Talk and Embodiment in Collaborative Virtual Environments

John Bowers
Department of Psychology
Manchester University, U.K.
+44-161-275-2599
bowers@hera.psy.man.ac.uk

James Pycock
Department of Psychology
Manchester University, U.K.
+44-161-275-2682
pycockj@hera.psy.man.ac.uk

Jon O'Brien
Department of Sociology
Lancaster University, U.K.
+44-1524-594186
soajeo@cent1.lancs.ac.uk

ABSTRACT

This paper presents some qualitative, interpretative analyses of social interaction in an internationally distributed, real-time, multi-party meeting held within a collaborative virtual environment (CVE). The analyses reveal some systematic problems with turn taking and participation in such environments. We also examine how the simple polygonal shapes by means of which users were represented and embodied in the environment are deployed in social interaction. Strikingly, some familiar coordinations of body movement are observed even though such embodiments are very minimal shapes. The paper concludes with some suggestions for technical development, derived from the empirical analyses, which might enhance interactivity in virtual worlds for collaboration and cooperative work.

Keywords

Conversation analysis, interaction analysis, body movement, embodiment, virtual reality, CSCW.

INTRODUCTION

Collaborative Virtual Environments (CVEs), where multiple individuals interact with each other in a computational environment rendered by Virtual Reality (VR) technology, are of emerging interest from a number of perspectives. Several writers have claimed [e.g. 2] that such environments may support collaboration and interactivity (especially geographically distributed collaboration) in ways which go beyond what is possible using more familiar meeting room or teleconferencing technologies. The development of CVEs has also drawn interest as a complementary alternative to video and 'mediaspace' [5, 12] research. CVEs may provide a shared spatial environment where (in principle) people can employ communicative resources which are unavailable to them in other technical systems. For example, participants can have a degree of control over what they view in a CVE which is not generally possible with mediaspaces supported by a fixed camera and monitor system, and our ordinary means for coordinating turn-taking in social interaction can be deployed rather than some technical means such as 'floor control' policies as is commonly the case in traditional conferencing systems. Furthermore, if users in a CVE are all embodied in it so that their location and orientation can be represented, then a degree of mutual awareness of each other's activity may arise or be readily supported [1, 2].

This paper attempts to subject such claims to preliminary examination by giving a characterisation of what CVEs are like as environments for cooperative work and social interaction and seeing how ordinary conversational mechanisms are exploited or transformed in such environments. To these ends we employ empirical techniques derived from Conversation Analysis or CA [e.g. 14], which to our knowledge has not been attempted before in studying CVEs or other VR technologies, though CA has had some influence in HCI and CSCW in enabling detailed studies of interaction in the workplace [10], the impacts of new computer based technologies on 'talk at work' [8] as well as in motivating technical design choices [3] and assisting in the analysis of the design process itself [4]. Accordingly, we seek to add to this literature while extending it to the study of a novel setting (a work-related meeting being conducted in VR).

Finally, we are concerned to show how methods of interaction analysis might contribute to the evaluation and hence future requirements of CVEs. User-oriented evaluative studies of VR systems are still overwhelmingly dominated by investigations of such matters as motion sickness [13] and the characterisation of phenomena in terms of individual perceptual psychology [15]. We wish to study CVEs so as to extend the base in reference to which VR systems should be evaluated and developed. Through characterising the nature of social interaction in a CVE at least in a preliminary fashion, we hope to make a start on the task of exploring the worth of distributed VR technologies for the support of cooperative work and social interaction.

THE VIRTUAL MEETING

We have been working closely with CVE developers for some time in a number of research projects. As their CVE technologies have become more robust, our collaborators have begun to use their own systems as environments for the support of distributed research meetings. We have been present as participant-observers at a number of these meetings and have begun to accumulate a corpus of videos and other materials for analysis. At the time of its occurrence, probably the most significant of these meetings took place on 28th March 1995 as a working meeting of COMIC, a European ESPRIT project devoted to basic

CHI 96 Vancouver, BC Canada
© 1996 ACM 0-89791-777-4/96/04..$3.50

research on CSCW. The significance of this meeting was that it involved real-time interaction in a number of virtual worlds between nine participants who were distributed across five sites and three countries (Germany, Sweden and the UK). According to many of the participants, this degree of international participation was a 'world first' and has been reported as such in a number of technical press articles.

The MASSIVE system [9] that was used for the virtual meeting supports multi-user interaction between distributed sites allowing participants to communicate over graphical, textual and audio media. The graphical interface provides a navigable 3D view of the shared virtual world and of other participants represented as simple graphical embodiments. For the current meeting the 3D view was presented on screen rather than immersively. The audio interface allows real-time conversation. The text interface provides a 2D plan view of a world and allows the exchange of text messages. MASSIVE employs client-server information distribution. The hardware used was exclusively Silicon Graphics for the interface clients with a Sun Sparc 10/51 at Nottingham University in the UK running server software.

