments: the use of Peripheral Lenses [13]; and the extended representation of activity in virtual environments [12]. Our intention is to combine these two designs in order to improve the visibility of a user's actions for their colleagues within the virtual environment. To this end, the interface and embodiment of the existing MASSIVE-2 CVE system [1] have been revised in the two ways described in the following sections.

### Peripheral Lenses

Topics involving the awareness of others' actions have received much attention within the study of collaboration through computing applications. Technologies designed to support remote synchronous interaction, such as media spaces, groupware systems, and virtual environments have been criticised for their inability to support awareness and monitoring of others' activities [7, 9, 12]. In [12], the ways in which a CVE can impede users' ability to make actions available for others is described. A key reason for these difficulties derives from the limited field of view provided by desktop CVEs [see also 8]. Head-mounted interfaces may provide more rapid and intuitive viewpoint movement but often suffer from limited fields-of-view through which similar problems have been observed [6]. Designers of interfaces to CVEs usually provide a field-of-view on the virtual environment at around a third of the human perceptual capability (~50–60 degrees). The reason involves potentially debilitating perspective distortions across the entire field-of-view that occur when rendering wide viewpoints on the limited space of a desktop display. For example, actions involving precise movement such as navigation and object manipulation may prove impossible with severely distorted views. High levels of distortion might also hinder Virtual Reality applications, such as medicine, architecture or the military, where visualising realistic physical situations or processes is essential.

Therefore, we aim to consider an alternative that may provide some enhanced support for awareness as compared with the previous interface. We intend to extend the field of view, whilst maintaining the detail of the focal view.

Peripheral Lenses [13] consist of two windows, which render views on a virtual environment to the left and right of the main view, with increased distortion allowing more visual information to be displayed within a smaller horizontal space. This interface technique was primarily introduced as a navigation aid, the focus of the proposal being concerned with providing users with peripheral vision to acquire geographical knowledge of the virtual space. Quantitative analysis of single user desktop virtual environments utilised search times as a measure of the technique. Although this particular aspect of the study proved inconclusive, Robertson et al. note that "Further studies are needed to understand exactly when Peripheral Lenses are effective" [13].

In pursuing this design, we hope to determine if this effectiveness might be in viewing and attending to other's actions within a collaborative environment. In this way, the use of peripheral lenses might alleviate the kinds of problematic interaction noted in our previous work and provide further support for awareness in CVEs.

Indeed, whilst a different area of use is targeted – a multi-user co-operative task compared with a single-user search task – similar effects are intended: to increase a user's effective field-of-view; and to provide distorted view of peripheral scenes.

Figure 1 displays part of the virtual world interface used in evaluation. As with the earlier system, this realisation of peripheral lenses avoids perspective distortions in the main field-of-view, conversely using them to provide peripheral vision at the sides of the desktop display.

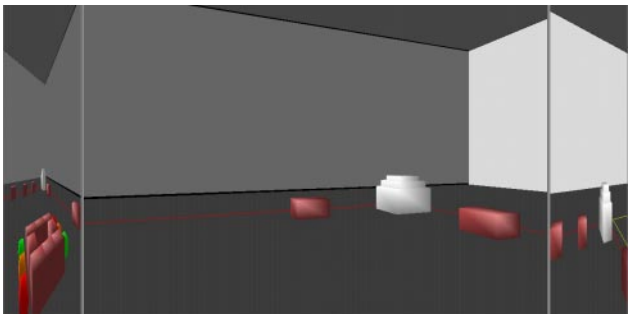

**Figure 1 - Peripheral lenses in MASSIVE-2**

This implementation of the concept differs from that of Robertson et al. in two ways:

- Users are unable to alter the horizontal lens angle, which is fixed at 60 degrees for each lens, effectively providing a constant viewport of 180 degrees;
- The ability to focus on either peripheral lens is provided through a technique which we coin 'peripheral glancing'. The interface allows users to momentarily swap distortion between the middle and peripheral views to visually attend their left or right.

When designing peripheral lenses for monitoring others' actions, we anticipate difficulties in assessing actions depicted in the distorted views. The glancing facility is provided to try and enable users to resolve these problems. 'Peripheral glancing' begins on depressing a button placed below the relevant lens. The distortion is immediately removed from the peripheral lens and added to the central view, and releasing the button reverses the process. An example of a user 'glancing' to the left (on the same scene as in Figure 1) is shown in Figure 2:

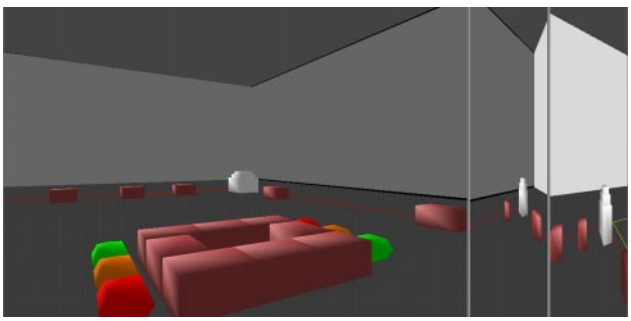

**Figure 2 - Peripheral glancing to the left**

It is our belief that the metaphor of glancing will give users a useful and intuitive sense of the interface and its affordances.

