Creating Coherent Environments for Collaboration

Christian Heath, Paul Luff, Hideaki Kuzuoka, Keiichi Yamazaki, Shinya Oyama
King's College London, University of Tsukuba, University of Saitama
christian.heath} @kcl.ac.uk, paul.luff@kcl.ac.uk, kuzuoka@esys.tsukuba.ac.jp, yamakei@post.saitama-u.ac.jp, oyama@esys.tsukuba.ac.jp

Abstract. Drawing on studies of experimental systems and everyday settings we explore the relationship between social interaction and its environment We show how interaction is inextricably embedded within its environment and discuss the ways in which innovative systems designed to support remote collaboration inadvertently fracture conduct from its setting, and undermine the participants abilities to produce, interpret and coordinate activities. We discuss the implications of these issues for the design of future systems, and, more critically, for studies of work and collaboration.

Introduction

A long-standing problem for those involved in the design and development of systems to support remote, synchronous collaboration, is how to provide users with a coherent environment in which to accomplish action and interaction. It has proved difficult to ensure that participants have compatible views of their respective domains, or even that through interaction they are able to establish, for all practical purposes, common standpoints to enable cooperative activities to be coordinated and accomplished These problems become increasingly exacerbated if one attempts to provide participants with resources with which to undertake seemingly simple actions such as pointing, reference, or manipulating objects and artefacts within the remote location

In this paper, we wish to explore some generic problems which arise in creating coherent environments for distributed collaborative work. In particular, we examine the use of a technology that has been developed in order to enhance the ways individuals interact with each other, orient to and manipulate objects in a video-mediated environment called GESTUREMAN (Kuzuoka, et al., 2000). Through a combination of technologies, including a mobile robot, wireless video communication and a remotely controlled laser pointer, GESTUREMAN is meant to provide resources for a remote participant to refer, point to and distinguish objects from each other in a remote domain. By examining how it is used by participants performing a quasi-naturalistic task we see how they make use of these resources. We consider the use of GESTUREMAN in the light of other systems aimed at enhancing distributed collaboration and participation, particularly recent attempts at augmenting video-mediated environments and collaborative virtual environments (CVEs) As with other systems which endeavour to provide a coherent environment for remote collaboration, we suggest that this system, though in quite a different way, fractures the environment(s) of action, and inadvertently undermines the participants' ability to produce, interpret and co-ordinate their actions. We then draw upon examples from everyday settings to show the ways in which actions and social interactions are inextricably embedded in their immediate environment. In particular, we consider how both the production of action and its recognition by others are dependent on relevant and occasioned features of the setting. We conclude by discussing the implications of the analysis for creating coherent environments for collaboration, and for our understanding of collaborative action and interpersonal communication.

Embodying Actions in Remote Domains

It has long been suggested that system designers might benefit from a more thorough understanding of what is known about groups and interpersonal communication. Unfortunately however, the dominant models and theories of interpretation and communication still tend to emphasise talk and non-verbal communication and largely disregard the ways in which action is embedded within its immediate environment In recent years, there has been a growing emphasis on both the body and embodiment (e.g. Robertson, 1997), and on objects and material artefacts (e.g. Latour, 1992), and yet to a large extent the ways in which communication and interaction are embedded in, and constitute, 'material realities' remain under-explored (despite important exceptions, e g. Goodwin, 1995; Hutchins, 1995).

In part, the predominant theories of communication which pervade the social and cognitive sciences may well have influenced the 'face-to-face model' which still pervades technical developments in CSCW and more generally computer systems

designed to support synchronous, remote collaboration (Heath, et al., 1995) Whilst capabilities have been introduced to allow participants to refer and point to objects, in particular documents, the principle emphasis on both basic and complex systems, is to provide remote participants with face-to-face views of each other and to largely disregard the local environments in which conduct is produced. We believe that this tendency has been unfortunate for CSCW, since collaborative work, and more generally social interaction, is largely accomplished in and through objects and artefacts, tools and technologies – be they documents, diagrams, models, on paper, screen, through keyboards, pen, mouse and the like. These material resources not only feature in how people produce actions, but also, and critically, in the ways in which they recognise or make sense of the actions of others.

In this regard, for some years we have been involved in successive projects concerned with developing support for synchronous, remote collaboration. These projects are primarily concerned with exploring and developing ways in which we can provide remote participants with the ability to invoke, use, discuss, and manipulate, objects and artefacts. These projects emerged from our earlier studies of the disembodied character of conventional media space (Heath and Luff, 1992b) and led to the Multiple Target Video (MTV) experiments (with Gaver and Sellen) (Heath, et al., 1995) and more recently to develop very different 'solutions' through CVEs (Hindmarsh, et al., 1998). As we have discussed elsewhere, neither expanding media space to include features of the remote participants' environments, nor CVEs in which a seemingly compatible, pseudo-worldly environment is provided to the participants, provides satisfactory support for 'object-focused' collaboration. Indeed, in both solutions and their various iterations, participants encounter difficulties in making sense of each other's conduct even when undertaking seemingly simple actions such as pointing to objects within a particular environment.

In this paper we wish to address the more generic difficulties which arise in building coherent environments to support remote collaboration. To do this we will discuss our recent attempts to build technologies to support object focused collaboration amongst remote participants, and continue by discussing some instances drawn from everyday settings.