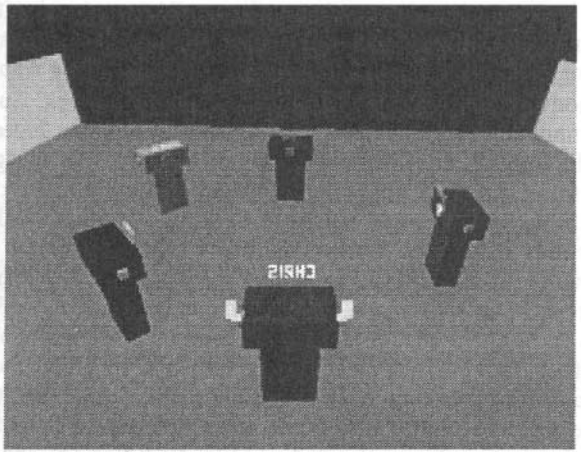

Figure 1. Blockies in a virtual meeting.

The user-embodiments or 'blockies' are made of simple 3D box-polygons with one square 'eye' on one vertical surface and the user's name 'suspended' above the top surface. As well as identity, this design affords a rudimentary sense of 'face', 'front' and 'back' which according to Goffman [6] are features of the human body of basic significance to social interaction enabling us to distinguish between, for example, talking to someone's face from talking behind their backs. Furthermore, the 3D view that a user has can relate to the blockie's face. Although different views are possible, the default is that one 'sees out of' the blockie's eye. This view, or one where one looks over the blockie's shoulders, are the ones typically employed by users. In this way, what other participants can see and where they are looking is often available from an inspection of their bodily orientation and, accordingly, a sense of mutual awareness can be sustained and transformed by aligning the blockies or moving them around. For example, 'full face encounters' [6] can be brought about by two participants aligning their blockies to face each other. Indeed, it has been precisely a consideration of what was the minimal geometrical object necessary to sustain basic interactional relations between participants which has informed the design of the blockies [1]. Insisting that the embodiments should be geometrically simple (yet still have interactional potential) is necessary because of the extreme computational complexity of distributed VR systems based on current technology.

The blockies also support minimal gesturing. They have 'ears' which can be 'flapped' in different ways (left one raised, right raised etc.). They can recline (or 'sleep'). This can be used, for example, to denote that the participant the blockie corresponds to has currently left their local machine and is not available for interaction. Gestures are controlled through simple key sequences and the blockie is moved by clicking the mouse on the 3D view or by using the arrow keys. Finally, the blockies have a 'mouth' which opens when a user's speech exceeds a certain amplitude threshold.

The overall business of the virtual meeting is well described by the following quote from the minutes which one of the participants distributed afterwards: "The time spent in the conferencing software, approximately an hour and a half, can be split up into three distinct periods. To begin with there was a fifteen minute mingling session where people arrived for the meeting and chatted socially ... This time was also spent making sure everybody could communicate with everybody else. At a quarter past two SB called the meeting to order and everybody trooped off to the designated meeting room where a pre-prepared agenda was awaiting on a noticeboard and ... introductions were carried out. CG then gave a tutorial on how to use MASSIVE ... [The items on the agenda were then discussed.] The formal meeting finished at about two fifty five, and after that much more informal communication took place." While some of the clients needed to be restarted from time to time, the server software and network connections were reliable throughout.

INTERACTION ANALYSIS

In this section, we report our analyses of the talk exchanged between participants in the virtual meeting. Specifically, we focus on two issues: the nature of turn taking in virtual environments and how the embodiments are used. Our interest is in the *interactional qualities* of CVEs, what these environments are like as arenas for social interaction as revealed in the moment-by-moment texture of talk and activity. Accordingly, our analyses work from transcript data which we analyse qualitatively and interpretatively.

Before we turn to some examples from our data, it is necessary to clarify the transcript conventions we have adopted. We employ an adapted version of the conventions devised by Jefferson and presented in a number of sources [e.g. 14]. Pauses and silences are notated by their length in seconds shown within round brackets: `(1.2)`. Talk which receives more emphasis than the surrounding speech is underlined: `someone else's turn`. Parts of the transcript where we are unsure of what is said, but are able to guess, are notated with round brackets: `thanks lennart (how) very eloquent`. Where we are unable to guess the round brackets are empty. Prolonged sounds are indicated by inserting colons, and concatenated speech, where the words are quickly run together, is notated by hyphens placed between the words: an::d this wonderful VR system-MASSIVE-that-we're-using is wot i writ. An audible in-breath is notated .hhh and hhh. denotes an audible out-breath. Overlapping speech is notated by means of square brackets positioned to indicate where the overlaps occur. Speech which is distorted by, say, some malfunctioning of the audio link is placed within curly brackets: i'm john {b}owers from manchester {uni}versity. Our comments within the transcript are enclosed within angled brackets: <laughter>. We shall discuss our conventions for transcribing the movements of the blockies later.