#### Representing Orientation and Conduct

When performing actions alongside objects within the CVE, certain features might render the task of perceiving the 'targets' of those actions problematic. These include: limited or potentially distorted perspectives; large virtual distances between participants; slow or simplistic gaze directing strategies; and lack of stereoscopic vision. Therefore, the CVE may hinder a user's ability to connect the talk and gestures of another to a particular object. As a result it may be difficult for them to recover the sense of the action.

Also, the use of pseudo-humanoid embodiments within a CVE may cause other participants to assume that a user has human-like capabilities in perceiving the virtual environment [12]. An avatar is effectively, in a CVE, the representation to others of the user's interface. It is the only means through which another's interface activity can be displayed and perceived. Avatars in CVEs are often designed using a realistic humanoid or pseudo-humanoid embodiment that aims to mirror the conduct of a real human. This may cause other participants to expect the associated user to perceive similarly to a human; yet the concepts of field-of-view and rapid movement/gaze direction may significantly differ between real and virtual environments. Thus problems arise in assessing what another can see. Given this problem, it also becomes difficult for a user to design actions to be seen by the other and to ensure that they can be seen.

In order to address these problems, we intend to exaggerate the visibility of actions relating avatars and objects, and embed the orientation and conduct of a user more firmly in the virtual world. Designing representations of participants in this manner is in some ways in direct conflict to the usual approaches employed in creating avatars in virtual environments. Using exaggerated representations might be said to attach more importance to the practical use of the representation, rather than its intended 'realism'.

In considering the ways in which we might construct an interface to a CVE, we therefore propose an approach whereby each potential interface action is shown not only to the user through the desktop display, but also to co-present users through the avatar representation and elsewhere. For the purposes of our analysis, key interface actions were represented through the environment and on the targets of those actions, as well as through the traditional approach of presenting changes on the avatar itself. It was decided to design an interface to, and avatar within, the virtual world which provided the following

actions: speaking; moving; field of view and peripheral glancing; pointing; and grasping/moving objects. However, this section will concentrate on the representation of the actions seemingly most relevant to the context of the task performed in trials: those of field of view, grasping and pointing.

*Visualising the Field of View.* The avatar can be seen to face a particular direction in the world. Therefore, one might assume that users could discern what another can see. However, there are reasons why we might wish to represent the field of view more explicitly to co-participants. Firstly, although the users are provided with a large field of view, the majority of this field is grossly distorted and unavailable to detailed inspection. Therefore, the general orientation of the avatar does not accurately display what is readily available in focus and what is distorted. Secondly, although the CVE interface provides a peripheral glancing capability, it may be unclear to another when this is actually being used and when it is not. An explicit representation of the field of view should highlight what a colleague, under any distortive conditions, can see. Therefore a representation of the field of view should show which lenses are currently distorted and undistorted, and thus not only what areas other participants see as distortions, but also which part of the field-of-view the user is currently attending using the 'peripheral glancing' capability.

In determining how to explicitly represent another's view to a user within the virtual environment, we refer back to [12], which states that visual representation could be achieved "by making the view frustrum visible as a wire-frame; or by highlighting viewed objects in some way (perhaps using lighting or shadows)." It was decided to explicitly outline the view frustrum using a wire-frame to affirm the view of the other, because lighting is often used in virtual environments for cosmetic or realism reasons. Lighting intended to provide reciprocal perspective might prove confusing or, at worst, indistinguishable from other lights. Whilst lighting can be identified by colouring, providing multiple participants with distinguishing representations might actually compound the complexity of interaction – it may be easier to identify overlapping lines of different colours than to distinguish overlapping volumes of light of an equivalent hue. Additionally, the graphical complexity required to render shadowing within the virtual environment means that changes in the viewpoint of the other would significantly slow the frame update rate of a user's view.

An example of a participant's representation of field-of-view within this implementation is shown in Figure 3. The dark lines bound the edges of the distorted lenses, and bright lines bound the undistorted view, changing position in the relevant direction if a user 'glances' to the left or right. As co-participants on the scene, it should be possible for us to identify which objects can be seen in both distorted and undistorted conditions.

