In Finn, K. Sellen, A. J. and Wilbur, S. B. (eds.), *Video-Mediated Communication*. Lawrence Erlbaum., pp. 323-347. 1997.

Reconfiguring Media Space:
Supporting Collaborative Work

Christian Heath
Paul Luff
University of Nottingham

Abigail Sellen
MRC Applied Psychology Unit, Cambridge
Rank Xerox Research Centre Cambridge EuroPARC

ACKNOWLEDGEMENTS

Part of the work reported in this paper was supported the EC RACE Programme (R2094) under the auspices of a research project entitled MITS (Metaphors for Telecommunication Services). The MTV experiments were carried out in collaboration with William Gaver of the Rank Xerox Research Centre Cambridge EuroPARC (now at the Royal College of Art). We are also grateful to Wendy Mackay for assisting with the design of the MTV experiments. We would also like to thank Mike Molloy, Paul Dourish and others at Rank Xerox Research Centre Cambridge EuroPARC, as well as Michel de Fornel, for their support and ideas during our studies of media space.

INTRODUCTION

Advances in telecommunications will undoubtedly have a profound impact on organisational life and collaborative work over the next decade. Their ability to enhance and transform distributed activities, as well as enriching how people work when with each other, has been well documented and we wait with some impatience to witness the extraordinary contribution of such technologies to our ordinary lives. Despite the optimism that greet successive innovations in telecommunications, it is not at all clear whether current developments provide satisfactory support for even the most simple or apparently straightforward collaborative activities. Indeed, the debates which have arisen concerning the actual contribution of experimental systems when deployed in organisations such as research laboratories, reveal perhaps not only the potential shortcomings of the technology, but our lack of understanding of the ways in which it might contribute to interpersonal communication and collaborative work. Even the very basic question as to the advantage of audio-visual communication over basic telephony remains subject to debate and curiosity. In fact, we have been reliably informed that in a large scale study of the use of video telephones by domestic users in a city in South Western France, many subscribers preferred to look at themselves whilst on the phone rather than the person with whom they were talking.

As yet therefore, we have relatively little understanding of the characteristics of video-mediated communication or the contribution that audio-visual technologies including telecommunications might provide for collaborative work. In this brief chapter we wish to discuss the contribution and design of "media spaces": computer-controlled networks of audio and video equipment intended to support collaboration among physically distributed colleagues. In particular, by exploring the actual use of media space technology and comparing the support it provides to the resources that people ordinarily rely upon when working together in more conventional environments, we wish to consider the requirements for developing more satisfactory technological environments for collaborative work. These requirements form the basis to a number of experiments in which we develop and evaluate prototype support for working together at a distance.

VIDEO-MEDIATED CONDUCT

EuroPARC's Media Space: Background and Setting

In common with several other system research laboratories, Rank Xerox have in place an audio-visual infrastructure in their EuroPARC laboratory in Cambridge. This infrastructure allows scientists and administrative staff to establish visual and audible contact with each other, or to view public areas such as the commons area and the conference room. EuroPARC's offices straddle three floors, and in part the technology was introduced to facilitate informal contact and sociability between organisational personnel. The system basically consists of a camera, 14" monitor, speaker, and microphone in each office, with larger monitors in the public areas. The monitor with camera is typically placed on top is positioned to one side, roughly at a 120 degree angle, to their workstation (Figure 1). A flat PZM, multi-directional microphone is normally positioned on the desk by the workstation, operated by a foot pedal.

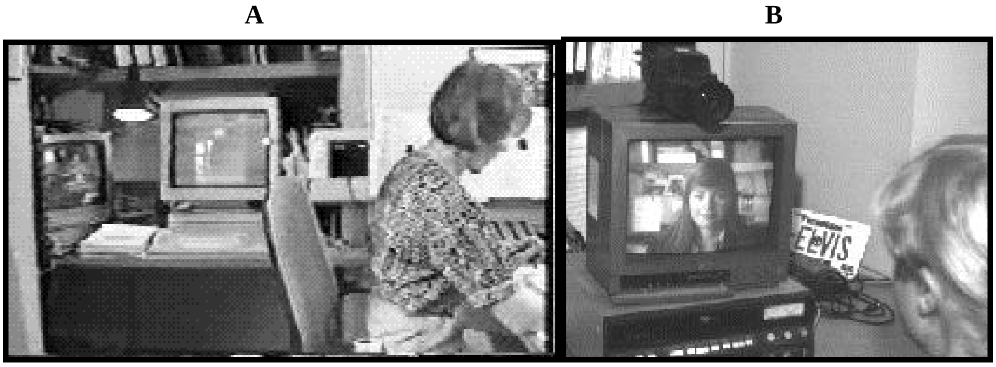


Figure 1. Two offices using the audio-visual infrastructure at EuroPARC. In A, the camera and monitor are to the left of the workstation and the microphone is multi-directional, consisting of a small, flat metal plate on the wall, operated by a foot switch. B gives more detail of the common relationship between camera and monitor positions.