Taking Turns at Talk

Turn taking is a basic and obvious feature of the organization of talk. Some aspects of how turn taking is managed by participants in ordinary conversation are analysed by Sacks, Schegloff and Jefferson [14] who propose what they call a 'simplest systematics' by means of which participants manage the exchange of speakers from turn to turn. In the analysis of turn taking, a basic distinction can be made between (i) turns which select who is next to speak by, for example, addressing a question (ahh lennart can you hear us?) or request (go on dave) to a specific, perhaps explicitly named, participant and (ii) turns which do not contain next-selecting components:

```
AB:     <mouth click> .hhh (2.2) <mouth click> .hhh-woahruh
        i'll go next (.) then if no one else is speaking.
        (0.6) uh i'm adrian bullock also from the
        university of nottingham.
```

In AB's turn here, it is AB who selects himself to speak, indeed he does so quite explicitly noting that 'self-selection' is exactly what he is doing (i'll go next). AB's turn itself contains no components which project who is next to speak after him. He introduces himself and then stops. At such moments, it is open for the other participants to either select themselves as next-to-speak or, for that matter, for AB to continue speaking prolonging his current turn.

Next Selected by Self

Our preliminary observations suggested to us that turn taking was often problematic in the virtual meeting. The transcript was marked by many long silences as participants seemed to wait for each other to say things. However, a closer inspection reveals that these silences are most prevalent when speakers have to select themselves as next-to-speak in the absence of any prior turn with next-selecting components. This is not to say that all such turn transitions break down. Rather it is to claim that the disfluencies in speech exchange are notably concentrated at transitions where speakers have to self-select. The following examples (from the early stages of the meeting where SB has invited participants to introduce themselves) reveal problems in speaker switches of this sort.

Example 1

```
SB:     someone else's turn. (1.2) thanks lennart (how)
        very eloquent. (1.2) maybe (eloquent.)
(12.0)
CG:     i'll 'ave a go then.
(1.2)
SB:     yeah please do.
(1.4)
CG:     i'm chris greenhalgh (0.4) also at nottingham (0.8)
        an::d this wonderful VR system-MASSIVE-that-we're-
        using is wot i writ.
```

Example 2

```
SB:     <laughter> how much does it cost?
CG:     oo to you nothing. bargain at double the price.
(1.0)
SB:     excellent.
(3.0)
AB:     <mouth click> .hhh (2.2) <mouth click> .hhh-woahruh
        i'll go next (.) then if no one else is speaking.
        (0.6) uh i'm adrian bullock also from the
        university of nottingham.
```

Example 3

```
SB:     anyone else?
(1.2)
AC:     {uh uh hello s-s-} can you hear me?
        [ (.)  ]=anyone? [ (.) ]=err
SB:     [ yeah=]         [ yea=]
AC:     it's andy at lancaster.
```

Example 4

```
(1.2)
( ):    (.hhh)
(4.0)
SB:     anyone else?
(2.8)
JB:     huhlow::. i'm john {b}owers from manchester
        {uni}versity currently v{i}sit{in}g sics. (0.8) i'm
        interes{ted in} lots-of stuff (0.4) {and i'm a
        c}apricorn.
SB:     <laughter>
```

In Examples 1, 3 and 4, SB does not explicitly name who is next to introduce themselves. Rather he invites contributions by means of anyone else? (Examples 3 and 4) or someone else's turn (Example 1), leaving it up to whoever is next to speak to select themselves. Such components are followed by quite lengthy silences. Indeed, in Example 1, after just over a second's silence, SB engages in some 'side talk' thanking LF in an ironically humorous fashion for his prior self-introduction. Another twelve seconds of silence follows before CG starts talking. In Example 2, SB and CG exchange some humorous remarks before SB utters excellent which can be heard as closing his exchange with CG. The three second pause that follows is then broken by AB selecting himself. In all of these examples, speakers only self-select after quite lengthy silences and with much preparatory activity (e.g. mouth clicks, in-breaths, protracted sounds, stammerings and so forth) or, as in Example 1, a minimal turn from CG (i'll 'ave a go then) which requires confirmation from SB (yeah please do) before CG continues.

This preparatory activity is often quite exaggerated as in Example 2 where a mouth click and an audible in-breath are heard from AB and then, after a pause of just over two seconds, there follows a second mouth click, another audible in-breath running into a vocalization transcribed as woahruh before AB explicitly self-selects. The interactional significance of such preparatory activity is worth noting. Audible in-breaths and the rest do not explicitly or fully claim a turn at talk in and of themselves. They could be interrupted by another participant immediately launching into a turn without such activity. Accordingly, such preparatory activity displays a participant's *readiness* to contribute as next-to-speak, without disqualifying others. Indeed, in Example 2, even *after* he self-selects, AB pauses very briefly (notated by (.)) after i'll go next and for about six tenths of a second after then if no one else is speaking. These are further junctures where another participant could have self-selected and claimed the floor to introduce themselves ahead of AB. These features of the examples suggest that self-selected turns are managed with considerable care by speakers - a matter borne out by the fact that AB in Example 2 is explicitly attentive to the possibility that others may speak ahead of him. The exaggeration of preparatory components in the virtual meeting is, we suggest, a means for managing turn taking at moments which can be problematic where, for example, in the absence of explicit next-selection, a number of speakers could start to speak simultaneously. Indeed, Examples 5 and 6 suggest that simultaneous self-selected turns are problematic for the smooth conduct of the virtual meeting and that the presence of audio distortion makes them especially hard to manage.