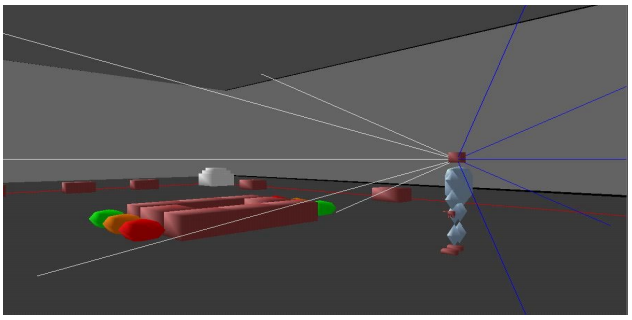

**Figure 3 - Explicit representation of field-of-view**

*Presentation of Grasping and Pointing.* The difficulties of recovering the sense of an action due to the 'fragmentation' of the space were noted in [12]. In particular users found to difficult to find a referent, for example, when they did not have the other's embodiment and the object both in view. The development of the interface presented should enable users to make sense of an action even when they can only see the object itself, or indeed, just environment between object and avatar. In our implementation, both embodiment 'arms' are extended to touch the artefact when portraying the grasping/moving of that object. As shown in Figure 4, the object in question is also 'wire-framed' in order to show its current state of 'being moved'. A mouse is used to select the correct artefact, and then manipulate its position 'on the ground'.

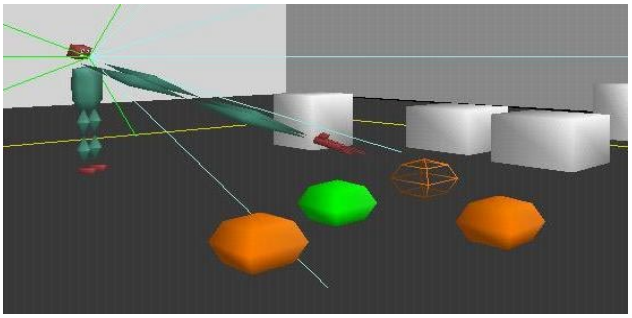

**Figure 4 - Explicit representation of grasping**

In contrast, only one arm is extended to point at artefacts or regions of the environment. In this way observers of the connecting environment between avatar and artefact may differentiate these actions, even if the target of the action itself is visually unavailable. There is no representation on the target of a pointing gesture, as this target is often indefinable – a region of space has a variable and perceived boundary, and cannot be pre-identified by the CVE application and therefore explicitly represented when 'pointed to'. However, extending and retracting the gesture is made possible by stretching the virtual 'arm' using the mouse, thus allowing a pointing gesture to focus on any feature of the virtual world.

## TRIALS

In studying the use of the interface, pairs of subjects were asked to perform a co-operative task through the revised interface described above. Users were asked to collaboratively *duplicate* a virtual world layout. An environment was provided with an architecturally abstract, yet symmetrical, structure, and a second area containing all the necessary 'building blocks' in an unstructured form. Participants were asked to re-organise the latter half of the world to conform to the same pattern and symmetry as the former. The world consisted of blocks and spheres of different shapes, sizes and colours. Many of the artefacts provided were only dissimilar in one of these aspects for two reasons: to assess problems that might occur when viewing similar objects through distorted viewpoints; and to encourage participants to actively use the referencing aspects of the interface to help the other discriminate between artefacts.

Communication in the audio medium was performed through a microphone and headphones. Each user sat in a separate room of the same physical location at a Silicon Graphics workstation, connected via a local ethernet. A mouse was used to carry out pointing, grasping, and navigation, by toggling between activity 'modes'. These actions were only available whilst the mouse was in the window rendering the virtual world (which covered the majority of the desktop display). In addition, the interface contained two buttons, for peripheral glancing in both directions, and a reporting facility for displaying the current activity mode.

Video and audio information were collected from the effective 'perspective' of each person involved. The type of data compiled consisted of the visual view of the desktop display each participant was attending, along with their voice in real time mixed with the other subject's voice as it was heard locally – in other words, after a delay in delivery across the network[1]. Users were also interviewed on their opinions and experiences of the system. All subjects were students with limited or no expertise in virtual environments, with a variety of age and previous intimacy, and of both genders. Six pairs of participants performed the task over approximately one hour per pair, including an acclimatisation period during which the interface controls could be learned. Despite its complexity, informal viewing seemed to indicate that novices fulfilled the remit of the task relatively easily. However, a more rigorous, qualitative study of the data collected during the trials was undertaken.

[1] Fast ethernet speeds meant that there was minimal delay in update between users. Thus we treat talk and visual conduct as occurring in 'real time' from either participant's perspective, although sometimes small discrepancies in timing occur.

## OBSERVATIONS

Whilst there have been relatively few studies of the ways in which CVEs support and enable interaction between users, some observational analyses have been conducted with these kinds of technology. Amongst these include studies by Bowers et al. [4, 5] of problematic interactions in and with an early CVE system, and the work noted earlier on object-focused collaboration in CVEs [12]. The application of qualitative studies has been particularly successful in informing the design of collaborative virtual systems. Whilst larger user studies may be more preferable, the incipient nature of the technology is such that that would be somewhat premature. The analysis is presented with exemplars of phenomena observed in interaction through this CVE interface. Fragments are intended to display the kinds of situation that seem to recurrently, or 'typically', occur throughout the collected trial data, and figures captured from the video material are used to illustrate these points. In cases where the limitations of resolution render video-captured images unclear, these figures have been annotated for clarification.