These concerns are also reflected by a series of technical innovations developed in a collaboration between engineers and social scientists at the Universities of Tsukuba and Saitama in Japan. In these, a number of systems to support remote collaboration with and around objects and artefacts have been developed; for example, extending media spaces to include cameras which can move when operated by remote participants and laser pointers which can indicate locations of interest in a remote space (Kuzuoka, et al., 1994; Yamazaki, et al., 1999). These techniques have been deployed in a series of evolving systems: GESTURECAM, GESTURELASER and GESTUREMAN. Such developments provide an invaluable opportunity for considering how to establish coherent distributed environments in which to undertake collaborative activities.

GESTUREMAN, for example, is the latest of these developments. It was constructed and configured at the University of Tsukuba and consists of a small robot that can move around a remote domain (see Fig. 1). In different configurations it has 2 or 3 cameras and a laser pointer. Images from these cameras and commands to the robot are transmitted through a high bandwidth wireless channel so that the remote participant is provided with high quality, real-time video.

In the remote domain a participant sits in front of 2 projection screens displaying the images from the cameras (see Fig. 2) The participant moves the robot by use of a joystick. The laser pointer is controlled by a mouse and pressing the mouse button makes the laser dot brighter. This is a way of distinguishing movement of the laser dot from pointing with it. The participants speak to each other via wireless headsets and microphones.

Fig. 1: The GESTUREMAN with a 'local' participant

In order to examine GESTUREMAN we undertook a variation of the FurnitureWorld Task which has been given to participants using the MTV configurations of video spaces and CVEs (Heath, et al., 1995; Hindmarsh, et al., 1998). In each case the task though similar had to be re-configured to match the technology under investigation. For the

Fig. 2: the 'remote' participant The screen in front displays the image from the right camera, the display on the left is from the left camera

GESTUREMAN task, one of the participants (the 'local' participant) was in a room with real furniture. We carried out the task with 5 pairs (2 Japanese and 3 English speaking pairs) and collected data for each from 6 cameras. The objects in FurnitureWorld included examples of both Japanese and Western furniture.

The task provided us with data where the participants: referred and pointed to different objects; distinguished different objects for each other; were mobile around the domain and could move objects around the space; and allowed them to make different uses of the space. It also provided us with materials with which to assess the different capabilities of the system for pointing and referring to objects with respect to media spaces and CVEs, as well as allowing us to collect some novel data which could inform our understanding of the interactions with tangible artefacts and in mixed realities.

Locating an object for another

The laser pointer of GESTUREMAN did provide a way for the remote participant to identify objects and distinguish locations for the local participant. In the following fragment Dave (the local participant in the room) and Candy (the remote participant) are discussing the space on the left hand side of the room and how to lay out the furniture. Candy has positioned the robot in the doorway of the room and it and Dave are oriented to one corner on the left hand side. In the course of this discussion Candy utters 'let me>let me look over here' and moves the robot so that it is oriented to the right hand side.

```
C· let me>let me look over here to see how much room
      (.)
C:  oh theres not so much room over here↓-------
D                                                where where you looking –
```


As the right hand side of the room (behind Dave) comes into view she says 'oh theres not so much room over here↓'. Whilst Candy is speaking Dave maintains his orientation to the left hand side but glances back and forth to the robot and then asks 'where, where are you looking'.

```
C  I'm (.) looking over to your>behind you theres you ca n(t)        yep there see>can you
D                                              over over here
```


Whilst Candy responds with 'I'm (.) looking over to your. .' Dave continues looking in front and even takes a small step forward. As she starts to reformulate her description of where she is looking ('>behind you theres') Candy moves her hand to the mouse and starts to manipulate the laser pointer. Only as Candy utters 'theres' does Doug spin around. On doing this he finds the faint laser dot on the wall on the right hand side and points to it down to the left whilst uttering 'over over there'. Candy replies with 'yep there see>can you see that little red light thing' whilst brightening the dot with the mouse pointer and moving it around the wall.

C   see that little red light thing=   theres not really any room over there------so
                                      =yeah yeah-

Dave follows the red dot on the wall and Candy utters 'theres not really any room over there'. They then reorient back to the left hand side of the room.

In this simple instance the remote participant uses the system to identify a location in the physical space, a location related to the matter the co-interacts are discussing. She manages to get Dave to re-orient to the location that she is talking about by moving the laser dot, brightening it and then moving it around the space.

The accomplishment of this shift in orientation is not entirely unproblematic. The talk through which it is accomplished involves various perturbations, restarts and reformulations. Candy has some difficulty identifying the location for Dave, it being 'over here', 'over to your...' and 'behind you'. In part, this is because Dave is also engaged in trying to identify the location, and even when standing still, he shifts the direction of his glance. Such utterances have therefore to be coordinated with the shifting orientation and participation of the co-interactant.

For Dave, being co-located with a little robot also presents various difficulties Even though the cameras are visible and it is possible to see (and hear) the robot moving, it is not clear where the remote participant is looking. Some sense of this can be obtained by glancing at the orientation of the robot in relation to the environment, as Dave does in the fragment above. But, as for the remote participant, this can be problematic if the robot is itself moving. Hence it is difficult to coordinate one's own activities with another when mediated by the robot and its pointing devices. However, the participants do manage to orient to a common object or location. In this and other instances, they do this by talking about the pointing, 'the laser light', the 'little dot' or the 'red thing', thus the pointing often becomes a topic of the discussion rather than a resource for collaboration In this instance what initially seems to be a passing remark about another area of the room ('theres not so much room over here') requires a number of activities before they can achieve a common orientation to that passing location ('theres not really any room over there') and then continue their discussion about the layout in the original area.