Example 5

```
(3.0)
SB:      [ (  )
JB:      [ (      )(fahlén)
(3.7)
LF:      okay. errm. (2.5) we have err (.) umm (.) a rather
         a serious echo here.
```

Example 6

```
(1.8)
LF:      {i think we ermm} (0.5)
KJ:      [ {(let's put) that upon} the agenda ] (1.0)
LF:      [ {(                             )} ] {uh }
KJ:      [ (     ) ]
LF:      [ {(      ] ) some kind of uh distortion (.) uhhhm
         (0.5) originating at K T H. you should probably
         lower yer input levels a bit.}
(1.0)
KJ:      okay i'll try.
```

In Example 5, after a three second silence, SB and JB start speaking together. JB's speech continues after SB's but neither of their turns are taken up by any other participant, ratner the talk dies away into a silence of nearly four seconds before LF complains about the audio quality. In Example 6, KJ and LF talk together and then after LF utters uh while KJ falls silent, they both restart together again with LF's complaints about the audio again heard after KJ's talk finishes. In cases like these, the overlaps leave unfinished interactional business with one or both overlapping turns failing to be taken up by others in subsequent talk. The overlaps in Examples 5 and 6 are hard to transcribe for the precise reason that the voices tend to mask each other and this is exacerbated by the poor audio quality LF complains about. Not only do they mask each other so as to make transcription difficult, they interfere so as to make it hard for participants themselves to hear the talk and tease apart overlapping contributions. We suggest that it is because of the problems involved in the management of overlaps of turns that self-selected turns often manifest the preparatory activity we have noted.

Note again the presence of various artifacts (e.g. distortion) in JB's talk in Example 4. In the next example, this is particularly intense. SB does not hear DE's eh:::m:. (0.5) yuh as preparatory to a turn at all!

Example 7

```
(2.0)
SB:      okay-is there any shy: person who hasn't spoken
         yet?
(3.0)
DE:      {eh:::m:.} (1.5) {yuh}
SB:      i think that's everyone.=
DE:      ={ouh}=
LF:      =steve-d'yuh-think-i think you could probably
         [ errr (0.9) ] raise your levels a bit?
DE:      [ {hallo::?} ]
SB:      who me? maybe.=
DE:      ={huhlo?}
AB:      <mouth click> yeah go on dave.
DE:      (2.5) {right. um. [ (.)  ] (first)ly can you hear
         me:?}=
LF:                        [ (  ) ]
         =you're much quieter [ (  )
AB:                           [ just about.
(2.8)
SB:      hi the[re.
AB:            [{ } breaking up (.) but you can just about
         make it out.
(0.8)
( ):     um.
(2.0)
DE:      {{(okay)}} (1.6) {{{(hiya (.) steve)(    [
         [
                   [           )}}}
JB:                                               [<laughter>
SB:      [ ah ha. um.
AB:              [nah.
AB:      nah-that's unintelligible.
```

This example suggests the problems in coordinating speaker switches when at least one participant has a considerably degraded signal may be profound. It also further underlines the point we have been making about difficulties with self-selection and that these may be intensified when one's signal is poor. Indeed, very few of DE's turns at talk are brought off in Example 7 without some overlap or interruption from the other participants.

Interestingly, DE's first attempt to claim a self-selected turn at talk in this example manifests very similar preparatory features to those we have already seen in Examples 1 to 4. It is unsuccessful presumably because SB does not hear them as having this significance or as being any different from the artifacts, pops and crackles and other background noises that can be heard on the audio channel. DE having the poorest audio connection is doubly disadvantaged: first in that his speech is easily masked by others in overlap, secondly in that his routine attempts to anticipate the problems of overlap (e.g. protracting an ehm or uttering a preparatory yup) are not heard as such!

It is important to emphasise that the difficulties with turn taking we have noted cannot simply be reduced to problems with audio quality. We observe substantial silences in examples where there are no (Examples 1 and 2) or few (Examples 3 and 4) audio problems. Indeed the audio connections within the site accommodating SB, CG and AB were of good quality throughout the meeting, yet these provide some of the most notable silences in our transcripts. Hence we argue that the problems of self-selection are exacerbated by but not solely attributable to poor audio quality.

The lengthy silences before speaker-switches, we suggest, reflect problems due to managing self-selection with minimal embodiments which have restricted gestural abilities. In this regard, it is of interest that DE does not attempt to compensate for interactional difficulties by any form of virtual gesture or change of body orientation at any moment in Example 7. The embodiments are very rarely used concurrently to aid speakers in designing their own turns or in eliciting turns from others at such moments (in contrast to the use of gesture and body movement in ordinary co-present conversation, see [7, 11], or, for that matter, as reported in the videoconferencing literature, see [16]). This is a point we shall return to.