### Recovering Action

The aspects introduced into the interface are intended to support awareness of others' actions. Fragment 1[2] shows how the new interface can be used to establish the actions of the other more easily than the earlier interface. Gerry inspects the completed version of the virtual room layout ("the red one") in order to provide both himself and his co-participant, Bob, information about how to position blocks relevant to the next stages of their task plan. However, Bob encounters a problem with a part of the layout considered by Gerry to have been completed.

Fragment 1

```
Gerry: I'll go down to the far end of the
       red one umm and tell you sorta
       whereabouts they need moving
Bob:   yeh umm d'ya see this pile I'm
       pointing at there?
Gerry: (2.0) ((glances left)) Oh yeh
Bob:   They need to be moved sorta to my
       left so I'm just gonna
       ((G stops glancing))[give em a push
Gerry:                      [Oh right
```

Both Gerry and Bob use certain tools provided within the interface to retrieve the kinds of information that are needed to resolve a mutual perspective on the scene.

After Bob asks whether Gerry can see the "pile" that he is pointing at, Gerry performs a peripheral glance to the left in order to get an undistorted perspective on the activity (Figure 5b).

[2] In transcribing users' talk, single round brackets indicate the length in seconds of pauses in talk, double round brackets provide additional comment, and square brackets indicate overlapping speech.

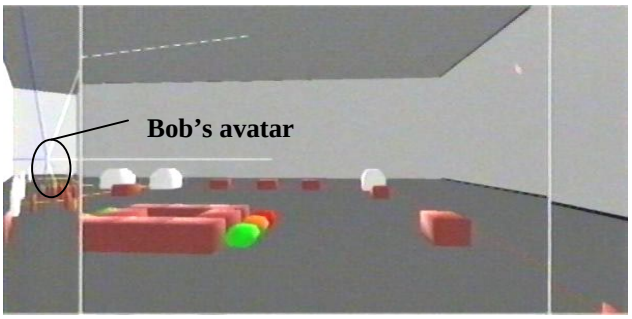


**Figure 5a – Gerry's view before glancing**

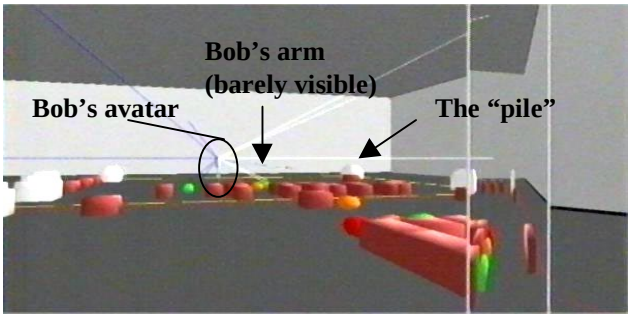


**Figure 5b – Gerry's view after glancing**

From, Gerry's initial view (Figure 5a), Bob's avatar is highly distorted and it would be difficult to discern its actions. Indeed, the avatar would even be hard to locate without the attached lines denoting the field-of-view. These give additional visual cues as to his location. So, Gerry is able to identify his co-participant's location through the combination of his own distorted lens and the other's visual representation. The peripheral glance enables him to divert his attention to the newly introduced action. If we imagine Figure 5a to have no peripheral lenses or explicit representation of field-of-view, we begin to see how limiting a standard CVE display proves in identifying the location and composition of the other's actions. However, with the peripheral resources provided, an awareness of the activity and subsequent attention change allows the sequence to continue without the possible problems of locating the pointing gesture or the pile of blocks.

Bob also draws on new aspects of the interface in this sequence. During the two-second pause after Bob's request, he extends his pointing arm towards the "pile" twice. He is aware that Gerry is some distance away, and that he may be unable to retrieve the target of the gesture. In attempting to facilitate recognition, the action is extended into the environment in order to reduce distance between gesture and artefact. Here we begin to see a participant using the extension of action to facilitate awareness of that action, in this case the visibility and intelligibility of a pointing gesture.

### Embedding Reference

Our previous work showed that when referencing objects, the 'fragmentation' of the space (between avatar and object) often disrupted the ability of users' to locate the relevant object. Fragment 2 shows how the extended representation of grasping an object allows co-participants to retrieve the action, which helps to overcome some of these difficulties. Graham and Brian are beginning to work on the edges of the symmetrical structure. Rectangular blocks are available in the world, which are supposed to be placed at specific edges of the design, based on their orientation. Brian begins by noting these differences.

