

**Disembodied Conduct:
Interactional Asymmetries in
Video-Mediated Communication**

Christian Heath and Paul Luff

Technical Report EPC-1992-119

*Published in:
Technology in Working Order: Studies of
Work, Interaction, and Technology,
Graham Button (ed),
Routledge, London, 1993,
pp 35-54.*

Copyright © Graham Button, Christian Heath and Paul Luff 1992.

Rank Xerox Research Centre
Cambridge Laboratory
61 Regent Street
Cambridge CB2 1AB

Tel:+44 1223 341500
Fax:+44 1223 341510

Disembodied Conduct: interactional asymmetries in video mediated communication

Christian Heath and Paul Luff

University of Surrey / Cambridge EuroPARC

What of the hands? We require, promise, call, dismiss, threaten, pray, supplicate, deny, refuse, interrogate, admire, number, confess, repent, confound, blush, doubt, instruct, command, incite, encourage, swear, testify, accuse, condemn, absolve, abuse, despise, defy, flatter, applaud, bless, humiliate, mock, reconcile, recommend, exalt, entertain, congratulate, complain, grieve, despair, wonder, exclaim,... There is not a motion that does not speak and in an intelligible language without discipline, and a public language that everyone understands.

Montaigne 1952 pp. 215-216

1 INTRODUCTION

Over the past thirty years, there have been numerous attempts to develop audio-visual technologies which provide real time access between geographically dispersed individuals; (cf. Egido 1990). As yet however, these attempts have met with relatively little success. The videotelephone and conferencing systems were early precursors of such developments, and in more recent years we have seen the ways in which audio-visual technologies can support a range of computer-based tools, such as shared meeting spaces (e.g. Watabe et al. 1990), shared text editors (e.g. Olson et al. 1990) and shared drawing tools (Bly 1988, Tang and Minneman 1991). These extraordinary technological innovations have been accompanied by a growing body of research concerned with the potential contribution of audio-visual technologies to cooperative work (including Abel 1990, Egido 1990, Gale 1989, Fish et al. 1990 and Smith et al. 1989). Despite this work, we still have relatively little understanding of the character of interpersonal

communication mediated through video technologies or the extent to which the media facilitate collaboration in the work place.

At EuroPARC, Rank Xerox Research Laboratories in Cambridge, we have been developing a media space to support cooperative work and informal sociability between physically distributed individuals within a single building. As part of these developments we have been exploring the ways in which personnel within the organisation utilise and communicate through a ubiquitous audio-visual infrastructure that provides for free-flowing accessibility between individuals in different physical locations. Preliminary observations suggest that whilst the technology provides real time audio and visual access between personnel, it transforms the visual conduct of the participants which in turn can be consequential to the delivery and receipt of talk. In particular, interpersonal communication through the media space appears to introduce certain asymmetries into the interaction which are as far as we are aware not found in more conventional settings. In this chapter, we wish to explore these asymmetries and briefly discuss their implications for developing a technology to support real world collaborative work.

1.1 Background and Setting

An important element in EuroPARC's recent initiatives is the development of an audio-visual infrastructure in its Cambridge offices. The infrastructure allows scientists and administrative staff to establish visual and audible contact with each other, or to view public areas such as a commons area and the conference room. EuroPARC's offices straddle three floors, and in part the technology was introduced to facilitate informal contact between organisational personnel. The system basically consists of a camera and 14" monitor in each office, with larger monitors in the public areas. There is some variation to the way in which members of the laboratory position their monitor and camera in relation to their workstation and within the overall environment of their office. Typically however, the monitor with camera seated on top is positioned to one side, roughly at a 120 degree angle, to their workstation which is usually operated by a flat PZM, *multidirectional* microphone is normally positioned on the desk by the workstation. It is operated by a footpedal.

Over the past three years the infrastructure has become increasingly sophisticated and we have experimented with various additional facilities which might enhance contact and cooperation between EuroPARC's personnel. A number of these developments have been designed to provide 'users' with more delicate ways of scanning the local environment or establishing connectivity; (see for example Borning and Travers 1991). Despite these technological developments, the most prevalent use of the system within the EuroPARC is to maintain an open video connection between two physical domains, typically two offices. These 'office shares' are often preserved over long periods of time, weeks and sometimes months, and simply provide two physically distributed individuals with the a sense of co-presence. Audio connections are normally switched off, until the two colleagues specifically wish to speak with each other.

As part of the introduction and development of the system, we undertook selective audio-visual recording of particular connections between individual offices. To diminish the potential influence of recording on the way individuals used the system, and to enable us to gain an overall picture of how frequently and for what purposes individuals used the connection, we undertook 'blanket' recording of particular connections for up to two or three weeks. This data corpus was augmented by field observation both of connections and discussions in the laboratory concerning the system. We also collected audio-visual recordings of experimental systems, and the use of related technologies in environments other than EuroPARC, for example the Xerox Television (XTV) link between Britain and the USA.

The analytic framework of the research draws from conversation analysis and cognate approaches. It focuses on the *in situ* or contextual character of human conduct and in particular the sequential and socio-interactional organisations which inform the production and intelligibility of social actions and activities. In the project at hand, the analysis was simultaneously driven by a range of substantive concerns which in part derived from the growing body of research concerned with the interactional organisation of visual and vocal conduct (for example, Kendon 1990, Erickson and Schultz 1982, Cosnier and Kebrat-Orecchioni 1987, Goodwin 1981, Heath 1986). The corpus of findings within these and related studies, provided a heuristic with which to compare some features of the organisation of interaction in videomediated presence, and through detailed case analysis began to reveal some curious features of communication though audio-visual technology.

2 VIDEO MEDIATED COMMUNICATION

We will begin by briefly discussing how audio-visual connectivity can contribute to collaborative work and informal sociability between physically distributed individuals and then proceed to reveal the way in which the technology can transform the significance of both visual and vocal conduct.

2.1 Gearing Conversational Initiation

Unlike a telephone or audio connection, video provides the opportunity for individuals visually to assess the availability of a colleague before initiating contact. More precisely, the video channel not only allows one individual to discern whether a colleague is actually in his or her office, but also to assess more delicately the state of his or her current activity and whether it might be opportune to initial contact. At EuroPARC, we have recently introduced facilities into the infrastructure which further support the possibility of momentarily glancing at a colleague before deciding whether it is opportune to establish engagement. In this way, video makes an important contribution not only to the awareness of others within a physically distributed, work environment, but also to one's ability to respect the territorial rights and current work commitments of one's colleagues.