These difficulties and their solution resonate with conduct found in other technologically mediated environments, such as media spaces and conventional collaborative virtual environments. The laser pointer is a way of relating the talk and activities of a remote participant to the ongoing activities and environment of another, much as the extending arm does on the avatar in Fraser's augmentation of the MASSIVE CVE (see Hindmarsh, et al., 1998). In the case of GESTUREMAN the pointer can be used to mark particular locations and general areas in a physical space

and, in the above instance and others, the participants do use the device in this way. However, as in a CVE a device to support pointing and achieving a common orientation with another itself becomes a subject of explicit talk and interaction and the operation of the device requires work on behalf of the participants to get the pointing done.

Securing a common orientation to an object

It is not only the remote participant who may face difficulties in attempting to have the other discover an object within their domain. The individual who is with GESTUREMAN may also have to undertake successive actions in order to have the robot, and the remote participant, locate a relevant object. The individual located in the furnished room has limited resources with which to identify the orientation and perspective of the other and in consequence may have to make various attempts to have an object seen and recognised. Consider the following fragment. As she tidies the books on the table, Jemima says 'we have a (.) we have a (.) we have a bookcase in he:re↑'. She then starts to locate the bookcase for Darren who is operating the robot. As she says 'so maybe I should (0.3) if you turn over that way↑ (0 3) to the left↑' she points in front of the cameras on the top of the robot

J: so maybe I should---if you turn over that way↑ --- to the left↑ - yeah↑

Jemima's initial pointing gesture first appears for Darren in the centre screen, but as she says 'turn' in 'you turn over that way↑' her hand moves into the left screen, the gesture thus occupying most of the two large screens in front of Darren. Jemima then pauses briefly and brings her arm back slightly. She then extends her arm fully again and reformulates the location 'to the left↑'. Darren then starts to move the robot to the left and as he starts this movement Jemima utters 'yeah↑'. As Darren spins the robot around, Jemima continues to hold out her hand saying 'do do you see it↑(0.3) do you see the book shelf'. Her hand now appearing just on the centre screen.

J   do do you see it↑--- do you see the book shelf------   yeah yes alright
D   oh right here - okay↓okay↓

Only when Darren appears to identify the bookcase with 'oh right here (.) okay↓okay↓' does Jemima withdraw her hand and start to discuss where to put the bookcase, moving around the robot as she does so.

Jemima manages to get Darren, through his use of the robot, to see the bookcase in the far corner of the room. She does this by accompanying her spoken directions with an extended pointing gesture. But this is not a straightforward gesture. Jemima reshapes it in the course of its production and through its animation successfully secures the appropriate reorientation from Darren (mediated by the robot). This initial movement of the robot is accompanied by a 'yeah' from Jemima, but whilst the robot is in motion Jemima continues to point towards the bookcase. In all she maintains this point for some 6 seconds and tries to secure some further response from Darren that he can see the target of the point ('do do you see it↑ (0.3) do you see the book shelf'). Only when she says 'oh right here (.) okay↓okay↓' does Jemima withdraw her hand and her talk about the bookcase can continue. The robot does provide resources for Jemima to make sense of Darren's participation in the ongoing event. From its orientation it is possible to have some sense of the 'field of view' of the remote participant and by its movement whether the remote participation is transforming his or her participation in an ongoing activity, that a particular trajectory has commenced. However, these capabilities provide a limited and fragmentary set of resources. Jemima cannot get a sense of how her conduct is seen by the other and the nature of the movement once again foregrounds the means by which a common orientation is accomplished rather than the object and locale towards which the orientation is being established.

Making sense of disembodied gestures

In the following fragment Candy (the remote participant) has positioned the robot in the doorway angled so it can see into the corridor outside the room with the left hand camera and into the room with the right She is trying to identify the object that is underneath some cushions and asks Dave: 'is that another table under those cushions'. As she says this

Dave is standing in the corridor just visible on the left hand side of the left image. Dave pauses and then points into the room as he says 'ah:::::: (0.6) you mean right right here↑'.
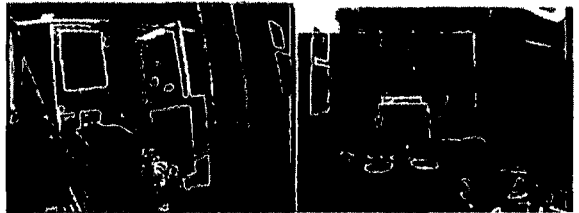
D   ah · ·------you mean right
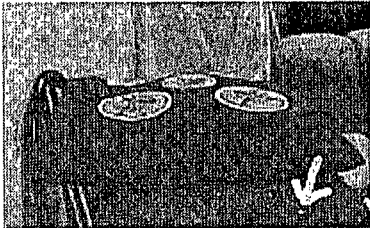   right here↑--

Unfortunately, there are a number of cushions in the room and the corridor and the laser light just happens to have been positioned near to a cushion that has been placed on top of another and both are on top of a chair in the room (see arrow). Dave takes the location of the laser light·as relevant to Candy's talk and draws on this to locate what Candy is referring to and points to those cushions. Candy then moves the laser light accompanied by.some perturbations in her talk· 'huh>errm--like oh'.