Fragment 2

```
Brian: I think there are two kinds of
       squares, some are made for
       horizontal edges and some of them
       are made for the vertical edges
       (0.2) so [err
Graham:          [Oh yeh that's true, so
        for example that one there should
        go to (0.3) there
Brian:  err yeh you're right
```

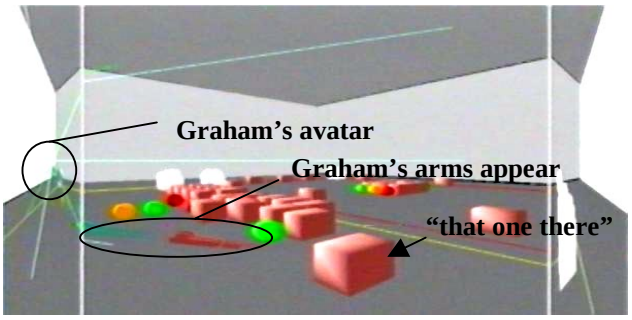


**Figure 6 – Graham's arms appear in Brian's view**

Brian keeps his avatar (and thus his view) stationary throughout this fragment and he has Graham's location available in his left peripheral lens (Figure 6). When Graham grasps an object to provide an example of his agreement, his 'arms' automatically extend towards it as he 'picks it up'. As Brian's undistorted view is oriented to the block, he can immediately re-assemble the action through its extended representation – Graham's extended arms and the subsequent wire-framing of the object. He is able to identify Graham's references to both the object ("that one there") and the relevant next location ("should go to there") without any displayed problems or queries. Without these representations, problematic interactions might arise. Using the previous system a grasping action only used a thin line, which an on-looker often could not see. Therefore, it could make it hard to connect the movement of object with an avatar. Here, the exaggerated arms provide a clear presentation of the action of grasping and the relevant block is seen quite unproblematically. The connection between the avatar and the object is readily and visibly available, such that the talk can be easily understood, without recourse to further questioning. So, the activity progresses without problems of referencing the object or the candidate location.

In our previous study it was found that in the case of three participants performing a design task, only one

other was required to acknowledge a referent, leaving the third without the resources to identify the action. So, it may be that in multi-party interaction, the kinds of exaggerated presentation of actions are even more important to retaining a flow of activity.

### Coping with Problematic Activity

In the previous two sections, we have shown how participants use features of their interface to the virtual environment as resources with which they support their own, and the other's awareness of action. Whilst these fragments are typical examples of the ways that subjects performed the task, there are some cases contained within the video data where users find problems with their use of these resources. In this section, we look at illustrative fragments of two particular recurrent obstacles: Fragment 3 looks at difficulties associated with the boundaries between interface lenses; and Fragment 4 discusses the ability to discern the emerging course of an action.

The following fragment contains a subject pairing that have experienced problems with positioning blocks individually. The CVE does not support stereoscopic vision, and perceiving the depth of some artefacts relative to others proves difficult. They have approached the problem through collaboratively positioning blocks by observing the task from orthogonal angles. Each participant adjusts the resulting structure according to which artefact's positional dimensions are best available from the individually adopted viewpoint. Gene is content with the block structure as viewed from his angle. His next step is to ask Barry whether they are correctly placed from the complementary view. However, the lens distortion hinders Barry's ability to confirm or deny the correctness of the block positions.

Fragment 3

```
Gene:  OK what do they look like?
Barry: mm they look pretty good (1.0)
       ((moves forwards and turns left))
       hang on, I'm not square on though
Gene:  (3.8) ((glances right)) (2.0)
       ((stops glancing)) (5.2)
       err do you wanna to square them up?
Barry: uh ((moving forwards again)) (1.0)
       think its just that middle one
```

The structure is initially depicted in Barry's left peripheral lens (Figure 7a). During the fragment he is attempting to get into a position from which he can view the structure perpendicular to one of its sides. However, he has some difficulty in quickly accomplishing this manoeuvre. At one point, the structure in the distorted lens appears to be in an appropriate position. It is necessary for him to obtain an undistorted view of the structure to deliver his account, so he turns to place it in the central lens. However, when he does this he still turns out to be at an inappropriate angle to the structure (Figure 7b). Barry then attempts a series of movements to re-orient his view correctly.

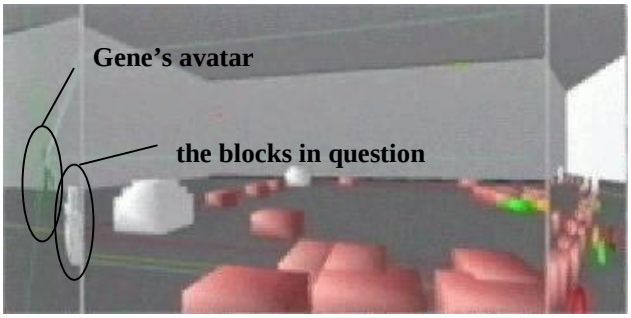


**Figure 7a – Barry's view before turning**

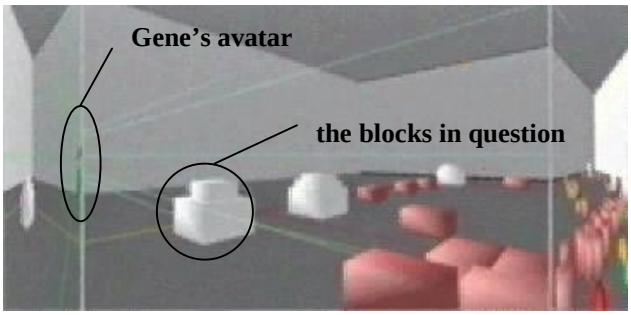


**Figure 7b – Barry's view after turning**

When turning around, or glancing, users have to deal with the sudden change in distortion at the boundary between the main and peripheral lens. The lack of a smooth transition between views seems to make it hard to assess the relative positioning of objects in the different lenses. So, as objects move across lens boundaries, views of them and their position in the world alter quite dramatically. For example, users may assume that anything viewed within the left peripheral lens is 'to the left'. However, because the undistorted view is 60 degrees, the objects in the lens are actually 'forwards and left' of their view.