It should be added that video technology also provides the possibility of unmotivated encounters, that is, colleagues happening to encounter each other, by for example, one person noticing another noticing them and initiating conversation. In this way the technology can begin to support the range of informal sociability which would normally be associated with an open office environment, whilst avoiding some its obvious disadvantages.

2.2 Coordinating Talk with Concurrent Activity

Once individuals have established contact with each other, video provides participants with the ability to coordinate talk with a range of other activities in which they might be simultaneously engaged. This aspect of video's contribution is particularly important to Computer Supported Cooperative Work (CSCW) where individuals are frequently undertaking screen based activities whilst speaking with colleagues. Mutual visual access provides individuals with the ability to discern, to a limited extent, the ongoing organisation and demands of a colleague's activities, and thereby coordinate their interaction with the practical tasks at hand. Moreover, mutual visual access provides individuals with the ability to point at and refer to objects within the shared local milieu.

Such facilities have become increasingly important in recent years as scientists have begun to develop shared real time interfaces (cf. Bly 1988, Olson et al. 1990). Recent experiments by Smith et al. (1989), and Olson and Olson (1991) have demonstrated the importance of providing video for participants to coordinate simultaneous screen based activities. Curiously the studies by Smith et al. (1989) tend to suggest that the video channel is employed during periods of talk about the task, or as they describe it "metatalk", rather than featuring within the coordination of the activity. The system's design, and facilities such as individual pointers, providing the necessary distinctions and referents for the participants jointly to accomplish the task.

2.3 Speaker Recognition and Turn Transfer

As part of our data collection, we gathered recordings of the use of multi-party audio-visual connections both at EuroPARC, and the Xerox Video Conferencing facility at Welwyn Garden City. Analysis of these materials reveals that participants are relatively insensitive to much of each other's visual conduct, and yet the video channel does seem critical to the organisation and flow of the meetings the systems support. The key contribution of video within these multi-party interactions appears to be the way in which it provides participants with the ability to recognize who is speaking and to systematically coordinate speaker change. In particular, video provides the possibility of participants utilising more flexible and informal procedures for coordinating speaker turns. For example, audio or telephone conferencing necessitates relatively 'formal' arrangements for allocating and exchanging the floor between multiple participants.

Despite the important contribution made by video to support collaborative work between physically dispersed individuals, it is important to note that the technological medium provides a communicative environment which markedly differs from actual physical co-presence. In the following we will sketch some of the more significant differences between human conduct performed through audio-visual media and actions and activities undertaken in face to face settings. These differences can raise certain problems for users, but these problems can, ironically, provide a strong foundation for screen or document based collaborative work. We will begin by discussing the way in which a look, and gaze, is transformed when performed through video and go on to discuss the articulation of talk and production of gesture in and through media space.

2.4 Looking through video

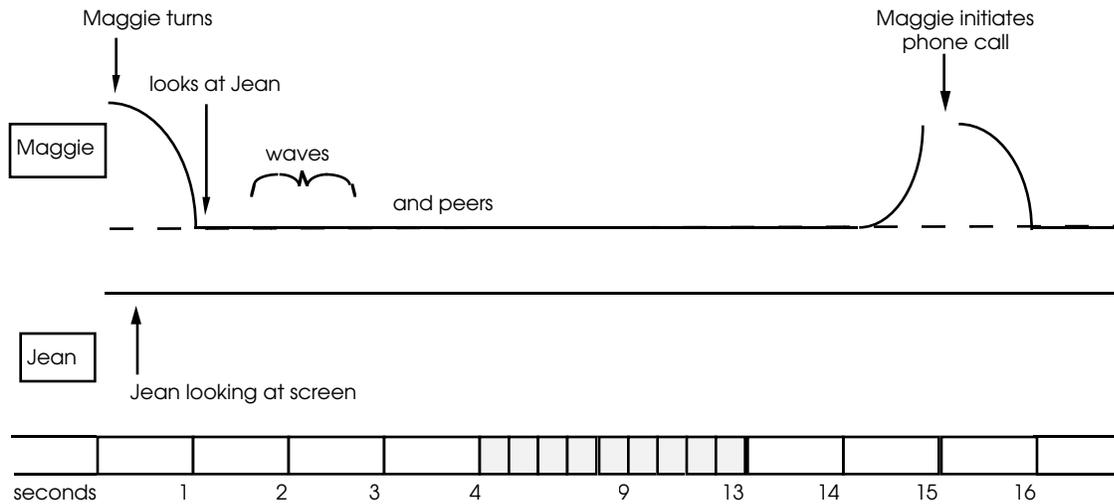
When one perceives another is looking at one, one perceives that the other intends something by one, or expects something of one. In a word, one is being taken account of by another. It seems reasonable to suppose that this will have marked arousing consequences, but what line of action it arouses another to take will depend upon the context in which the look is perceived.

Kendon 1990 p. 51

In many ways personnel in EuroPARC who maintain open video connections treat the visual availability of a colleague on the screen as if the other is co-present, though perhaps at some distance. They presuppose that they are able to exchange glances with each other, wave to initiate contact and remain sensitive to another's conduct. They assume that the medium provides mutual accessibility and the ability to initiate contact, to move from disengagement to conversation, as they might if they were actually co-present. For example, in the following fragment, drawn from a recording of a video connection between a scientist and a member of the administration, we find Maggie, the scientist, attempting to initiate contact with Jean by looking and waving¹.

¹ For the purposes of presentation in this paper it is useful to map an individual's visual conduct relative to the other 'video partner'. Turning towards the other is indicated by an arc moving towards the dashed line, turning away by an arc moving away and gaze towards the other is indicated by a continuous line along the dashed line.