C huh>errm--like oh   =errm no I'm not over there yet hold on------err
D: oh yeah right there↓=

In the first of the mouse movements the pointer comes temporarily to rest on the bottom cushion. Dave again takes this as relevant and says 'oh yeah right there↓='

Candy then accompanies her next movements of the laser pointer along the wall and out into the corridor with an explicit account of what she is doing with the laser pointer '=errm no I'm not over there yet hold on'. Only when the laser light comes to rest near the cushions on the chairs in the corridor can Candy ask 'out here in the hallway↑ these things?' to which Dave replies 'errm no it is another one of these chairs↓ we've got a whole mess of these (0.2) these black chairs↑'.
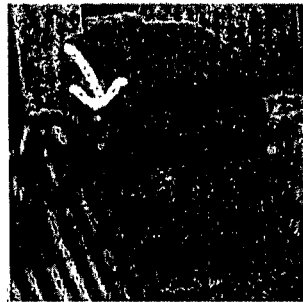
The laser pointer does provide a resource for a remote participant to locate objects and locations in the environment of another. It is invaluable in relating one's own actions to features in the remote environment. This simple mechanism, a visible dot operated with a mouse, allows remote participants to tie their talk and actions to particular objects and locales. It also provides individuals in one locale with a way of identifying what a participant located elsewhere is referring to and thus enables them to establish a common orientation to objects and places in the environment.

The laser does however have shortcomings. Perhaps most critically, unlike a gesture, the light of the laser lies on the surface of the landscape of the remote environment. So for example, as in the instance above, the participant in the

furnished room may assume a relationship between the laser light and the actions of the remote participant, even when it is not relevant, leading the participants on occasions to have to address these misunderstandings. Lying on the surface, the laser light inevitably has a potential relationship with objects in the world, and can lead participants to assume a connection where no relevant relationship exists Moreover, unlike gestures which can be shaped in the course of their production, and the co-participant's conduct, the laser light lies largely undifferentiated in the furnished domain. What then appears to be a obvious solution to relating talk to particular objects in a remote environment, by providing a simple tie between the activities of a remote participant and the local environment of another, renders activities such as pointing highly problematic and invokes potential relationships where none are relevant.

As mentioned above, in previous studies of media spaces, enhanced video-mediated interaction and CVEs, participants of these quite different technologies appear to face parallel difficulties. Despite efforts to provide 'common spaces', 'symmetric environments' or resources for pointing and reference, these technologies can be seen to inadvertently fracture the relationship between conduct and the environment in which it is produced and understood. Ironically, the more we attempt to enhance the environment to provide participants with a range of abilities and resources, the more we may exacerbate difficulties in the production and coordination of action.

Given these difficulties faced by individuals in these technologically-mediated environments, perhaps it is worth briefly considering the resources on which participants rely to coordinate their actions in more conventional settings, and in particular the ways in which interaction is inextricably embedded within the immediate environment. We will draw examples from various settings including control rooms, medical consultations and museums. Considering examples drawn from everyday environments illustrates the ways in which the production and intelligibility of conduct is dependent upon the participants' mutual access to the immediate environment and can suggest why the use of a simple remote-controlled pointing device may lead to difficulties.

# Embedded Action in Everyday Settings

Invoking the environment

As part of their daily activities personnel frequently refer to objects and artefacts within their immediate environment. These objects and artefacts may range from diagrams in paper documents, through to complex multimedia displays, they can be presented on screens or projected onto the walls of the work setting. To a large extent these occasioned references to features of the local environment, even where

they themselves are changing or in flux, rarely engender serious difficulties or problems. The smooth and unproblematic ways in which participants invoke, discover, discuss, and respond to features of the immediate environment, becomes more impressive when one considers the complexity of contemporary workplaces; environments such as control centres are littered with an array of objects and artefacts, in different media, with variable rates and quality of change.

Consider the following fragment in which one participant simply asks another where a particular object is to be found within the immediate environment. The fragment is drawn from a control centre on London Underground in which participants are provided with numerous tools and technologies to oversee traffic and passenger movement. The setting, a site of previous research (Heath and Luff 1992a)is currently being re-studied as part of a project concerned with the design and deployment of image recognition systems. We join the action as John receives a phone call asking him where the train number 225, is currently positioned.

    J.   Hello there (3 2) Two Two Fi_ve Let me have a look for it (0.4)
 → J   Two Two Fi ve : ? ( ) He's around here somewhere
    G   Two Two Five s at er (0 6) It's up there?
            (0 3)
    J   Oh he's in the shed<he's in Queen's Park ( ) °hh Sou th Sheds

Whilst John utters 'Two Two Fi:ve:::? (.) He's around here somewhere', into the handset, he simultaneously looks at various screens along the console. During the utterance, John turns and looks at the monitor to the left of Graham; a monitor which shows the timing of the trains over the section of line for which Graham is responsible (south of Piccadilly Circus) As John turns towards the monitor to his left, Graham follows his
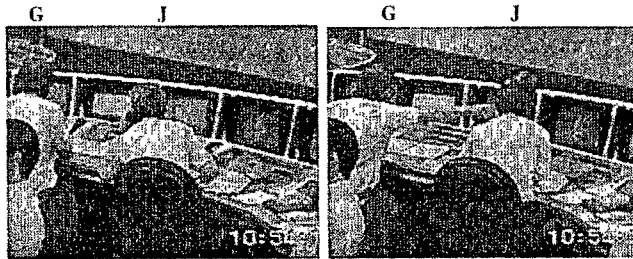
J Two Two Fi_ve me have a look   G   It's up there?