Although this is a problem for individual users of the peripheral lens interface, it has consequences for the interaction in the above fragment. The difficulties for Barry mean that the collaborative task is prolonged until he is able to position himself adequately.

We now turn to extended representations of actions. There are occasions within the video data when users find it difficult to assess the emerging course of an action. In particular, we will consider the course of pointing gestures. When pointing out an object for the other, one might expect that the actual production of the gesture is quick and straightforward. However, with this CVE interface, the production of an extended pointing gesture can take some time. This is due to the potentially large virtual distances involved and the slow speed of the system. Therefore, it may take some time before the gesture reaches its target. Were these trials to have been conducted over wide-area networks, this situation would

have become more conspicuous, due to additional network delays.

In Fragment 4, Gemma and Barbara are selecting which blocks they wish to use in the corners of the structure. This fragment is taken from early on in their task, and the area is still somewhat cluttered with similarly shaped blocks. Having agreed upon the use of one block, they move on to discussing the "one next to it".

Fragment 4

```
Gemma:   there's also one next to it
         errm (.) hang on
Barbara: ((B points to the correct block))
         just there
Gemma:   err nope further (0.2)
         ((starts to point)) errrr (0.5)
         hang on oh oh oh oh
         ooohhh[hhhh
Barbara:       [oo there's your arm
Gemma:   that one (0.2) can you see what
         I'm pointing to?
Barbara: um (0.2) yeh I think so
```

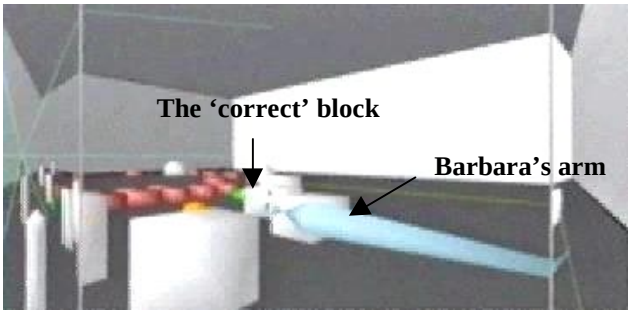


**Figure 8a – Barbara's view as she points**

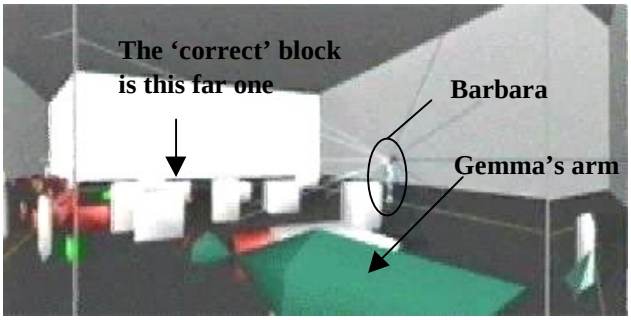


**Figure 8b – Gemma's view as she extends her point**

Barbara is in close spatial proximity to the block in question, and therefore is able to produce a pointing gesture to the correct block fairly quickly and during her utterance. However, Gemma is unable to see where Barabara's avatar is pointing. This is due to occlusion of the gesture by other blocks, and may also be because the point is not extended far enough to touch the relevant block in a crowded scene of similar artefacts (Figure 8a), hence Gemma's "further" comment. Gemma then proceeds to attempt her own referential gesture (Figure 8b). However, because of her relatively large distance from the object, it takes time to extend her gesture to the correct target. She displays these difficulties through her talk, in some ways vocally 'animating' her virtual arm as it moves towards the target object ("hang on, oh oh oh oh ohhhhhhh").

It is noticeable from the video data, however, that Gemma is not simply displaying an unfinished action in these noises. As the pointing arm extends, it produces a jerking movement rather than smoothly flowing outwards. At each jerk, the arm momentarily comes to rest by different blocks that lie between Gemma and the target block. She uses each "oh" noise to show that these blocks are not the relevant blocks, and the pointing gesture is still on its way. So, each "oh" sound displays that, whilst an object could be a target for this point – because it is an object, and because it is an object of the kind we have been discussing – it is, in fact, not the relevant one for this gesture. Therefore, Gemma attends to the jerking movement produced by the system by using her voice to portray to Barbara that the gesture is still in progress.