Fragment 1

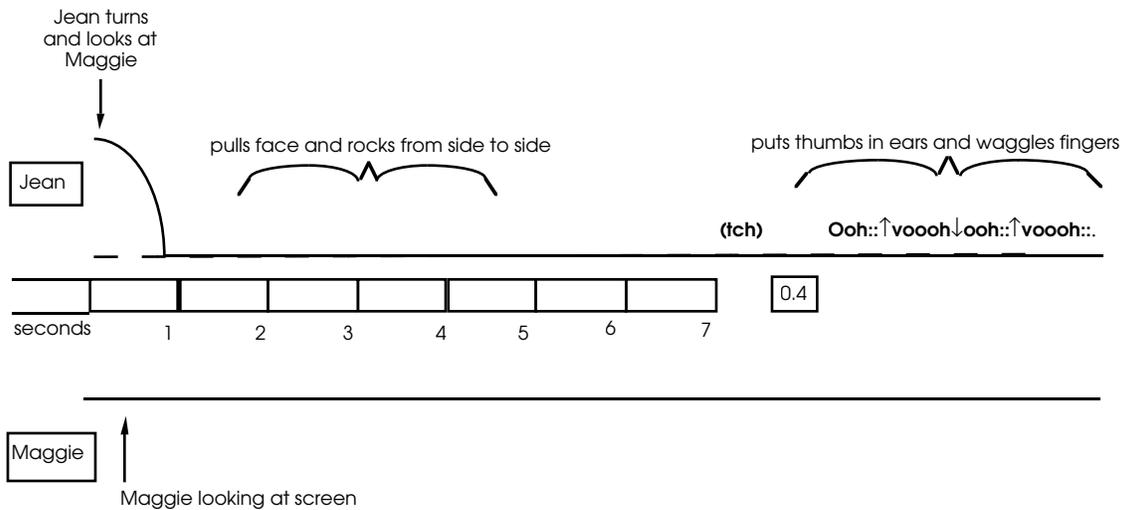


Maggie turns towards Jean and then waves. For more than ten seconds she stares at Jean but receives no response. Finally, Maggie looks away to her phone and dials Jean's number, summoning Jean to the telephone. Only when Jean replies to Maggie's greeting do the parties establish visual contact. In attempting to contact Jean, Maggie assumes that their mutual visual accessibility provides the possibility of attracting the other's attention in the way she might if she were physically co-present. Despite the failure of her initial glance to engender a response from her colleague, Maggie presupposes that various visual actions will eventually serve to attract attention. By recycling her shift of gaze towards Jean, gesturing, and even staring, Maggie presupposes that her visual conduct will be effective in the way that it might if they were in a similar socio-spatial arrangement in co-presence. Even the shape of her gesture, the waving hand criss-crossing her line of regard and forming a potentially noticeable element within the mutual domain not only fails to engender a response but actually passes unnoticed.

In this and numerous other instances in the data corpus, we find participants presupposing the effectiveness of the resources they might ordinarily use to establish mutual engagement, even after their actions have met with successive failure. They assume that a glance or wave will have the impact it might if they were co-present with the other, and will systematically upgrade their attempts to engender a response from a colleague despite the other's apparent insensitivity. In video mediated presence, personnel at EuroPARC initially assume that the medium provides an environment akin to co-presence, in which relatively insignificant features of bodily comportment, such as looking at another, can serve to gain another's attention and provide for the progressive movement into a state of talk and interaction. However, presence mediated through video appears to interfere with the ability of a look, or even a wave, to engender response from a potential co-participant. Though users assume that the technology will support delicate ways of moving from a state of incipient talk to mutual engagement, they discover that these forms of visual conduct are not necessarily supported by the technology.

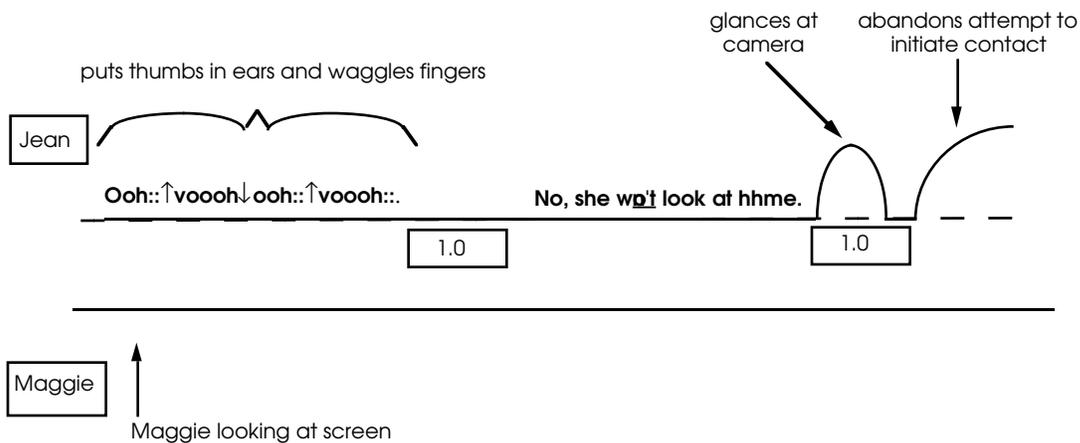
In the following instance drawn from a recording of the same connection a couple of days later, the users face similar problems establishing mutual contact through video².

Fragment 2 Transcript 1



Prior to the beginning of this fragment Jean returns to her desk. As she sits down in her chair, she turns and acknowledges Maggie, in the way she might if she were reestablishing co-presence. Neither her reemergence within the scene, nor her initial glance serve to engender any response from her colleague and in consequence she begins a series of gestures through which she attempts to gain Maggie's attention. She turns to her colleague, places her thumb on her nose, and rocking from side to side, waggles her hand at her colleague. This elaborate performance fails to attract notice and, continuing to look at her colleague, Jean places both thumbs in her ears, waggles her hands and utters "Ooh::↑vooh↓ooh::↑vooh::.." Finally, thrusting her face towards the monitor (the other), she abandons her attempt to initiate contact, uttering "No, she won't look at hme".

Fragment 2 Transcript 2



² The transcription system for talk was developed by Gail Jefferson; details of the system can be found in various monographs and collections including, Atkinson and Heritage (1984), Goodwin (1981) and Heath (1986).