Next Selected by Prior Speaker

In contrast to the problems we have noted at moments where speakers have to self-select to claim a turn at talk, turn transitions are much smoother where the identity of the next person to speak is clear either because they have been explicitly named or due to some other contextual feature. In Example 8, AB gives JB instruction as to how to use the mouse button after he has been struggling to pass through a gateway to another virtual world. Similarly, in Example 9 KJ and SB discuss two different modes for 'focus' a technical feature of the MASSIVE system. Finally, in Example 10 SB notices the presence of LF and starts a discussion about the organization of a conference panel.

Example 8

```
AB:      (you can use your right mouse button for lateral)
         translations john.
JB:      i can do what?
AB:      use your right mouse button (.) in the middle of
         the screen just to move you side ways.
<JB moves sideways>
AB:      wrong way.
SB:      oh he's gone. <through the gateway>
```

Example 9

```
(KJ):    (  )
SB:      (did you) change focus?
(0.5)
KJ:      so what is the difference between narrow and
         directed
SB:      directed is a sort of cone (.) in front of you erm
         i can't remember the angle it goes out to. narrow
         is like a really thin tube
(0.5)
KJ:      ok
(0.5)
SB:      you can try this thing out in the audio gallery and
         (you'll) probably notice the effect fairly well
KJ:      alright
```

Example 10

```
SB:      oh there you are. ok so w-what happened about this
         panel thing what's the (scam)?
LF:      umm nothing much but erm we're still missing erm
         erm: the erm er spatial awareness versus ((    ))
         versus media spaces kind of panel ya?
SB:      arham are we liable t-to do one of these?
LF:      ya <conversation between SB and LF continues for
         nearly a minute with many turns readily exchanged>
```

In all these examples, the speech exchange between participants is fluent and not marked by the hesitations, overlaps and other perturbations noted before. In each of them, though, who is next to speak is clear either as a result of explicit naming or the identification of 'you' even if several other parties are present. When speaker switches can be governed in this way, fluent conversations can occur between two parties in the virtual environment. (It should be noted that these examples involve fluent exchanges between people at distant sites: SB and AB are in Nottingham, the others in Stockholm. This suggest that the problems we have documented are interactional and not just due solely to technical reasons like network delays which were rarely notable in the meeting.)

Interpreting Silences and Absences

We have suggested that there are differences between turn transitions where the prior speaker has selected who is next to speak (these are carried off relatively fluently) and those where speakers have to self-select (these are the sites of the main breakdowns of fluency in the transcript). As noted by Sacks et al. [14], these different transition types give rise to differences in the significance of silences in talk. When the prior speaker has selected who is next-to-speak, any silence after next-selection is *attributable*. That is, it is 'owned by' the person who has been designated as next to speak. It is *their* silence and their next speech action is *noticeably absent*. Depending on the circumstances, a continued silence after next has been explicitly selected may licence inferences about the person who is remaining silent. For example, in a law court, persistent silence under cross examination may lead to suspicions of guilt or other culpability. On the other hand, when the prior speaker has not selected who is next-to-speak, such attributions are not standardly available and the silence is not uniquely 'owned'.

When conversations are technically mediated, technical failure can be inferred as a source of attributable silence. Indeed, in the data we have, technical failure (rather than some socially significant attribution like evasiveness or rudeness) is *invariably* first considered as accounting for an attributable silence. In Example 11, SB first asks a question about whether another meeting is intended. AB replies to this saying that a future meeting should use the DIVE system. SB continues by asking a question which is hearably directed at two of the current meeting's members who are also developers of DIVE. However, this receives no reply and after a one second silence, SB explicitly names (and aligns his embodiment in the virtual world so as to face) the two the question is addressed to. Again, a long silence follows and AB checks on LF's ability to hear. Another long silence: whereupon SB notices and brings to the attention of others that KJ has typed a message in the text window saying he cannot hear anything. The point of this example is that when next-to-speak has been explicitly selected and no reply is heard, this is interpreted here as arising from an inability to hear due to an audio failure.

Example 11

```
SB:      are we going to have another one of these?
AB:      we ought to have one with dive.
SB:      yer (0.5) what's-what's-what's the status of that
         stuff? (1.0) he says looking at lennart and kai.
(5.0)
AB:      ah:: lennart can you hear us?
(5.0)
SB:      (hi) if you look at the text window (.) kai says
         he's not hearing anything
```

Example 12

```
AB:      can you hear me kai? (2.5) hello?
CG:      he may be a corpse now. he's disappeared off my
         text (  as far as  )
```

Similarly, in Example 12 AB's repeated failure to elicit a reply from KJ is accounted for by CG. Amongst the VR research groups involved in this meeting 'corpse' is used to refer to an embodiment which lingers in the virtual world even though the local system which manages interaction with it (in this case the interface client running on KJ's machine) may have crashed or become disconnected. In both examples, noticeably absent 'seconds' in an 'adjacency pair' [14] like question-answer are first interpreted as indexing *technical problems and not accountable social behaviour* (e.g. rudeness etc) on the part of whoever has been selected to speak. In contrast, the silences in Examples 1 to 4 where SB did not select who was next to speak, are not uniquely attributable and hence, if anyone does happen to have audio connection or other technical problems, they will pass unnoticed and possibly for some time to follow. To put this point another way, silence can be hard to interpret and technical problems can be difficult to identify, even those originating from one's own local environment. A local failure can pass unnoticed for some time simply because an appropriate moment for the problem to be identified does not occur in the interaction. We shall return to the design implications of this in discussion at the end of this paper.