Interestingly, Barbara does not even see the arm until it comes into her main view ("oo there's your arm"), as the arm is too thin to be seen in the distorted peripheral lens. She only has access to Gemma's activities through her vocalisations [cf. 12].

## IMPLICATIONS

Our observations indicate that the interface described supports certain kinds of awareness in interaction through CVEs. However, there are cases where the interface causes difficulties for interaction between the users. In summary:

- Peripheral Lenses and extended representations of action are often used as a successful resource in overcoming problems observed in object-focused interaction through CVEs;
- Peripheral lens distortion can disrupt both a user's own sense of, and their notion of the other's, orientation to actions and features within the environment. Distortion may cause representations of actions performed by the other to be unavailable or visually indistinct;
- Extended visual representations of action may still be insufficient within certain contexts. Observations indicate that in cases where the emerging course of an action may be confusing, users cope by explicitly describing or vocally animating these aspects

Whilst our proposals seem to have met with some success in constructing an interface which supports awareness and interaction in CVEs, our observations lead us to propose that certain issues must be addressed further.

### Peripheral Lenses and Field-of-View

Awareness of actions seems to be supported through the use of Peripheral Lenses and Peripheral Glancing. In particular, the combination of the peripheral lenses and the lines denoting the other's field of view, seem to provide useful resources for locating the position of the other. It is interesting to note that, whilst the lines representing a user's viewpoint are rarely used to discern what another can see, they are often used as a very successful device for judging the approximate location or view of the other. For example, when participants are large virtual distances apart, or facing diametrically opposite directions, these lines are often followed to locate the other's avatar. Similarly, when the other is in a distorted lens the avatar itself is almost imperceptible, but the lines work well to extend that embodiment and pinpoint the other.

So, the lines had unanticipated benefits in terms of locating the other in space – a critical problem in the previous experiment. Nevertheless these lines did not seem so successful in terms of providing the resources for users to assess what the other has in view. The lines were originally intended to convey the viewpoint of a colleague in both distorted and undistorted conditions. Perhaps the use of a transparent frustrum, or the use of shadows might prove more successful in this respect. It might also be interesting to discover if these representations are additionally used in locating another's avatar or action. Of course, when designing representations for applications, the actual visual representation must attune to the nature of the task. For example, a CVE application supporting large numbers of users could quickly become crowded with badly designed representations. Imagine a virtual medical operation application using the representation employed here to show the other's view. Confusion between artefacts and representations might well ensue.

Peripheral Lenses seem to provide an improved sense of the location and on-going action of other users embodied in the virtual world. However, distortion can render the location of features of the world with respect to one's viewpoint orientation difficult to perceive. Representations of the other's action may also be unavailable within a lens. We might contemplate reducing the distortion in the Peripheral Lens to alleviate these misconceptions, either through a reduced fixed value, or through allowing a user to vary the lens angle as in [13]. However, as the horizontal space on the desktop display is limited, this would require a similar reduction in the user's field-of-view.

Another possibility is to address the ways in which representations of action are viewed in distorted conditions. Although this does not explicitly address problems with orientation to artefacts within the environment, it might allow users to maintain knowledge of the actions being performed with and around those artefacts. We might achieve this through solely rendering certain features of the world within peripheral lenses, such as avatars and objects relevant to the task at hand. Another possibility would be to design the representation of actions through environment and on features of the world specifically with the knowledge that they might have to be re-assembled in distorted conditions. For example, visual cues such as colour changes or flashing may be more obvious than inanimate representations of action. These visual cues might also enable the location of the avatar from representations, and thus improve the sense of the connection of body and object in viewing distorted actions.

All of these factors will influence our continued research into the design of virtual embodiments and interfaces for practical, object-focused interaction.

### Course of Action

An essential feature of collaborative object-focused interaction in CVEs is the ability to recognise the connection between the other's avatar and the relevant artefact. As displayed in Fragment 2, extending representations of actions can alleviate the need for explicit prefatory referential sequences prior to some activity regarding the relevant object. However, there are cases where extended representations of actions such as pointing can cause problems. In the case of Fragment 4, the emerging course of the action is unavailable in its representation. In that case, the user highlighted its incompleteness through her vocalisations. It would seem a rational step to somehow acknowledge the course of an action within that action's representation.

It is unlikely that we might employ the CVE application itself to determine and represent the course of an action. For example, understanding which region a point might indicate, or the position an object being moved might finally reside, would be indefinable by the application. Only the particular user might determine this, and even then only on a contingent moment-by-moment basis. We might, however, display this inability in the extended representation of the user's action, or the potential targets of that action.