It is interesting to note that the gestures used by Jean in her attempt to establish mutual contact, allow her to exaggerate her visual orientation towards Maggie. The waggling fingers balanced on the end of her nose that flicker across her line of regard, coupled with movements from side to side, exaggerate her bodily and visual alignment towards the other. Similarly, the hands to the side of the head, and the moving fingers, underscore her orientation towards her colleague, broadening the visual appearance of her head and its particular alignment. Despite her theatrical movements, Jean's attempts to gain her colleague's attention fail and a few moments later she abandons her efforts to initiate contact.

As in fragment 1, the user presupposes the effectiveness of their visual conduct through the media, assuming that a glance and then, more dramatically a series of gestures, will 'naturally' engender a response from the potential co-participant. However, the glances, their accompanying gestures and the power of the look to ordinarily attract the attention of another appears to be weakened when performed through video rather than face to face. In neither these, nor the many other instances we have examined, is there evidence to suggest that the potential co-participant is deliberately disattending the attempts to attract their attention. Rather, the looks and gestures of their colleagues simply pass unnoticed as if their appearance on a screen rather than in actual co-presence diminishes their performative and interactional impact. The sequential and interactional significance of such actions is undermined by the medium. In consequence, the relatively delicate ways in which individuals subtly move from disengagement to engagement in face to face environments, especially when they are in a 'state of incipient talk', appear to be rendered problematic in video mediated co-presence.

The relative impotence of a look or gaze when mediated through video is also consequential to the ways in which individuals preserve privacy. In co-presence, the interactional significance of a look, coupled with our ability to remain sensitive to the actions of others, even though they may be outside the direct line of our regard, constrains our looking at others and their opportunity to look at us. Privacy relies upon the power of the look to engender action and peripheral awareness, even amongst individuals who may be co-present yet disengaged. By undermining the performative impact of a look and our ability to remain peripherally sensitive to the conduct of the other, as it appears on the screen rather than in co-presence, audio-visual technology can interfere with constraints and competences, the socio-interactional organisation, which provide the foundation to privacy in 'public' domains. As we saw in fragment 1, the failure of a look to engender a response, can inadvertently lead an individual to stare at a colleague, to watch another's actions where the other is unaware that they are receiving the attentions of their colleague.

As individuals have become accustomed to the technology and its fallibilities, they have become increasingly sensitive to such problems and in some cases have developed a range of practical solutions to enable them to preserve each other's privacy. Even so, by posing a threat to the ordinary ways in which we maintain privacy in 'public' domains, the technology, at least as it is currently configured, can continue to generate a range of difficulties and make people uneasy about using the media space. In this light

we have recently been exploring technological solutions to some of the problems which derive from the ways in which video can undermine relatively minor, yet significant, aspects of our socio-interactional organisation.

2.5 The articulation of talk and recipient insensitivity

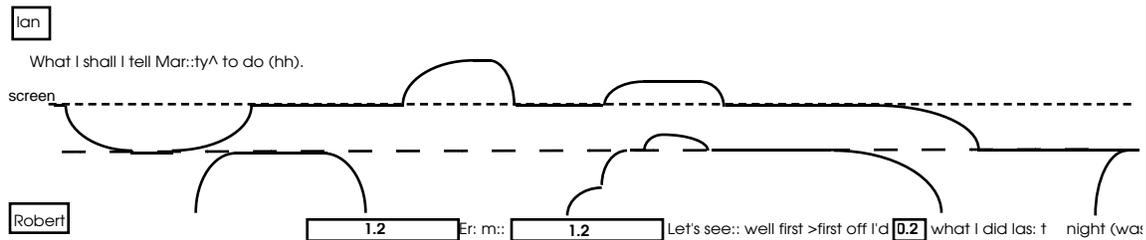
The relative ineffectiveness of gaze and other forms of visual conduct in video mediated co-presence is not only consequential to ways in which users are able to establish mutual engagement. It can also generate 'difficulties' for the articulation of talk and for the communication of information embodied in gestures and other forms of bodily movement. The following example drawn from an extensive collaboration between two scientists over a period of a few weeks. In it we find that the relative inability of looking and gesturing to have an impact on the co-participant is consequential to the delivery of an extensive answer. The collaboration was supported by an open video and audio connection between two offices at EuroPARC. We join the action as Ian initiates contact with Robert by enquiring what he should tell Marty to do.

Fragment 3, Transcript 1 (AB 41:5, 17:51:52)

1 I: What I shall I tell Mar::ty^ to do(hh).
2 (1.2)
3 R: Er:°m::
4 (1.2)
5 R: Let's see:: well first >first off I'd (.2) what I did
6 las: t night which seemed to (work) was send it tw::ice
7 under different names:: <an then she did a (di::p:).
8 (1.6)
9 R: en then she: could clean up the er::: (.8) line
10 noi:se.
11 (...)
12 (2.3)
13 R: °thhh
14 (.3)
15 I: O.k ay
16 R: (Such a hack)

At the outset it can be noticed that Robert delays his reply to Ian's question firstly by pausing, then by producing "Er:°m::" (line 3), and then once again by pausing (line 4). Even when he does begin to reply, the actual answer is not immediately forthcoming. It is as if the speaker is deliberately delaying the gist of his response. The gist of the reply is delayed, pushed away from the beginning of the speaker's turn, by virtue of the preface "Let's see::" and various forms of speech perturbation, including a sound stretch ("see::"), a 0.2 second pause (line 5) and consecutive restarts "well first >first off I'd (.2) what". The speaker's actions and his apparent difficulty in beginning his reply may be systematically related to the conduct of the (potential) recipient, and particularly with Robert's inability to secure his co-participant's gaze. A more detailed transcript might be helpful.

Fragment 3, Transcript 2.



Ian's original question is produced as the two colleagues are independently engaged in individual screen based activities. As he initiates contact by asking the question, Ian momentarily glances at Robert and then turns back to his screen. By the time Robert reorientates, he finds Ian engaged in dealing with some text. In the light of having a potential recipient who is apparently engaged in another activity, Robert uses an utterly conventional device for attempting to secure co-participant alignment; he delays the delivery of the reply and produces an object which projects more to follow but withholds the actual reply; (cf. Goodwin 1981, Heath 1986).