Embodiment, Talk and Movement

How are the embodiments used by participants in the virtual meeting and how are their movements coordinated with concurrent talk? A first observation is that participants do move the embodiments around to get a better view of those they wish to interact with or attend to. The provision of an audio channel does not *require* them to do so in order to speak to each other, but that we have examples of people navigating the embodiments in this way suggests that, in a simple sense, interactants are seeking to become 'face engaged' [6] when exchanging talk. Accordingly, the embodiments do seem to have a social *interactional* role and not merely a role in determining the view an individual has of the virtual world. Note how SB and LF establish contact in Example 13 before starting a lengthy exchange.

Example 13

```
SB:        (   ) lennart
LF:        yar (steve)
SB:        oh what happened about um the the panel for ecscw?
(LF):      (   )
SB:        c-could you turn (the) mic up a bit?
LF:        (   )
SB:        (   ) where has he gone lennart?
(2.0)
LF:        ar where's steve?
SB:        hi i'm i'm here behind kai i think
LF:        ye
(4.0)
<LF and SB move around the environment and finally turn to
face each other>
SB:        oh there you are. <continues as Example 10>
```

Do we have evidence from the recordings of the virtual meeting of any more subtle uses of virtual body movements to accompany ongoing interaction? To examine this, we first consider whether body movements exist in the virtual world as an accompaniment to one's own talk and then look at whether participants are able to coordinate their movements with those of others and their talk.

We remarked above that participants seem to rarely use virtual body movements to aid the design of their own turns, even when constructing turns by means of talk alone is problematic (as in Example 7). Indeed, it is rather rare in the virtual meetings we have studied for people to complement their own turns at talk with any concurrent movement of their embodiment. This may not strike one as surprising as talking down a potentially troublesome audio channel may be difficult enough without having to engage in simultaneous mouse movements to get one's embodiment to move! However, it presents a stark contrast with ordinary talk where a whole array of body and facial movements, gazings and changes in overall deportment can accompany and aid the design of turns at talk [7, 11]. Example 14 transcribes the body movement from Example 1. Here, CG raises and lowers the ears on his embodiment. These gestures span his breaking of the long silence we have already noted and aid the construction of his self-selected turn. This is however the only example we have yet found of gesture being used to aid the design of a speaker's concurrent turn. (It also introduces our transcript conventions for showing movement. The beginning and end of the movement are shown underneath the concurrent talk and described in italics on a line after that.)

Example 14

```
<The virtual body movements in Example 1>
           ( 12.0 )
CG:              ¶-
CG:        raises both ears

CG:        i'll 'ave a go then.
CG:        ------¶
CG:        and lowers them
```

While we have few examples of participants designing their own turns with the aid of some movement from the virtual embodiment, mutually coordinated movements between participants and movements coordinated with the speech of others are more commonly found in our data. Consider Example 15 which transcribes the body movements of a number of participants while LF is introducing himself to the meeting. We transcribe on the spot turnings about with ∞----∞ the length and the position of the symbol corresponding to the analogous position in the talk above it. We transcribe translation movements (xyz-displacements) with ^-----^. A verbal description of each movement is given just below each movement-transcription. A period (.) at the beginning of a line is used to match up lines of transcribed movement where no body movement occurred with the corresponding line of talk.

It is notoriously difficult to adequately and clearly transcribe body movement [8, 11] but we hope our conventions will become clear as we now explicate this example. Early in the example, AC turns towards AB just while LF utters `fahlén`. He then stays facing AB for the rest of LF's turn. A 0.6 second pause follows, at the very start of which AC begins to turn back towards SB, continuing this movement over a brief `uh hum` from SB and stopping the movement when LF begins to hesitate (`er:`) in his next turn. For his part, reciprocally, AB turns towards AC, again beginning the movement at a hesitancy in LF's turn (the initial `uhm:`). Immediately once AC has finished turning back towards SB and started to move away from the group, AB also turns back towards SB, starting this movement during LF's `er:`. AB makes a further movement towards SB while (again) LF is hesitating and pausing with `er:m (1.2)`. SB then makes a movement towards LF which starts during a 0.6 second pause in LF's talk. Following the start of SB's movement, LF continues talking with `and um:: i've been involved in these things for a long time`. When LF begins to utter `for a long time`, SB reciprocates AB's slightly earlier turn towards SB, before finally returning towards LF, initiating this movement again during a one second pause in LF's talk.