Over the past three years the infrastructure has become increasingly sophisticated, and we have experimented with various alternative configurations 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 & Travers, 1991; and Gaver, Moran, Maclean, Lovstrand, Dourish, Carter & Buxton, 1992). Despite these technological developments, the most prevalent use of the system within 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). They provide two individuals based at different parts of building with continual video access to each other. Audio connections are normally switched off until one of the participants wishes to speak with the other.

Methodological Considerations

As part of the introduction and development of the EuroPARC media space, we undertook audio-visual recording of connections between individual offices. To diminish the potential influence of recording on the way people used the system, and to enable us to gain an overall picture of how frequently and for what purposes individuals used connections, we undertook "blanket" recording of a particular connection for up to two or three weeks. This data corpus was augmented by more conventional 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.

Whilst our analytic orientation to the audio-visual materials and field observations was relatively catholic, it drew in part from recent developments in the social sciences, in particular, ethnomethodology and conversation analysis. Our central concern was with the accomplishment of ordinary activities in and through media space technology and in particular the ways in which the participants themselves produced and recognised visual as well as well as vocal actions within the developing course of their interaction. Thus, nonverbal behaviour is not treated in isolation from the talk with which it occurs or the context in which it arises. Conduct in interaction, whether visual, vocal, or a combination of both, is addressed with regard of the actions it performs *in situ* within the local configuration of activity. The meaning, or better, the sense of a particular actions is embedded in the context at hand, accomplished in and through a social organisation which provides for the production and intelligibility of activities in interaction.

In examining video-mediated communication, we had a particular interest in the ways in which participants were able to organise each others' involvement in the course of various activities and to coordinate their vocal and visual actions. We were also driven by an interest in various substantive concerns such as: how individuals established mutual engagement; the extent to which they were able to remain (peripherally) aware of each others' activities and immediate environments; and whether video connectivity provided a suitable medium for accomplishing object-focused (such as screen or paper-based) collaborative tasks. Analysis developed on a case by case basis, in which we began by transcribing particular fragments of data, identify potential phenomena such as sequences of actions, and assembled collections of candidate instances. Comparing and contrasting the organisation of specific activities across numerous instances found both in data of video-mediated communication as well as more conventional face to face interaction, provided ways in which we could begin to delineate a body of observations and findings (cf. Heath and Luff, 1991; Heath & Luff, 1992b).

Observations

Given the various, sometimes contradictory arguments concerning the significance of video to communication between physically distributed individuals, it is perhaps worth beginning by summarising the conclusions from our analysis concerning the contribution of real-time visual access to informal sociability and collaborative work. There are three such contributions worth mentioning.

The first is that, unlike a telephone or audio connection, at EuroPARC video provides the opportunity for individuals to assess visually the availability of a colleague before initiating contact. More precisely, the video channel not only allows an 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 initiate contact. The infrastructure supports 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.

Second, 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 some 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 designers have begun to develop shared real-time interfaces (cf. Bly, 1988; Olson, Olson, Mack & Wellner 1990). Recent experiments (Olson & Olson, 1991; Smith O' Shea, O' Malley, & Taylor, 1989) have demonstrated the importance of providing video for participants to coordinate simultaneous screen-based activities.

Third, the video channel provides participants in multi-party conversations with the ability to recognise who is speaking and to "track" the thread of the conversation. This is of particular importance where video-conferencing facilities support multi-party interactions and where each connection involves more than single participant. In our analysis of multi-party audio-visual connections both at EuroPARC and the Xerox Video-conferencing facility at Welwyn Garden City, we noted that video plays an important part in the allocation and coordination of speaker turns. The advantages of video in helping to identify and discriminate amongst speakers is also supported by experimental studies of multi-party video-conferencing systems (Sellen, in press).