Withholding the reply fails to engender any reorientation from Ian, and following "Er:°m::", Robert begins to progressively shift his gaze towards Ian, as if attempting to encourage a reorientation whilst avoiding actually staring at his potential recipient. Both the withholding of the reply, and the subtle shifts in Robert's orientation fail to encourage any display of reciprocity from the co-participant. Robert begins the preface "Let's see::" and looks directly towards his colleague. The alignment of gaze towards the co-participant, the preface, the sound stretch, the pause and the restarts are all devices which are regularly used to secure recipient alignment at the beginning of a turn. Finally the pause appears to engender a response from Ian, and following his realignment of gaze from the screen towards his colleague, the speaker begins the gist of his reply with little evident perturbation or difficulty.

It is apparent therefore that in fragment 3 the respondent has various difficulties in securing the relevant form of co-participation from the potential recipient, ironically the party who initiated the interaction in the first place. The potential recipient displays little orientation to the speaker's successive attempts to secure his gaze. It is possible that the difficulties faced by the speaker in attempting to secure a realignment from the recipient derive from the relative ineffectiveness of his visual conduct and in particular, the apparent inability of the co-participant to notice the successive shifts in orientation undertaken by the speaker. Elsewhere (Goodwin 1981 Heath 1986), it has been demonstrated that a realignment of gaze by a speaker towards a potential recipient can be a critical element in utilising perturbations in talk or even gestural activity to engender a co-participant's reorientation. In video mediated interaction, the relative scale and presentation of a speaker's more delicate shifts in orientation on a TV screen may pass unnoticed and thereby undermine the performative impact of conventional devices to elicit gaze.

In passing a further point should be mentioned. To provide individuals with the ability to vary their position whilst speaking with colleagues through the media space, we deliberately used multidirectional microphones (PZM) to provide audio connections. These multidirectional microphones

are designed to conceal relative changes in the direction of a sound within a circumscribed domain. In consequence, they mask changes in the sound level of a speaker's voice which ordinarily allow a potential recipient to infer when a speaker is realigning their gaze. Thus, the relative ineffectiveness of a speaker's shift of gaze to engender a response in video mediated interaction may not only derive from the relative inaccessibility of visual conduct, but also from the absence of changes in tone and loudness of the voice.

2.6 The visibility of gesture

It is often suggested that one of the important contributions that video can make to interpersonal communication, derives from the way in which it can convey gestures and other forms of bodily movement. It has been found for example that participants do coordinate relatively subtle elements of their bodily comportment (Smith et al. 1989, Olson and Olson 1991) and that the technology allows individuals to discern information conveyed through gesture and such like. In our own research, we have found interactants using gestures as they would in face to face interaction, not only to convey information but to organise how a recipient should participate during the delivery of turns at talk. In the following fragment, for example, Robert is attempting to deliver an extensive description of an interface to Ian. As he delivers the description Robert produces a series of gestures which appear to both illustrate certain features of the interface whilst simultaneously attempting to engender some form of response from Ian.

Fragment 4

1	R:	there's: two degrees of freedom you can	
2		move it in X an Y:	← side to side gesture
3		(0.3)	
4	R:	if there are more than two degrees of	
5		freedom you can select which variables	
6		were to be manipulated:	← open palm movement
7		(0.5)	
8	R:	which (will) remain fixed	← flat palm movement
9		(0.3)	
10	R:	an then manipulated at (.)	
11		two::(.) <three variables by	← glance towards Ian
12		the control: icon.	← finger movement
13		(1.2)	
14	R:	er:: is: this correct	
15		(1.0)	
16	I:	Well:: (.) not quite.	

The series of gestures which accompany and embody the description are designed to illustrate certain features of the interface and to circumscribe certain segments of information within the description itself. The gestures are shaped to provide a visual portrayal of the objects or events mentioned in the talk (see for example, Birdwhistell 1970, Bull 1983, Ekman and Friessen 1969, and Schegloff 1984). There is little evidence however, that despite looking at the speaker throughout the delivery of the description, Ian is able to retrieve relevant information from the visual conduct of Robert. Part of the difficulty in utilising the gestures to elaborate the talk and provide the gist of Robert's description, derives from way in which the gestures disappear out of camera range and leave the recipient with the speakers outstretched arms but little indication how they form part of the overall gesture. However, even where the gesture remains largely within the screen, there is a certain way in which the movement appears to be disembodied from the talk it is illustrating. For whatever reason, the illustrative and iconic elements of the gesture appear to lose their performative impact as they are articulated through the medium.

The gestures are also designed perhaps to shape the way in which the potential recipient should participate during the course of the description. Indeed, the movements accompanying the utterances, for example the open palm gesture with "an then manipulated at..." (lines 10-12) and the thrust towards the recipient with the flat palm at "fixed" appear to designed to elicit some form of acknowledgement. Even the illustrative gesture with "x an y" might be designed to elicit some form of response from the recipient whilst simultaneously elaborating the talk with which it occurs. The gestures fail to engender either a visual or vocal response from the recipient either during each utterance or immediately post their delivery. Indeed curiously, despite Ian's orientation towards the monitor, the gestures seem to pass unnoticed, failing to elicit any form of response from the recipient. It is as if the gestures lose their performative impact. The recipient is not simply withholding response to the speaker's actions, but rather the gestures themselves appear to lose their ability to engender sequentially relevant action from the 'co-participant'.

The relative inability of the speaker's visual conduct to effect some response from the recipient during the production of turns at talk, is found elsewhere, amongst different users within the data corpus. Even relatively basic sequences that recur within face to face interaction tend to be absent from the materials at hand; for example when a speaker uses a movement to elicit the gaze of a recipient and coordinates the production of an utterance with the receipt of gaze. Speakers continue to gesture and produce a range of bodily behaviour during the delivery of talk in video mediated communication, yet their visual conduct fails to achieve sequential and interactional significance. The recipient remains unaffected, and in the light of an apparently recalcitrant recipient, speakers systematically upgrade their demands, their gestures and movements becoming more theatrical as they fail to shape the relevant forms of co-participation within the production of the activity. In the case at hand, fragment 4, the movements are relatively gross even though the recipient is actually looking towards the speaker, but

even here gestures and visual reorientation prove ineffective; the speaker's conduct is visible, but impotent.