Example 15

```
<SB is facing the others who are looking towards him. After
some side-talk with LF, SB invites LF to introduce himself>
LF:        (1.6) um: (2.0) i'm um (.) lennart fahlén from the
           swedish institute of computer science
AC:                                           ∞----∞

AC:        turns to face AB
```

```
         ( 0.6 )
AC:      ∞------
AC:      turns to face SB
SB:      uh hum=
AC:      -------
AC:      continues turning
LF:            =uhm: (1.0) and er: in my spare time i try:
         to manage this (    ) virtual reality group in
         here-at er:m (1.2) sics as it's called (0.6) and
         um:: i've been involved in these things for a long
         time (1.0) so i'm starting to (        ) (1.0)
         that's very apparent i think
AC:      ----------------------∞ ^-------------------------
         -----------------------------------------------
         -------------^ ∞------∞
         .
         .
         .
AC:      finishes turning to face SB then backs away from
         the rest of the group then turns towards AB
AB:               ∞---------∞       ∞-----------------------
         ---------∞
                     ∞--∞
         .
         .
         .
AB:      turns towards AC then returns to face SB then turns
         further towards the rest of the group
SB:      .
         .
                                                  ^-----
         -^                                      ∞--------
         ------------∞                            ∞---
         -------------------∞
SB:      moves towards LF then turns towards AB and then
         turns back towards LF
```

We suggest that this example on close examination compellingly shows *exchanges* of movement between first AC and AB and then AB and SB during LF's introduction of himself. To the extent that the blockies can exchange glances at each other, this is what they are doing here (even though for a blockie a 'glance' involves a whole body movement!). Thus, participants do seem to be able to use the blockies to enter into coordinated movements with one another. The blockies have an *interactional* significance and are not merely navigation devices. What is more these virtual body movements seem to be very precisely coordinated with LF's ongoing talk. It is in the pauses and hesitations in LF's talk that we see the majority of the movements initiated. Thus, strikingly, participants seem to be able to coordinate movements with each other while interleaving this with the talk of a 'third party'. It is also noticeable that SB's approach to LF initiated during a 0.6 second pause provokes LF to further add to his turn of self-introduction. In all these respects, we are witnessing residual though noticeable phenomena reminiscent of the coordination of action in ordinarily embodied conversation. In short, even if only to a limited extent, participants do seem to be able use the blockies in ways derivative from their ordinary interactional competencies.

DISCUSSION

We owe it to Heath et al. [12] to have drawn our attention to the following quote from a 1972 lecture of Harvey Sacks: "The technical apparatus [Sacks had in mind the telephone] is, then, being made at home with the rest of our world. And that's a thing that's routinely being done, and it's the source for the failures of technocratic dreams that if only we introduced some fantastic new communication machine the world will be transformed. What happens is that the object is made at home in the world that has whatever organisation it already has". We feel that our analyses of talk in virtual environments clearly exemplify Sacks' intuition that the new and technologically unfamiliar is often made to be 'at home' with our familiar world. If ever there were technocratic dreams, they have been thoroughly invested in VR! But what we see is the ordinary apparatus of conversation and the social interactional coordination of body movements being moulded and adapted to what the virtual environment affords so that participants can carry on as best they can with their business at hand.

Thus, we see systematic ways in which participants try to resolve or anticipate turn-taking problems, the elementary coordination of body movements between participants, the coordination of movements with ongoing speech, the utilisation of the bodies to engage others and initiate talk, amongst other phenomena. While, of course, we have concentrated on data from just one virtual meeting and it must be acknowledged that more meetings and more examples are required to develop yet more convincing generalisations, it is fascinating that what we have noted so far in the virtual world is in some way familiar.

Though familiar, this is not to say that improvements should not be made to systems such as MASSIVE to enhance their abilities to support multi-party collaborative activity in a virtual environment. Our interaction analytic techniques have been able to highlight problems, some of which may be amenable to technical solution or assistance. Let us discuss three classes of possibilities.

(1) We noted that it is possible for many technical failures to pass unnoticed for some time, simply because those moments which make them clear (a sufferer of technical failures being selected as next-to-speak, yet not responding) may not have arisen in the conduct of the meeting. This suggests to us that CVEs should support *local troubleshooting* because it may be very hard to bring to the attention of *others* that one is experiencing a local failure. This has implications for the overall distributed architecture that a system might exploit. An architecture must be not only robust in the face of local failure but it must also support graceful distributed recovery from local failures. A local failure must be remediable *at that site* and not require initiation from a remote site where participants may be unaware of the failure. Additionally, it should be possible to bring about such recovery by means within the expertise of any participant. These are actually quite demanding requirements and we believe they follow from our observations of how the structure of technically-mediated social interaction may lead to problems identifying ongoing failures in distributed systems such as MASSIVE.

(2) The overall design of virtual worlds should be considered in terms of how they afford social interaction and not just in terms, say, of their navigability, capability for presenting masses of information, or their thrilling aesthetics. The kinds of objects that we insert into a virtual world should be selected and designed with social interaction in mind. For example, a meeting table may be a simple device for people to gather around while affording them means for coordinating their talk, views of each other and mutual bodily orientation. Indeed, such a device may have aided our participants in solving some (not all) of

their turn-taking problems by suggesting a 'round the table' sequence for talk. Quite simple devices (e.g. a table as polygon on the base plane) may often be the most important from a social interactional standpoint yet their inclusion in the virtual world is easy to forget. Although the MASSIVE system can support a variety of 'meeting furniture', only a noticeboard was included on this occasion.