gaze and they look at the screen together. John has not only encouraged Graham to help him look for something that moments ago was not explicitly relevant, but encouraged him to look at, and momentarily inspect, the information on a specific screen, in a certain way. As John turns away, failing to find the train, and uttering the word 'around', Graham continues the search by first looking at the fixed line diagram and then to the monitor directly in front of John. Uttering 'Two Two Five:s at er: (0.6) It's up there', Graham points to the image on the screen directly in front of John, and at which John is now looking.

In pointing to the object, Graham is sensitive to where John has already looked and to his current orientation. The point is designed to contrast the particular monitor with the others, and to enable John to follow the trajectory of the pointing finger to the screen which displays a complex array of trains and traffic information.

As John turns towards the screen, Graham moves the pointing finger closer to the monitor to enable John to discriminate the information displayed on the screen. The focal object is made progressively and increasingly visible, with regard to the ways in which Graham aligns his gesture and talk with the shifting orientation of his colleague. John is able to locate the object and responds with 'Oh he's in the shed<he's in Queen's Park (.) °hh Sou:th Sheds:' John finds the object in question, train 225, by virtue of Graham's search and point. He informs the caller where the train is, and simultaneously displays to Graham that he has both found the object and located its 'exact' position.

Simply encouraging a co-participant to look at and discover a particular object within a common domain therefore is a complex and emerging event. It is sensitive to the reason for finding the object, the orientation of the co-participant, the preceding actions, which in this case visibly fail to locate the object. The gesture and the accompanying talk is highly sensitive to the individual's standpoint *vis-a-vis* the object and the co-participant. As the gesture emerges it is fashioned with regard to the shifting orientation of the other, initially demarcating the region in which the object may be found, and then more specifically, the actual location of the object with regard to other objects within the surrounding area Without access to the co-participant's immediately preceding conduct, and the relationship of that conduct to its surrounding environment, and without access to the recipient's conduct as he is beginning to discover the referent, it would be difficult for the person pointing to produce the activity in a relevant way. Just as the recipient relies upon his access to the other, and the other's progressively emerging actions *vis-a-vis* visible features of the ecology to determine the sense and intelligibility of the action, in a nutshell, to 'get the point'

One can see why participants may encounter difficulties in referring and pointing to objects using a laser. For the remote participant, it is difficult to determine the orientation of the other with regard to the furnished world and to be able to keep track of how the shifting orientation is sensitive to that world and the actions of the other. More critically perhaps, as the remote participants refer and point to objects, they are often unaware of how the other's conduct emerges during the course of that action, so that they are unable to reshape their talk and pointing with regard to the emerging actions of the recipient. For the individual in the furnished room, it is often difficult to determine the orientation of the remote participant (for whom their referent is designed), more difficult still to determine how the other sees them in the furnished world. It is not surprising therefore that even relatively simple actions such as reference and pointing involve the participants in relatively extended sequences of action though which they render visible their own orientation and attempt to reconcile their respective views on the world.

Entailing action and animating objects

Encouraging co-participants to notice or examine something within the immediate environment is often accomplished in a less explicit and obtrusive fashion. People happen to find themselves looking at particular documents or noticing screen changes, and in consequence dealing with events which might otherwise have passed unnoticed. Colleagues encourage each other to notice particular phenomena or events, and as we and others have discussed elsewhere are highly sensitive to the concurrent activities of colleagues even though they may engaged distinct and seemingly unrelated tasks (e.g. Heath and Luff, 1992a).

In control centres, as in other environments, the participants abilities to maintain a sense of what others are doing, and to encourage them to notice changes that might otherwise pass unnoticed, derives in part from the familiarity with the local environment. Despite the seeming complexity of the settings, displays, tools and artefacts remain in stable locations. Particular phenomena, such as alarms, incoming information and other events are revealed by screens or documents that remain in the same relative location, such that particular types of activity rely on particular tools and technologies (the use of which may well be visible to others in the same domain). In consequence, the use of a particular keyboard or document, may provide resources to enable people both to discover and make sense of what others are doing within the immediate environment, and form the basis to coordination and collaboration In a sense therefore, despite their complexity, and the changing character of materials on-screen and the like, the immediate environment provides resources to enable people to discover or make sense of the actions in which others may be engaged.

For the matters at hand, it is important to note that participants have a complex array of resources for having others notice, and react, to actions, objects and the like which may happen to arise within a particular setting. Pointing and showing are the most explicit way of establishing co-orientation to some potentially relevant feature within the local milieu. In interaction and collaboration individuals are highly sensitive to the ways in which co-participants may notice, and encourage others to notice, objects and events within the local milieu. Where the critical aspects of such noticings is for the other to examine and determine for themselves the potential relevance and significance of the object in question In a sense therefore participants embed potential action within the local milieu, and rely upon others to make sense of those noticings by virtue of the ability to scrutinise the environment for potentially relevant objects and events. In GESTUREMAN, as in others forms of media space, these noticings, embedded actions, largely pass unnoticed, since the remote participant is unable to 'connect' the participant's action to the relevant feature of the domain. Conduct becomes disembodied.