3 MEDIATED CONDUCT AND CO-PRESENT ACTION

3.1 Communicative asymmetries in video mediated presence

Video mediated presence reveals asymmetries in interpersonal relations which, as far as we are aware, are neither found within face to face interaction nor in other technologically mediated forms of communication such as telephone calls. Indeed, even in the light of the growing corpus of literature concerned with asymmetries within various forms of institutional language use and interaction, such as the news interview (Greatbatch 1986), the courtroom (Atkinson and Drew 1979), or the medical consultation (West 1985), the distribution of communicative resources are peculiar in video mediated presence. In institutional environments we find the incumbents of preestablished roles, such as doctor and patient, having differential access and influence to activity types throughout the course of an event. By contrast, in video mediated communication, the asymmetries tend to parallel the categories 'speaker' and 'hearer' and are in continual flux during the conversation. The asymmetries interfere with the very possibility of accomplishing certain forms of conduct.

The technology provides the participants with mutual visual access, and yet undermines their ability to successfully perform gestures and other forms of bodily conduct. A speaker, for example, is able to monitor the visual conduct of a co-participant and remain sensitive, even during the production of a single turn at talk to the behaviour of the recipient. However, the resources the speaker might ordinarily use to shape the way in which the recipient should participate, gestures and the like, are unreliable when performed through the medium. It is as if the technology undermines their sequential force and thereby their impact within the interaction and, as we have seen, it is not only whilst people are speaking that their visual conduct may be ineffectual. The relatively subtle glances through which individuals can ordinarily reestablish mutual engagement, even gross attempts to attract the attention of another, can be undermined by the medium through they are performed. So, on the one hand, video mediated presence provides individuals with the ability to mutually monitor each others visual appearance and conduct, but on the other hand fails to provide a reliable medium for the production and recognition of nonvocal actions and activities. The technology appears to interfere with the *local, sequential significance* of a range of visual actions, revealing an alternating imbalance between the participants as they shift between the role of speaker and hearer.

3.2 Incongruent environments of action

As yet it is still unclear why audio-visual technologies may fail to support visual conduct and whether it is possible to develop a system which would

deal with the difficulties that we have discussed. However, it is possible that the asymmetries we find in videomediated interaction derive from the incongruent environments in which the participants' conduct is produced and received.

Much of the work performed through gesture and more generally body movement, is accomplished on the periphery of the visual field; indeed, visual conduct and the ways in which it organises co-participation relies upon being 'seen but unnoticed'. It 'glosses' or masks its own operation and is frequently designed to work on the margins of the perceptual field of the recipient. In the case of video mediated communication, the recipient's access to the other's gesture or bodily activity is their appearance on a screen. The gesture is either viewed directly, within the totality of the screen's contents or, if the recipient is looking to one side, constitutes one element of the screen's overall image. In the first case seeing the image as a whole destroys the relative weighting of gesture or bodily activity in relation to the rest of the speaker's appearance. In the second, it is found that recipients are largely unable to differentiate various elements of a speaker's bodily comportment if they are viewing the screen on the periphery of their vision. Only occasionally are relatively gross movements, such as the other standing or blocking the screen, noticed and 'noticeable'. In consequence, much of the delicacy which features in the design and performance of certain forms of bodily activity is lost to the recipient.

These difficulties become more severe when one considers that the camera and monitor inevitably distort visual conduct. The technology inevitably transforms the spatial and temporal organisation of a movement. So for example, the graduated progression of a gesture, designed for example to elicit the gaze of a recipient, its movement towards the periphery of the other's visual field is transformed when it appears on the screen of the co-participant. The object received is not the object produced.

The way in which the technology transforms our ability to peripherally monitor the conduct of the other and distorts the appearance of bodily conduct might also explain why video mediated gaze or looking is often ineffectual. Unless the recipient is looking directly at the screen, it is unlikely that he or she is able to discern relatively small changes of head and eye movement of the co-participant. Moreover, it is possible that the relative inability to discern changes in another's gaze direction may well explain why conventional devices for securing recipient alignment such as pauses, sound stretches and other forms of speech perturbation seem to lose their impact when performed through video. Various studies (Goodwin 1981, Heath 1986) suggest that in face to face interaction, these devices sometimes rely upon the accompanying visual orientation of the speaker to engender the relevant responses from the co-participant. The technology renders the look ineffective by virtue of the recipient's inability to peripherally discriminate small changes in bodily orientation of the other when it appears on a screen rather than face to face.

It is also worthwhile to consider the way in which an action is produced and the problems faced by, for example a speaker, in communicating through video. In video mediated presence, the camera and monitor inevitably delimit and distort access to a co-participant. An individual's view of the

other is from a particular angle and precludes access to a large part of the their body or local environment. In consequence the ability to design a bodily movement such as a gesture, so that it, for example, operates on the periphery of the visual field of the co-participant becomes extremely problematic. It is not only that a participant has 'limited' access to the other and their current conduct, but also the access the participant has is distorted by the camera. Moreover, the limited access to the other also means that a participant is relatively unaware of changes within their local environment with which their visual conduct may well be competing. For example, it is not unusual in the materials gathered within EuroPARC to find individuals undertaking screen based activities as they speak to a colleague through the media. The inability of the co-participant to see the other's screen, or perhaps more importantly the other's involvement within the activity in which they are engaged, (which might for example be embedded in the use of a workstation), undermines the ability to design contextually relevant actions and activities. These problems become more severe, when one recognizes that in contrast to physical co-presence, the individual cannot change their own bodily orientation in order to adjust their perception of the recipient or his local environment.

Perhaps however, the foundation to many of these difficulties derive from the participant's inability to discern how their own conduct appears to the other. In consequence, it is difficult to envisage how an individual can attempt to modify their conduct in order to achieve the relevant impact. So for example, if a gesture is unsuccessful, in say eliciting the gaze of the co-participant, the speaker is unable to discern how the movement might have appeared to the other in order to redesign the action accomplishing the relevant performative impact. It is not surprising therefore, that in reviewing the data corpus, one finds numerous instances of individual's upgrading, even exaggerating particular gestures so as to achieve impact on the conduct of the recipient. Unfortunately, these attempts inevitably transform the action the speaker is attempting to accomplish and also frequently fail to engender any response from the co-participant.