Despite these contributions, our observations of video-mediated communication suggest that this kind of technological medium provides a communicative environment which markedly differs from physical co-presence. In the following we will sketch some of the more significant differences between human conduct performed through technological media and actions and activities undertaken in face-to-face settings.

The insignificance of a look

In recent years a growing body of research has noted the ways in which looking at another person not only serves to provide certain information, but is itself a social action which engenders a response from the person who is being looked at (cf. Kendon, 1990; Goodwin, 1981; Heath, 1986). For example, all of us have been aware of being looked at by another within in a public setting such as restaurant or a train, and felt the discomfort that another's gaze can cause. We are perhaps less aware of the ways in which we attempt to distract the other and avoid their gaze, by becoming 'preoccupied' with a book we are reading or shielding our eyes with a gesture. On the other hand, a look may serve to encourage another to return their gaze, and having established a mutual orientation, allow the participants to move progressively into conversation or more generally focused interaction. Indeed, in research we conducted some years ago concerned with the medical consultation, we found the doctor would often initiate the business at hand in direct response to the patient turning towards him. In those particular cases, in the passing moment between the preliminaries of the consultation and discussing the patient's reason for visiting the doctor, a momentary look was enough to engender the doctor's initiating utterance. More generally, it has been found, throughout a range of ordinary settings, that turning and looking at another serves to elicit a response and most frequently a return of gaze from the person who is being looked at. Turning towards another, like other forms of bodily conduct like talk can be sequentially or interactionally implicative, generating a position (in immediate juxtaposition) where a particular action(s) is relevant, and if absent is 'noticeably' or 'accountably' absent (cf. Schegloff 1972).

One of the interesting aspects of the ways in which gaze can serve to engender action, is that the person who is looked at, and responds, is not necessarily looking at the other. Whether in conversation or simply walking down the street, individuals are able to monitor peripherally the local environment, and in particular, others within the local milieu, and where necessary or appropriate, respond. Indeed, it would seem that our ability to remain sensitive to and monitor people's behaviour outside the direct line of our regard is a critical element of the ways in which we produce and coordinate our actions with each other, whether they are produced through visual conduct or talk.

When we began to look at video-mediated communication, there seemed to be some curious differences in the ways in which the participants looked at each and responded to each other’s looks. In particular, we began to notice that whilst people would turn towards each other in the way they might in a co-present setting, their looks would often pass unnoticed by the person at whom they were looking. Moreover, it was not that the other was simply ignoring the look, since we knew from our research on face-to-face interaction that in ‘declining’ another’s gaze the recipient would frequently produce various actions to enable him to avoid the gaze of the other. Rather, in the cases at hand, one person would simply not notice that the other was looking at him, even where the other person upgraded, or exaggerated the look. The social and interactional significance of the look appears to be undermined by the technology; the look loses its sequential relevance.