In many cases, participants do not simply have someone look or notice something within the local milieu, but fashion, or even transform the object or artefact. In turn, these transformations serve to encourage particular actions, facilitate

certain experiences, and form a foundation to subsequent collaboration. It is worthwhile considering an example from a rather different type of setting, in this case a museum. It is drawn from the Materials Gallery in the Science Museum London and in particular an exhibit which is designed to provide visitors with a sense of the form and consistency of different types of liquid. This exhibit consists of a large glass structure which contains a series of tubes which hold different forms of liquid. At each corner of the exhibit is a pole. On the top of three of the poles are buttons which when pressed cause the liquids to rise. As a young boy (B) begins to press one of the buttons he is joined by his sister (G). As she joins him, he tells her 'I'm making it go up'. She turns and looks at the relevant tube in the exhibit. As she turns, her brother begins to gesture upwards in line with the rising liquid and at the same time goes 'Oooh::::::::::↑' rising to a crescendo, and as the liquid falls continues with 'Oooh::::::↓' and allows the gesture to fall downwards. His sister, watching the liquid, and trying to wrestle the button off her younger brother, delivers an appreciation of the liquid's motion by producing a falling sound 'Mmmmmm↓'. The boy's gestures and accompanying sound do not simply draw attention to part of the exhibit, but exaggerate the liquid's movement, transforming it momentarily into a spectacle, a moments drama that the children experience and appreciate together A feature of the environment is given a flavour and significance through the ways in which their actions reflexively animate its movement and function.

The ability of the young boy to invoke and animate features of the rise and fall of the liquid, relies upon his access to his sister and her emerging orientation towards the exhibit. Once again we can see how the shifting relationship between a participant and (relevant) features of the ecology is critical to the ways in which an individual is able to highlight an object and engender particular forms of experience. Moreover, the gesture and the accompanying sounds are shaped, within the course of their articulation, not only with regard to the object, the rise and fall of the liquid, but to the sister's emerging orientation towards the exhibit The exhibit's sense and significance therefore, at this moment within the interaction, arises through the moment-to-moment co-participation of the children, the exhibit itself is critical to the production, intelligibility and coordination of their conduct.

In different ways therefore, we can see how the participants' actions give particular sense and significance to features of the immediate environment, and how various objects and artefacts provide resources for the production and intelligibility of conduct and interaction. The participants do not simply draw attention to features of the environment, but entail actions in particular objects or artefacts In these cases we can see how participants respond to the object or the artefact, rather than the

activities of the other, even though the object has been given its determinate and occasioned sense through the other's conduct. Action is *transposed* and embedded within the immediate environment; the participant's talk and gestures, their interaction and collaboration are inseparable from particular objects and artefacts, and the ways in which they, at some particular moment, are constituted as relevant. The reflexive relationship between action and the environment is a critical feature of the participants' conduct and collaboration.

Emerging material relevancies in the environment

The relevance of the local environment to action and interaction is not limited however to occasions in which the participants refer to or invoke a particular object or artefact. The local environment pervades the ways in which people make sense of each other's conduct and produce their own actions and activities. The environment provides resources for rendering the actions of others intelligible or sensible. Without access to the environment in which the actions and activities are produced the relevant sense of the conduct can be irretrievable. So for example, someone's activities like standing and walking across a room is intelligible by virtue of the fact that you are familiar with the setting and can interconnect the movement with the physical features of the environment. These actions are intelligible through their occurrence within a commonly accessible and available world. To. consider these actions independently of their local environment would be to undermine the sense of significance for the participants and of course to disregard the resources which feature in their production.

For example, consider the following instance from the beginning of a medical consultation in general practice.
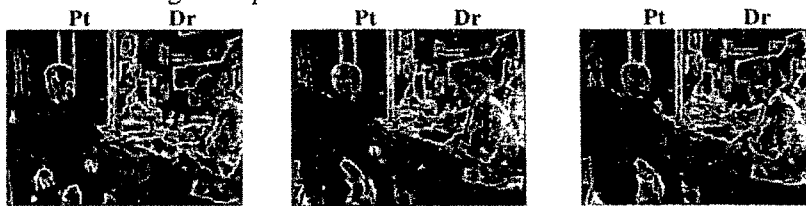
Dr   What's up?   (4.2)   Pt   I've had a bad eye   in there

The patient enters the room and walks towards the chair to one side of the doctor's desk. The doctor is glancing at the medical record and as the patient crosses the room, the doctor utters 'what's up'; a euphemism for 'what can I do for you'. The patient withholds a response for four seconds or so and then, pointing to his eye, replies with 'I have got a bad eye::: in there '

. Just as the doctor times his query to the patient's pattern of movement and prospective arrival at the chair, so the patient is sensitive to the doctor's conduct in producing his response. The doctor's use of the record, his shifting gaze up and down the page provides resources not only to account for his seeming disinterest in the patient, but in anticipating the upcoming completion of the activity. At one point,

soon after the patient sits down, the doctor smoothes the page and takes hold of a date stamp; the patient then glances up, opens his mouth, and finding his co-participant once more looking back towards the document, licks his lips, looks away and closes his mouth. A second or so later, the doctor turns from the record to the patient, and patient begins to speak, pointing to the blister on his eye.