The technology therefore, at least as it is currently designed and configured, provides physically distributed individuals with incongruent environments for interaction. Despite this incongruity, individuals presuppose the effectiveness of their conduct and assume that their frame of reference is 'parallel' with the frame of reference of their co-participant. Participants presuppose for the practical purposes at hand, an interchangeability of standpoints, a reciprocity of perspectives. Schutz in his classic studies of the phenomenology of the social world suggests that such a presupposition is a keystone of socially organised conduct.

Now it is a basic axiom of any interpretation of the common world and its objects that these various co-existing systems of coordinates can be transformed one into the other; I take it for granted, and I assume my fellow-man does the same, that I and my fellow-man would have typically the same experiences of the common world if we changed places, thus transforming my Here into his, and his - now to me a There - into mine.

(1962 pp. 315-6)

In video mediated presence however, camera and monitor transform the environments of conduct, so that the bodily activity that one participant produces is rather different from the object received by the co-participant. The presupposition that one environment is commensurate with the other undermines the very production and receipt of visual conduct and explains perhaps why gesture and other forms of bodily activity may be ineffectual in videomediated presence. The very presupposition which underlies socially organised conduct generates a range of difficulties for individuals who are attempting to produce and coordinate social actions and activities within the media space.

4 DESIGN CONSIDERATIONS

The observations above raise a number of issues for the design and development of technologies which provide individuals within a dispersed physical environment with virtual access to each other.

At EuroPARC, we have begun by attempting to address the issue of privacy in video mediated co-presence whilst preserving the ways in which technologies can support informal access between colleagues within the dispersed office environment. One development has been the introduction of a complex configuration of sounds which conventionally signify to others particular forms of action. For example, colleagues can be alerted to afternoon tea in the 'commons' through the sound of cups and saucers rattling and tea being poured. More relevantly, when one party chooses to glance at another, or even scans through the building via video, glancing to see who is present, individuals who have agreed to making themselves accessible to the colleague in question, hear the sound of a door opening as they momentarily appear on the other's monitor. In order to introduce symmetry into these fleeting interactions, recipients can themselves return a glance and look at the person who is looking at them. However, it is extraordinarily difficult to simulate the delicacy and range of actions that individuals ordinarily employ in initiating mutual contact with each other, even if it consists of no more than a passing glance, and we are still some distance from achieving the subtlety and non-incursiveness of the various ways in which colleagues momentarily establish contact with each other in an open environment. It is critical however that we explore ways in which we can preserve a balance between the privacy of the individual and mutual accessibility in the office environment, without being led to develop increasingly formal solutions to support informal sociability. To this end systems that incorporate even restricted forms of video access, such as the video snapshots of the Polyscope system described by Borning and Travers (1991), have incorporated features to provide for mutual availability and visibility.

The asymmetries within video mediated presence may not necessarily be detrimental to the accomplishment of certain tasks and activities; indeed the relative insensitivity to another's conduct may have important advantages over office environments in which employees are physically co-present. In an open office environment employees can find themselves continually subject to the demands of colleagues and sensitive to relatively small changes

within the local environment of 'goings on'. In contrast, video mediated presence and the asymmetries it introduces, allows individuals who are in one sense co-present to distance themselves from each other and thereby concentrate on individual tasks and activities. Even when individuals are mutually engaged, video mediated presence allows the participants to simultaneously undertake a variety of more or less related tasks and activities whilst simultaneously coordinating their actions with each other. Audio-visual technologies may make a particularly important contribution in computer supported cooperative work where participants need to coordinate a range of screen based activities whilst simultaneously accomplishing specific tasks and activities. Video mediated presence allows the individual to distance himself from the moment by moment demands of his colleagues whilst preserving mutual availability, allowing users to witness and coordinate tasks and activities whilst remaining relatively insensitive to the potential demands and interruptions of a shared physical environment.

Taking into account the potential advantage of certain asymmetries in video mediated presence for collaborative work, we have been undertaking a number of small experiments which attempt to deal with the relative ineffectiveness of visual conduct. We were interested in exploring ways in which we might at least increase an individual's sensitivity to the visual conduct of a colleague, to see if this might help solve the problem of privacy. In one experiment we replaced the standard 14" monitors with 26" screens placed on their sides for several of the staff in the laboratory. This was to see whether the relative increase in the size of the image of the other, would have a corresponding effect on the individual's sensitivity to the visual conduct of another. Preliminary analysis of the materials suggests that during periods of disengagement, individuals are more sensitive to relatively gross movements of the other, such as joining or leaving the scene. Occasionally an individual might notice another turning towards them and during conversation there is evidence to suggest that the gaze of a participant becomes more noticeable and effectual. In contrast, there is little evidence to suggest that the performative impact of gestural activity and other forms of body movement is more secure with the larger, reconfigured monitors, given that the users environments remain incongruent it is unlikely that a simple adjustment to image size will solve the difficulties which arise both in the design and receipt of visual conduct.

The relative increase in mutual sensitivity with the introduction of larger monitors can however generate problems for users. It has been found that individuals become increasingly conscious of the conduct and changes within the environment of their colleagues, and yet, unlike co-presence are unable to differentiate the status of such movements and find themselves unavoidably drawn to look at all a whole range of largely irrelevant actions and activities. The larger images of the other and their local environment, do not necessarily allow the user to peripherally monitor and assess the significance of changes in screen's contents. The technology, or rather the other and their local environment, becomes increasingly obtrusive within the user's milieu providing a stronger sense of co-presence, yet undermining the individual's ability to surreptitiously monitor and discriminate the actions and activities of their colleagues. Coupled with the obtrusiveness of the other with larger

monitors, is an certain self-consciousness as users appear to become increasingly sensitive of their own appearance in the local environment of their colleague.