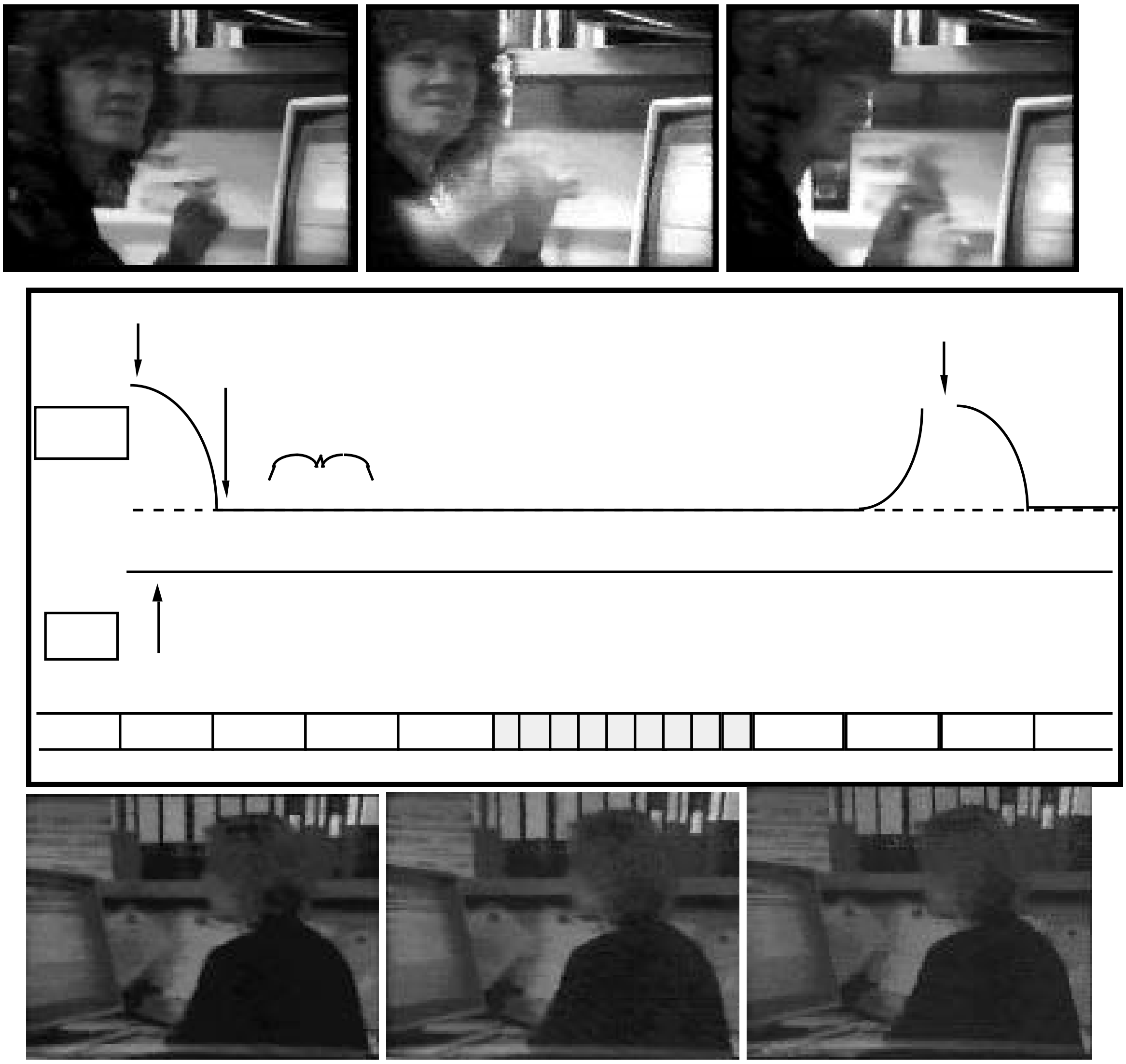


Figure 2. A fragment of activity between a scientist, Maggie (top), and a member of administrative staff, Jean (bottom). For presentation purposes, the gaze direction of each of the participants is indicated by a line. Gaze towards another is shown by the line moving towards a central dashed line; gaze away by their line moving away. In the above fragment, Jean’s gaze remains fixed on the screen whilst Maggie first looks towards Jean and then, several seconds later, turns away.

For example, in the fragment illustrated in Figure 2 drawn from a recording of a video connection between a scientist and a member of the administration at EuroPARC, we find Maggie, the scientist, attempting to initiate contact with Jean. To do this, 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 fragment 1, therefore 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 and their accompanying gestures, 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 this, 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. It is as if the sequential significance of such actions is weakened 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.

In consequence, at least during the initial introduction of the media space into EuroPARC, we found that users frequently had to resort to more formal or explicit ways of initiating mutual engagement. Unlike co-presence where participants can delicately and progressively move, step by step into a state of focused interaction, users would summon the other with a noise, as you might when calling someone on the telephone.

The articulation of talk and recipient insensitivity

Looking at another not only serves to initiate interaction, but plays an important part in the production and coordination of talk within an encounter. We have already mentioned how a look may encourage another to talk; gaze serving to display that the participant is not simply available but is also prepared to listen or ‘receive’. During the course of talk speakers themselves are sensitive to the gaze of the person to whom they are talking and draw various inferences from the direction of the other’s gaze concerning their involvement in the activity at hand. Indeed, it has been found that speakers have various devices for encouraging a co-participant to turn towards the speaker. These devices include speech perturbations such as pauses, the elongation of sounds, and various forms of self-editing and repair. They also include body movements such as gestures (cf. Goodwin 1981, Heath 1986). A critical element of the use and success of these various devices, turns on the speaker’s ability to encourage the recipient to reorient, by looking at the co-participant. The speaker’s gaze in many cases works with these various devices to establish a reorientation from the co-participant and thereby establish heightened involvement in the activity at hand.