(3) While the blockies do afford various *social* interactional phenomena (and are not *merely* navigational aids or simple interface devices), it is worth reminding ourselves of the subtlety of interaction and participation which is much more readily possible with a real human body. The hands, the arms, the head, the neck, the torso permit a number of different orientations with respect to each other as well as with respect to co-interactants' bodies. In this way, we can glance without moving our heads, or turn our heads without moving the rest of our bodies.

Importantly, the coordinated flexibility of our eyes and heads enable us to look around without turning our backs on anyone. This, together with all the other kinds of embodied distinctions which are available for investing with interactional significance, is not available to the blockies. The only way for a blockie to 'glance' is by changing its whole bodily orientation. Accordingly, the blockies are considerably constrained in *just how* they can display their attentiveness to others and also in *just how* they can gesture or engage in whole body movements to aid the design of their own turns or to partake in a finely co-ordinated stream of talk. What perhaps is remarkable is that the minimal embodiments offer any interactional affordances at all. Nevertheless, introducing articulations (in the physical sense!) to the embodiments does seem to be worthwhile: e.g. so that a 'looking-around' can be distinguished from a 'turning-away'. Current VR systems essentially treat action in a virtual world as a matter of navigation or object manipulation. However, in CVEs where participants are interacting with one another, perhaps one should consider the direct support of actions of a 'higher-order' than mere movement, actions of *social interactional* significance (like approaches, turnings, glances and maybe some under collaborative control like 'form a circle' and so forth). In future work, we wish to study whether such higher-order actions can be sensibly added to the repertoire of interaction techniques available to the blockies and participants who 'inhabit' them. In this way, we may also be able to help users employ gestures to aid the concurrent construction of their turns - something which is currently problematic.

Of course, adding any further complexity to the blockies has to be reckoned with in the light of technical issues such as computational and network-transmission performance. We feel, though, that a viable and systematic research strategy for developing useful CVEs is to incrementally add further sophistication to very simple embodiments *as and when* analysis reveals that it is called for in the support of social interaction. Interestingly, this goes against the grain of many VR research trajectories which are devoted towards photorealistic body renderings and whole body movement detection. But, unless the social interactional significance of the body is understood, such developments may be not only unduly computationally expensive (especially when one considers distributed collaborative VR systems) but also lacking in social scientific motivation.

ACKNOWLEDGEMENTS

This research was supported by COMIC (Esprit BRA 6225) and VirtuOsi (DTI/EPSRC Initiative in CSCW).

REFERENCES

1. Benford, S., Bowers, J., Fahlén, L., Greenhalgh, C. & Snowdon, D. User Embodiment in Collaborative Virtual Environments, in *Proc. CHI'95*, ACM Press, 1995.

2. Benford, S., Bowers, J., Fahlén, L., Mariani, J. & Rodden, T. Supporting Co-operative Work in Virtual Environments. *The Computer Journal*, 37, 8 (1994), 653-668.

3. Bowers, J. & Churcher, J. Local and global structuring of computer mediated communication. In *Proc. CSCW '88*. ACM Press, 1988.

4. Bowers, J. & Pycock, J. Talking through design. In *Proc. CHI'94*, ACM Press, 1994.

5. Gaver, W. The Affordances of Media Spaces for Collaboration, in *Proc. CSCW'92*, ACM Press, 1992.

6. Goffman, E. *Forms of Talk.* Blackwell, Oxford, 1981.

7. Goodwin, C. *Conversational Organization.* Academic Press, New York, 1981.

8. Greatbatch, D., Luff, P., Heath, C. & Campion, P. Interpersonal communication and human computer interaction, *Interacting with Computers*, 5 (1993), 193-216.

9. Greenhalgh, C. and Benford, S. MASSIVE: A Virtual Reality System for Tele-conferencing, to appear in *ACM Transactions on Computer Human Interfaces.*

10. Heath, C. & Luff, P. Collaboration and control. *CSCW*, 1 (1992), 69-94.

11. Heath, C. *Speech and body movement in medical interaction.* CUP, Cambridge, 1985.

12. Heath, C., Luff, P. & Sellen, A. Reconsidering the virtual workplace, in *Proc. ECSCW'95*, Kluwer, 1995.

13. Presence. Special Issue on Simulator Sickness, *Presence*, 1.3 (1992).

14. Sacks, H., Schegloff, E. & Jefferson, G. A simplest systematics for the organization of turn-taking in conversation. *Language*, 50 (1974), 696-735.

15. Slater, M. & Usoh, M. Representations Systems, Perceptual Position and Presence in Immersive Virtual Environments. *Presence*, 2.3 (1994), 221-234.

16. Tang, J. & Issacs, E. Why do users like video? *CSCW*, 1 (1993), 163-196.