The sense and significance of the doctor's conduct does not simply derive from his visual orientation, but rather from the emerging relationship between his bodily orientation and features of the immediate physical environment. The doctor's delay in aligning towards the patient is legitimised by virtue of the record's use, and his manipulation and looking at the document provides the patient with the resources to make sense of the conduct and attempt to coordinate his action. The bodily communication, non-verbal behaviour, of the doctor, is only intelligible for the patient (and of course the researcher) by virtue of how he can interweave the document, the object, with the participants' conduct; just as the patient's conduct for the doctor is recoverable by virtue of his own looking at and manipulating the record. To dissociate the talk and bodily conduct of the participants from relevant features of the environment, renders the action 'absurd', that is 'inharmonious; out of harmony with reason or propriety'. Surprisingly however, research on interpersonal communication, some of which has had an important influence on systems to support synchronous distributed working, disregards conduct and interaction from the environment in which it is produced and made sensible.

Discussion: embedded action

In the light of the materials presented here and a range of related studies of the workplace (e.g. Harper, et al., 1989; Suchman, 1996), we can begin to see ways in which the local environment features in the production, intelligibility and coordination of action and interaction. It is perhaps worthwhile outlining a few of these issues·

- participants reference features of the local environment, and encourage others to inspect, examine and look at objects and artefacts; the orientation and standpoint of the co-participant(s) and their emerging action and activity is critical to reference, and the discovery and perception of the objects;
- participants animate, and in other ways transform features of the environment for another, through gesture, touch and talk; the interplay of action and the environment giving the occasioned sense and significance of an object;
- actions can be entailed or embedded in particular features of the local environment; the object can serve to encourage and engender particular action and activities and serve as a 'centre of coordination';
- participants 'read' or make sense of the actions of others through the ways in which they interweave conduct with particular features of the immediate

environment; the embeddedness of action in the environment allows participants to discover why and what others are doing;

• the environment consists of a constellation of objects, artefacts, tools and the like; participants make sense of the actions of others, and produce their own actions with regard to each others access to and perspective on the 'occasioned' configuration of features.

The interdependence of action and the environment is critical to the ways in which participants are able to make sense of, and coordinate their actions with others. These foci are continually emerging moment-by-moment within the course of the participants' action and interaction. They are reflexively constituted in and through the participants' conduct.

The relationship between action, coordination and the local environment has profound relevance for our attempts to support synchronous collaboration amongst individuals based in different locations. In developing systems to support remote collaboration we attempt to interweave and/or create new environments, in which participants can produce actions within a setting which, in part, is accessible and intelligible to each other However, once we begin to create new environments to enable people to interact and collaborate with each other, we fracture the relationship between action and the relevant environment, and thereby engender difficulties, which may render even the most seemingly simple form of activity problematic.

Media spaces and CVEs vividly demonstrate the significance of environment to action and point to the shortcomings of certain approaches to interpersonal communication, language use and social interaction. They powerfully demonstrate how the very intelligibility of another's conduct relies upon participants' abilities to interweave environmental features with occasioned courses of action, just those features that are reflexively constituted in and through the participants' conduct. They reveal, for example, how the most seemingly trivial action, such as pointing to an object, involves the complex and occasioned interdependence of the participants' bodies, bodily orientation, bodily conduct, talk, and visible and material properties of a scene. They demonstrate how spatial and temporal features of conduct are so easily ruptured when attempting to interweave different scenes, scenes which fracture the production and receipt of action, and environments in which conduct is 'placed'. Small scale, naturalistic experiments with a number of technologies reveal how even such an apparently straightforward activity, like a point, can be problematic when access to particular features of the environment is either unavailable or transformed in some way. The experiments with GESTUREMAN, for example, reveal how a technology can be used for such an activity, but how it fractures the participants' conduct from its setting, and hence reveals the work required in getting such apparently simple activities done.

The significance of such small scale experiments can lie beyond the observations concerning a particular system or technology. Whilst not rigorous or generative as in a conventional experiment, they provide resources for what Garfinkel (1967),

borrowing from Spiegelberg, once characterised as 'aids to sluggish imagination'. By placing participants within unfamiliar environments and asking them to undertake seemingly simple activities, the experiments can serve to powerfully illustrate the resources on which they ordinarily rely in the practical accomplishment of action and interaction in more conventional environments Garfinkel suggests that "I have found that they produce reflections through which the strangeness of an obstinately familiar world can be detected" (p. 38).

These naturalistic experiments provide important resources to recognise the importance of a range of phenomena, competencies, resources and the like, which to some extent have been treated as epiphenomenal in social science research. In turn, we believe that such resources are critical to both the understanding and design of systems to support distributed collaboration. The development of experimental systems in CSCW to support synchronous collaborative working, throw into relief not only how little we know about the ways in which action and interaction is ecologically dependent and embedded, but how predominant orientations within the disciplines largely direct analytic attention away from the material, from the object, and in particular the reflexive constitution of the physical world in human conduct. Those concerned with developing technologies in CSCW have every right to question whether contemporary social science delivers on just the issues that they face when building systems to support synchronous, distributed collaborative work.