The gaze of the speaker and the person to whom he is speaking therefore is consequential to the production of talk. For example, the speaker may delay the onset of an utterance until they have secured the gaze of the recipient. Or, the speaker will stall the production of an utterance, withholding the gist of the talk, until they establish a reorientation from the recipient. Or, in some cases, a speaker will abandon the projected course of an utterance, and even a sequence of utterances such as a narrative, in the light of the failure to establish an appropriate orientation from the person to whom they are speaking. The production and articulation of talk in face-to-face interaction is embedded in, and interrelated with, the visual conduct of the participants; both talk and body movement inform the accomplishment of action and activity in interaction.
In consequence, the ineffectiveness of a look in video-mediated co-presence can be consequential for the production of talk and more generally the interaction between the participants.

In the following example we find the relative ineffectiveness of looking generating difficulties for the emergence of a conversation between two scientists at EuroPARC who are discussing a networking problem. We join the action as Ian initiates contact with Robert by enquiring what he should tell Marty, a colleague in the United States, to do:

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::ff:).
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 hak)

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; indeed, the gist of the reply appears 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 in particular his apparent difficulty in beginning his reply may be systematically related to the conduct of the (potential) recipient, and in particular with Robert's inability to secure his co-participant's gaze. The more detailed picture in Figure 3 might be helpful.

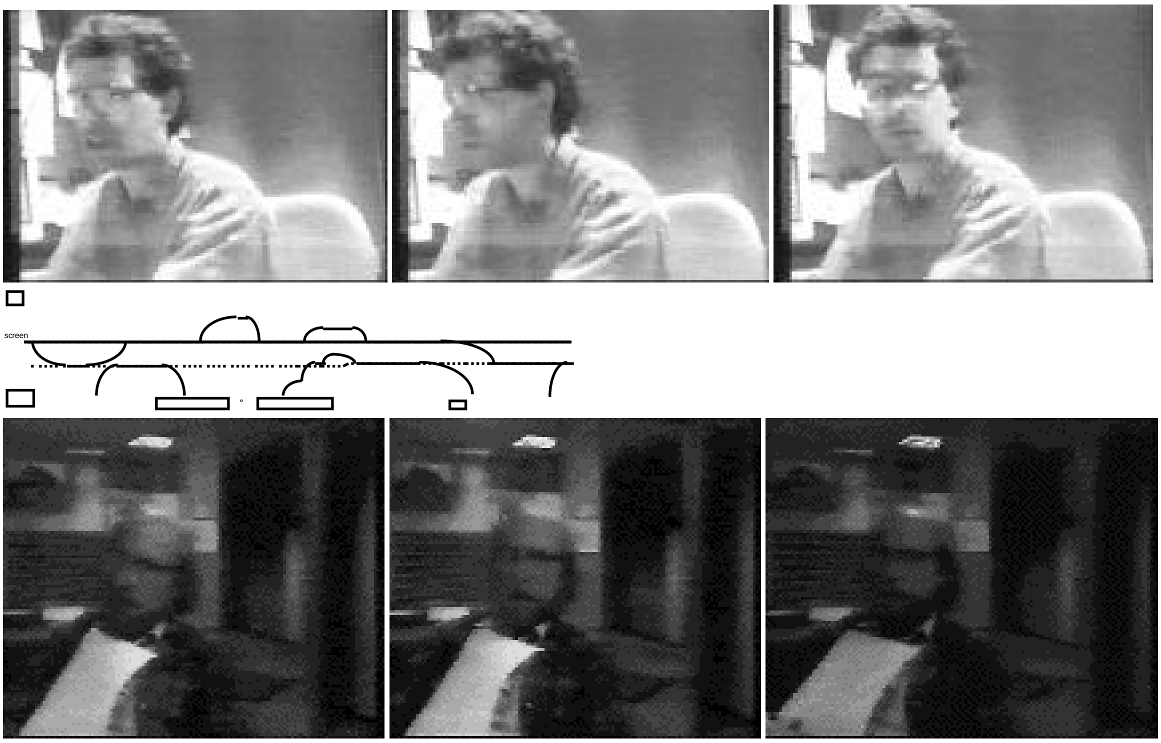


Figure 3. A fragment of interaction between Ian (top) and Robert (bottom) over a video-mediated "office-share" connection. Note, Ian's and Robert's talk is presented above and below the gaze line respectively. Intervals between utterances are in tenths of a second.