The representation of an extended action could show the incompleteness of the action through a wire-frame, or semi-transparent, representation. For example an avatar's arm might be wire-framed whilst being extended or retracted, just as an object was, for the duration of a grasp in this implementation – although this particular suggestion might have the unwanted effect of making action even harder to perceive in a distorted lens. In representing potential targets, we might display a range of most likely options available at any particular

moment. For example, whilst extending or retracting a pointing gesture, a number of most likely targets could be displayed. The fact that a number of options exist *at all*, regardless of whether they, all along and in the first place, are correct suggestions could display to the other the incomplete nature of the action. Of course, this is not a straightforward issue to address, as these solutions may add greater confusion for the participants. For example, if different objects become highlighted by the system as potential targets, they may get treated prematurely as actual targets, thereby adding a further disruptive element. This is very much a matter for future experimentation and development, and thus we continue to explore the different ways of displaying actions more clearly in CVEs.

## FUTURE WORK

Supporting awareness in interaction seems to be a crucial factor in allowing tasks to be effectively accomplished through CVEs. We believe that this study of awareness in collaboration has important implications for the design of CVE systems. However, whilst qualitative analyses in 'experimental' contexts have been beneficial in discovering problems with and solutions for CVE interfaces, these kinds of ethnography excel in naturalistic environments. The study of CVE use in 'real' applications might provide not only rich and varied empirical data on collaborative practices, but also move these kinds of technology from experimental interfaces to becoming successful communication applications. Nevertheless, if this position is to be attained, a basic understanding of the key problems and issues for interaction through distributed interfaces is critical to the eventual deployment of CVE systems in commercial environments.

## REFERENCES


1. Benford, S., Greenhalgh, C. and Lloyd, D. Crowded Collaborative Virtual Environments. In *Proceedings of the ACM Conference on Computer-Human Interaction (CHI'97)*, Atlanta, GA, USA, pp. 59-66.
2. Benford, S., Bowers, J., Fahlén, L.E., Greenhalgh, C. and Snowdon, S. User Embodiment in Collaborative Virtual Environments, in *Proceedings of the ACM Conference on Computer-Human Interaction (CHI'95)*, Denver, Colorado, USA pp. 242-249.
3. Benford, S. and Fahlén, L. A Spatial Model of Interaction in Virtual Environments, In *Proceedings of the Third European Conference on Computer-Supported Co-operative Work (ECSCW'93)*, Milan.
4. Bowers, J., Pycock, J. and O'Brien, J. Talk and Embodiment in Collaborative Virtual Environments. In *Proceedings of the ACM Conference on Computer-Human Interaction (CHI'96)*, Vancouver, Canada, pp. 58-65.
5. Bowers, J., O'Brien, J. and Pycock, J. Practically Accomplishing Immersion: Cooperation in and for Virtual Environments. In *Proceedings of the ACM Conference on Computer-Supported Co-operative Work (CSCW'96)*, Boston, MA, USA, pp. 380-389.
6. Fraser, M. and Glover, T., Representation and Control in Collaborative Virtual Environments, In *Proceedings of the Fifth U.K. Virtual Reality SIG Conference (UKVRSIG'98)*, U.K.
7. Gaver, W., Sellen, A., Heath, C and Luff, P. One is Not Enough: Multiple Views in a Media Space, In *Proceedings of the ACM Conference on Human Factors in Computing Systems (INTERCHI'93)*, Amsterdam, The Netherlands, pp. 335-341.
8. Greenhalgh, C. and Benford, S. Virtual Reality Tele-Conferencing: Implementation and Experience. In *Proceedings of the Fourth European Conference on Computer-Supported Cooperative Work (ECSCW'95)* Stockholm, Sweden, pp. 165-180.
9. Gutwin, C. and Greenberg, S. Design for Individuals, Design for Groups: Tradeoffs Between Power and Workspace Awareness. In *Proceedings of the ACM Conference on Computer Supported Co-operative Work (CSCW'98)*, Seattle, WA, USA pp. 207-216.
10. Heath, C., Jirotka, M., Luff, P. and Hindmarsh, J. Unpacking Collaboration: the Interactional Organisation of Trading in a City Dealing Room. In *Proceedings of the Third European Conference on Computer-Supported Co-operative Work (ECSCW'93)* Milan, pp. 155-170.
11. Hindmarsh, J. The Interactional Constitution of Objects. unpublished PhD. thesis, U. of Surrey 1997.
12. Hindmarsh, J., Fraser, M., Heath, C., Benford, S. and Greenhalgh, C., Fragmented Interaction: Establishing Mutual Orientation in Virtual Environments. In *Proceedings of the ACM Conference on Computer Supported Co-operative Work (CSCW'98)*, Seattle, WA, USA pp. 217-226.
13. Robertson, G., Czerwinski, M. and van Dantzich, M., Immersion in Desktop Virtual Reality. In *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST'97)*, Alberta, Canada, pp. 11-19.
14. Smith, R. B., Hixon, R. and Horan, B., Supporting Flexible Roles in a Shared Space. In *Proceedings of the ACM Conference on Computer Supported Co-operative Work (CSCW'98)*, Seattle, WA, USA pp. 197-205.
15. Steffik, M., Bobrow, D., Foster, G., Lanning, S. and Tatar, D., WYSIWIS Revised: Early Experiences with Multiuser Interfaces, *ACM Transactions on Office Information Systems, 5, 2*, pp. 147-167.