In developing audio-visual technologies to support collaboration between physically dispersed individuals, we have tended to establish systems which provide a form of mutual access which is relatively invariable. However, when one considers the extraordinary range of tasks and forms of sociability that the technology might support, it becomes increasingly apparent that we need to provide users with the ability to systematically vary their mutual accessibility. For example, consider the forms of screen based collaboration reported in the experiments undertaken by Smith et al. (1989). For accomplishing many of their tasks users do not need visual access to each other, and yet at critical moments when they are discussing the problem they are trying to solve, a face to face orientation seems to be particularly useful. Or, consider the ways in which an individual delicately shifts the orientation towards a colleague as they mutually collaborate on some plans or drawings. It is precisely this issue that is addressed by recent developments in shared drawing tools at PARC. These offer particularly novel approaches to varying the accessibility of a video-projected collaborator (Tang and Minneman 1991).

The technology will support formal and informal collaboration, despite its potentially inherent problems, if users can systematically control, even in the developing course of an activity their audio and visual accessibility to each other and local environments of action. The success of the technology, its ability to facilitate rather than undermine a range of tasks and forms of communication, will depend upon whether we are able build a virtual presence which provides the delicate and systematic processes of interpersonal coordination found in real world, everyday, work environments.

ACKNOWLEDGEMENTS

We should like to thank Robert Anderson, Victoria Bellotti, Paul Dourish, Bill Gaver, Marina Jirotko, Allan Maclean, Tom Moran, Gary Olson and Judy Olson for their comments on earlier versions of this paper and also thank colleagues and numerous visitors to EuroPARC for discussions on some of the issues it raises. Part of the research reported here was undertaken under the auspices of an ESRC/MRC/SERC Research Grant concerned with the Social Organisation of Human-Computer Interaction.

REFERENCES

- Abel, M. (1990). Experiences in an exploratory distributed organisation. In Galegher, J., Kraut, R. E. & C. Egido, (Eds.), *Intellectual Teamwork: Social and Technological Foundations of Cooperative Work* (pp. 489-511). New Jersey: Lawrence Erlbaum Associates.
- Atkinson, J. M. and P. Drew (1979) *Order in Court*. London: Macmillan

-
- Atkinson, J. M. and J. C. Heritage (eds) (1984) *The Structures of Social Action: Studies in Conversation Analysis*. Cambridge: Cambridge University Press
- Birdwhistell, R. L. (1970). *Kinesics and Context: Essay in Body-Motion Research*. Philadelphia: University of Pennsylvania Press.
- Bly, S. A. (1988). A Use of Drawing Surfaces in Different Collaborative Settings. In *Proceedings of CSCW '88*, (pp. 250-256). 26th-28th September, Portland, Oregon.
- Borning, A. and Travers, M. (1991). Two Approaches to Casual Interaction over Computer and Video Networks. *Proceedings of CHI '91* (pp. 13-19). New Orleans, Louisiana..
- Bull, P. (1983). *Body Movement and Interpersonal Communication*. Chichester: John Wiley and Sons.
- Cosnier, J., & Kebrat-Orecchioni, K. (1987). *Decrire la Conversation*. Lyons: Publicitie Universitie Lyons.
- Egido, C. (1990). Teleconferencing as a technology to support cooperative work: Its possibilities and limitations. In Galegher, J., Kraut, R.E. & C. Egido, (Eds.) *Intellectual Teamwork: Social and Technological Foundations of Cooperative Work* (pp.351-373). New Jersey: Lawrence Erlbaum Associates.
- Ekman, P. & Friessen, W. V. (1969). The Repertoires of Nonverbal Behaviour: Categories, Origins, Usage and Coding. *Semiotica* 1, 49-98.
- Erickson, F. & Schultz, J. (1982). *The Councillor as Gatekeeper*. New York: Academic Press.
- Fish, R. S., Kraut, R. E. & Chalfonte, B. L. (1990). The videowindow system in informal communication. *Proceedings of the Conference on Computer Supported Collaborative Work* (pp. 1-11). Los Angeles, California.
- Gale, S. (1989). Adding audio and video to an office environment. *Proceedings of the First European Conference on Computer Supported Cooperative Work* (pp. 121-133). London.
- Goodwin, C. (1981). *Conversational Organisation: Interaction between a Speaker and Hearer*. London: Academic Press.
- Greatbatch, D. (1988). A turn taking system for British news interviews. *Language and Society*, 17,3.
- Heath, C. C. (1986). *Body Movement and Speech in Medical Interaction*. Cambridge: Cambridge University Press
- Kendon, A. (1990). *Conducting Interaction: Studies in the Behaviour of Social Interaction*. Cambridge: Cambridge University Press
- Montaigne, M. Eyquem de (1952). *Essays, Book II, no' 12* Trans, C. Cotton. Chicago Encyclopedia Britannica.
- Olson, G. M. and J. S. Olson (1991). User-Centered Design of Collaboration Technology. *Journal of Organisational Computing*, Vol 1, No. 1, 61-83.
- Olson, J. S., Olson, G. M., Mack, L. A. & Wellner, P. (1990). Concurrent editing: the group interface. In *Proceedings of Interact '90 - Third IFIP Conference on Human-Computer Interaction* (pp. 835-840). Cambridge, Aug. 27-30.
- Schegloff, E. A. (1984). Iconic gestures: locational gestures and speech production. In J. M. Atkinson & J. C. Heritage (Eds) *The Structures of Social Action: Studies in Conversation Analysis* (pp. 266-96). Cambridge: Cambridge University Press.
- Schutz, A. (1962). *The Problem of Social Reality: Collected Papers I*. Natanson, M. (Ed.) The Hague: Martinus Nijhoff
- Smith, R. B., O'Shea, T., O'Malley, C., & Taylor, J. S. (1989). Preliminary experiments with a distributed, multimedia problem solving environment. *Proceedings of the First European Conference on Computer Supported Cooperative Work* (pp. 19-35). London.

-
- Tang, J. C. and Minneman, S. L. (1991). VIDEOWHITEBOARD: Videoshadows to Support Remote Collaboration. *Proceedings of CHI '91* (pp. 315-322). New Orleans, Louisiana.
- Watabe, K., Sakata, S., Maeno, K., Fukuoka, H. & Ohman, T. (1990). Distributed multiparty desktop conferencing system: MERMAID. *Proceedings of the Conference on Computer Supported Collaborative Work* (pp. 27-38). Los Angeles, California.
- West, C. (1985). *Routine Complications: Tasks and Troubles in the Medical Consultation*. Bloomington Indiana: Indiana University Press.