Withholding the reply fails to engender any reorientation from Ian, and following "Er:°m::", Robert begins progressively to 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 recipiency 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. The pause appears to engender a response from

Ian, and following his realignment of gaze from the screen to his colleague, the speaker begins the gist of his reply.

It is apparent therefore that, in this example, 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. The difficulties faced by the speaker in attempting to secure a realignment from the recipient may 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. Unlike face-to-face the relative scale and presentation of a speaker's more delicate shifts in orientation and gaze in a media space appear to pass unnoticed and thereby undermine the interactional significance of conventional devices to establish mutual orientation.

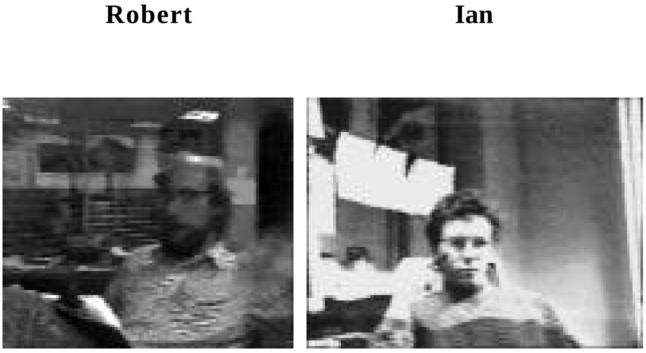
In passing, a further point should be mentioned. To provide individuals with the ability to vary their position whilst speaking with colleagues through the media space, we deliberately used multi-directional microphones to provide audio connections. These multi-directional microphones are designed to conceal relative changes in the direction of a sound within a circumscribed domain. In consequence, they mask changes in the sound level of the voice as a speaker changes his or her orientation. These changes may allow another to discern whether his colleague is changing his physical orientation, for example when the other is turning towards him. Thus, the relative ineffectiveness of a speaker's shift of gaze to engender response during the course of utterance, may not only derive from the accessibility of their visual conduct, but also from the absence of changes in tone and loudness of the voice.

The impact of gesture

Other forms of bodily conduct movements, ranging from relatively gross shifts in postural orientation through to minor head nods and the like also accomplish, often in combination with talk, social actions and activities. They are 'locally' or contextually significant, serving to engender or provide opportunities for particular actions from a co-participant. We have already mentioned, for example, how gestures can be used to elicit the gaze of the person to whom one is speaking, and how the utterance itself may be delayed until the gesture's accomplished its particular work. In some cases we find gestures accomplishing two or three actions simultaneously; the sequential relevance of the movement engendering particular actions from the co-participant at different points within the emergent of the activity. The significance of the participants' bodily conduct is accomplished within the developing course of the interaction and achieves its particular sequential significance then and there within the local configuration of action.

In video-mediated interaction, speakers use gestures as they might in face-to-face interaction. For example, in the following we find a lengthy description of an interface in which the speaker appears to encourage his recipient to participate

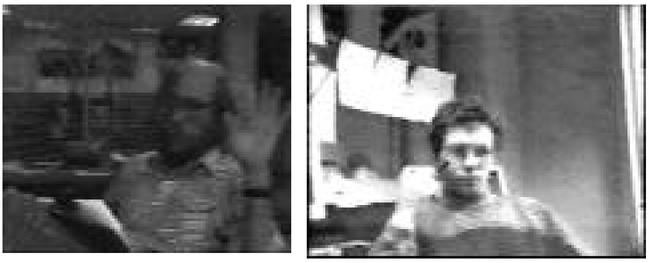
1 R: there's: two degrees of
2 freedom you can move it in
3 X an Y:::.



side to side gesture

4 (0.3)

5 R: if there are mo:re than
6 two degrees of freedom
7 you can select which
8 variables
9 were to be manipulat<u>ed:</u>



open palm movement

10 (0.5)

11 R: which (will)
12 remain fixed ← *flat palm movement*
13 (0.3)
14 R: and then manipulated at (.) ← *glance towards Ian*
16 two:: (.) <three variables ← *finger movement*
17 by the control: icon.
18 (1.2)
19 R: er:: is: this correct
20 (1.0)
21 I: Well:: (.) not quite.