We are currently engaged in exploring ways in which the observations and findings discussed here can be drawn upon to inform the design and development of future systems to support synchronous collaborative work, particularly amongst distributed participants. The interdependence of action and environment suggests a number of concerns and considerations which, in the light of social and technical research, we believe are relevant to the design and development of future systems For example, consideration could be given to how to:

- provide participants with the ability to determine the location, orientation and frame of reference of others ;
- provide resources for participants to determine their standpoint with regard to other participants and the space(s) in which they and others are located,
- provide resources through which participants can discriminate the actions of others which involve shifts in orientation and reference to the space and a range of objects, artefacts and features;
- consider ways in which participants can refer to, invoke, grasp, manipulate, address, discuss, and in various ways animate properties of the space, and coordinate such actions with the real-time conduct of others;
- consider how participants can be provided with, and themselves preserve a stable constellation of relevant objects, artefacts and scenes within the space(s), so that they can produce and interpret actions and activities with respect to a presupposed coherent and stable environment.

As well as being of relevance to designers endeavouring to build new environments rendered through video, computer graphics and projection technologies, the issues raised in this paper may also be of interest to those developing collaborative systems by other means, whether these are standard shared applications presented on conventional displays or enhanced forms of remote collaboration supported by novel techniques. Considering how actions are embedded in the environment both in the production and for the intelligibility of actions may suggest a rethinking of the resources which co-participants can utilise to establish common frames of reference. Indeed, the problem of providing adequate support for users to determine common standpoints, foci or points of reference has been an incessant problem for designers of collaborative technologies (e.g. Tatar, et al., 1991). This may in part be due to a limited conception of how such activities are accomplished with respect to the local environment and the emerging and ongoing actions of co-participants. The support for establishing a 'common referent', for example, being principally considered with regard to the individual undertaking the action and the particular object of concern, rather than with respect to the ongoing activities of other participants whilst it is being accomplished and how features of the environment are utilised in its production.

These considerations pose complex problems for those of us involved in the design and development of systems to support distributed collaborative work, more complex than merely providing participants with the ability to establish common foci and points of reference. Through a consideration of the use of GESTUREMAN and of more mundane everyday examples we can begin to see that providing the participants with the ability to point at and refer to particular (shared) objects within a domain is one aspect of a more fundamental and critical problem. The problem as we have demonstrated is not simply how people can detect and identify particular objects, but rather how they can establish and maintain a relevant 'connection', 'relationship' between the co-participant (even an avatar) and the environment in which that person (or representation) is located. The reflexive relationship between conduct and ecology, poses a critical problem both for the participants, in producing, making sense of and coordinating their conduct, and for those of us engaged in the design and development of novel collaborative environments. In this regard, we have recently begun a new programme of work in which we are attempting to develop media spaces in which participants have mutually compatible access to both each other and the respective domains and ecologies of objects and artefacts This programme of work also involves extensive studies of more conventional environments in order to begin to understand a little more about the ways in which participants reflexively coordinate their perspectives on particular domains during the course of social interaction.

Acknowledgements

References

Garfinkel, H (1967) *Studies in Ethnomethodology* Englewood Cliffs, NJ Prentice-Hall

Goodwin, C (1995) Seeing in Depth, *Social Studies of Science* 25 2, 237-274

Harper, R , Hughes, J. and Shapiro, D (1989). 'Harmonious Working and CSCW Computer Technology and Air Traffic Control', in *Proc of ECSCW'89*, 73-86

Heath, C C and Luff, P (1992a). Collaboration and Control Crisis Management and Multimedia Technology in London Underground Line Control Rooms, *CSCW Journal* 1 (1-2), 69-94

Heath, C C and Luff, P (1992b) Media Space and Communicative Asymmetries Observations of Video-Mediated Interaction, *HCI Journal* 7·315-46

Heath, C C , Luff, P and Sellen, A (1995) 'Reconsidering the Virtual Workplace Flexible Support for Collaborative Activity', in *Proc of ECSCW'95*, 83-100

Hindmarsh, J , Fraser, M , Heath, C C , Benford, S and Greenhalgh, C (1998) 'Fragmented Interaction Establishing mutual orientation in virtual environments', in *CSCW'98*, 217-26

Hutchins, E L (1995) *Cognition in the Wild* Cambridge MA MIT Press

Kuzuoka, H , Kosuge, T and Tanaka, M (1994) 'GestureCam A Video Communication System for Sympathetic Remote Collaboration', in *Proc of CSCW'94*, 35-44

Kuzuoka, H., Oyama, S., Yamazaki, K and Suzuki, K (2000) 'GestureMan A Robot that Embodies a Remote Instructor's Actions', in *Proc of CSCW 2000*

Latour, B (1992) Where Are the Missing Masses? The sociology of a few mundane artifacts, in *Shaping Technology/Building Society*, Bijker and Law (eds), 225-258 MIT Press

Robertson, T (1997) 'Cooperative Work and Lived Cognition A Taxonomy of Embodied Actions', in *Proc of ECSCW '97*, 205-220.

Suchman, L (1996) Constituting Shared Workspaces, in *Cognition and Communication at Work*, Engestrom, Y and Middleton, D (eds), 35-60 Cambridge University Press

Tatar, D G , Foster, G and Bobrow, D G (1991) Designing for conversation Lessons from Cognoter, *IJMMS* 34 (2), 185-209.

Yamazaki, K , et al (1999) 'GestureLaser and GestureLaser Car Development of an embodied space to support remote instruction', in *Proc of ECSCW'99*, 239-258