Figure 4. A transcript of a conversation over an office-share connection between Ian and Robert.

The description itself is accompanied by a series of iconic or illustrative gestures through which Robert shows the operation of an icon in the interface. These include a side to side gesture occurring over "X an Y:::." (line 2) followed by an open palm movement from side to side with "were to be manipulat<u>ed:</u>" (line 6). Robert moves his palm down and flat as he utters "remain fixed" (line 8) and moves his second and third fingers down as he says "two:: (.) <three variables" (line 8). These gestures illustrate the ways in which the variables might be

manipulated. Robert has the gaze of the recipient during much of this extract but only turns towards his co-participant during the final part of the description.

On the one hand, the speaker's gestures appear to be designed to provide a visual portrayal of the objects and actions mentioned in the talk (see for example, Birdwhistell, 1970; Bull, 1983; Ekman and Friessen, 1969; Schegloff, 1984). There is little evidence however, either in this fragment, or in numerous other instances of iconic gestures in video-mediated interaction, that the illustrative component successfully provides the co-participant with relevant or sequentially significant information. On the other hand the gestures also appear to be designed to encourage Ian to participate more actively in the description. They fail however to transform the way in which he is participating in the talk; indeed he provides little indication, despite various opportunities and encouragements to display that he is actually following the emergent description. Consequently, the speaker, who has been unable to encourage the recipient to indicate whether he agrees, disagrees or fails to follow the description, is then faced with having explicitly to elicit confirmation and clarification.

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, 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, tend to be absent from the materials at hand. Speakers continue to gesture and produce a range of bodily behaviour during the delivery of talk in video-mediated communication. Yet visual conduct largely fails to achieve its performative impact or sequential significance.

Asymmetries in Video-Mediated Conduct

The foregoing analyses suggest that social interaction within a media space reveals asymmetries 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 that in the medical consultation or in the courtroom, the distribution of communicative resources are particularly peculiar in video-mediated presence. For example, in institutional environments we find the incumbents of pre-established roles, such as doctor and patient, having differential access and influence to activity types throughout the course of an event. By contrast, in the materials at hand, the asymmetries parallel the categories of speaker and hearer and are in constant flux as the conversation and different forms of participation, emerge. The asymmetries undermine the very possibility of accomplishing certain actions and activities.

Despite providing participants with the opportunity of monitoring the visual conduct of the other, and gearing the production of an activity to the behaviour of the potential recipient, media space and video-conferencing systems can undermine the interactional significance of non-vocal action and activity. On the one hand, the speaker, or more generally the interactant, has visual access to his or her co-participant. The speaker is able to monitor how the co-participant behaves whilst speaking to him or her and remains sensitive to the recipient's behaviour, state of involvement, and such like. But on the other hand, the resources upon which a speaker ordinarily relies to shape the ways in which a co-participant listens and attends to the talk appear to be interfered with by the technology. In video-mediated interaction, the speaker has visual access to the other but may be communicatively impotent.

Incongruent Environments of Action

Asymmetries in video-mediated conduct appear to arise in the light of two, interrelated issues.

Firstly, 'recipients' have limited and distorted access to the visual conduct of the other. The other's conduct is available on a monitor which not only distorts the shape of a movement, transforming its temporal and spatial organisation, but also presents the image of the other *in toto*. It destroys the relative weighting of different aspects of an individual's conduct. Moreover the presentation of the other on a conventional monitor undermines the possibility of peripherally monitoring the different aspects of the co-participant's conduct.

Secondly, an individual's limited and distorted access to the other and the other's immediate environment, undermines their ability to design and redesign movements such as gestures in order to secure their performative impact. These problems become more severe when one recognises that in contrast to physical co-presence, a person undertaking an action, such as speaking, cannot change their own bodily orientation in order to adjust their perception of the recipient and the local environment. The speaker is unable to see how his or her actions appear to the other, and in consequence, has relatively few resources to enable him or her to modify conduct in order to achieve a performative impact. It is not surprising therefore, that in reviewing the data corpus, one finds numerous instances of upgraded and exaggerated gestures and body movements as speakers attempt to achieve some impact on the way that others are participating in the activity, literally, at hand.

Our observations of EuroPARC's relatively sophisticated media space, as well as more conventional video-conferencing systems, have shown that such systems provide users with incongruent environments in which to communicate and collaborate. 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. This presupposition of a common frame of reference and a reciprocity of perspectives is, as Schutz (1962) and others have pointed out, a foundation 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.
>
> (Schutz, 1962, pp. 315-6)

In video-mediated presence, camera and monitor inevitably 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 production and receipt of visual conduct and suggests why gesture and other forms of bodily activity may be ineffectual when mediated through video rather than undertaken within a face-to-face, co-present environment.

COLLABORATIVE WORK IN ORGANISATIONAL SETTINGS

The difficulties that people have when using media spaces, and perhaps the general lack of enthusiasm for the technology, becomes clearer when one considers how collaborative work is organised in more conventional environments. Alongside our research of media space use, we have undertaken a series of naturalistic studies of work, interaction and technology in a range of organisational settings. These settings include line control rooms in London Underground (Heath & Luff 1992a, forthcoming), primary health care (Greatbatch, Luff, Heath & Campion 1993), architectural practices (Luff & Heath 1993), and news agencies (Heath & Nicholls, forthcoming).

Whilst the settings encompass a broad range of tasks and technologies, the studies reveal that collaborative work in more conventional environments appears to reveal generic features which are relevant to how individuals organise their own activities and coordinate their own contributions, in real time, with others. So for example, if we take the London Underground Control Rooms we find that personnel have developed a body of informal and tacit practices for distributing information to each other and coordinating simultaneously, multiple activities. These practices allow personnel to distribute information to colleagues and to monitor each other's activities, whilst apparently engaged in a single, individual task. In this way, personnel coordinate their actions with each other and sustain a mutually compatible sense of the 'business at hand' whilst managing their own specific responsibilities within a complex division of labour. The studies reveal the ways in which co-present collaborative work relies upon a complex body of interactional practices through which personnel 'peripherally monitor' and participate in each others actions and activities. Indeed, in such settings, it becomes increasingly difficult to demarcate the ‘individual’ from the ‘collaborative’ as personnel mutually sustain multiple activities which ebb and flow within various forms of co-participation and production.

More generally the studies reveal various aspects of collaborative work which are of relevance both to the design and deployment of media space as well as understanding the current limitations of such technologies. Of particular relevance are the findings that:

- Both focused and unfocused collaboration is largely accomplished not through direct face-to-face interaction, but through alignment towards the focal area of the activity, such as a document, where individuals coordinate their actions with others through “peripheral monitoring” of the others involvement in the activity “at hand”. For example, much collaboration is undertaken side by side where the individuals are continuously sustaining a shared focus on an aspect of a screen or paper based document, such as a section of an architectural drawing;
- Collaborative work relies upon individuals subtly and continuously adjusting their access to each others’ activities to enable them to establish and sustain differential forms of co-participation in the tasks “at hand”;
- Collaborative work involves the ongoing and seamless transition between individual and collaborative tasks, where personnel are simultaneously participating in multiple, interrelated activities;
- An individual’s ability to contribute to the activities of others and fulfil their own responsibilities relies upon peripheral awareness and monitoring; in this way information can be gleaned from the concurrent activities of others within the “local milieu”, and actions and activities can be implicitly coordinated with the emergent tasks of others;
- Much of the interaction through which individuals, produce, interpret and coordinate actions and activities within co-present working environments is accomplished using various objects and artefacts, including paper and screen-based documents, telephones, and the like. The participants activities are mediated and rendered visible through these objects and artefacts.

These observations stand in marked contrast to the support that media space provides for collaborative work. It becomes increasingly apparent, when you examine work and collaboration in more conventional environments, that the inflexible and restrictive views characteristic of even the most sophisticated media spaces, provide impoverished settings in which to work together. This is not to suggest that media space research should simply attempt to 'replace' co-present working environments, such ambitions are way beyond our current thinking and capabilities. Rather, we can learn a great deal concerning the requirements for the virtual office by considering how people work together and collaborate in more conventional settings. A more rigorous understanding of more conventional collaborative work, can not only provide resources with which to recognise how, in building technologies we are (inadvertently) changing the ways in which people work together, but also with ways in which demarcate what needs to be supported and what can be left to one side (at least for time being). Such understanding might also help us deploy these advanced technologies.

DEVELOPING THE WORK SPACE

On the basis of the research we have described, we can begin to discern the limitations of a media space, at least as it is currently conceived. In part, these problems emerge as a result of assumptions which appear to inform the design of media spaces and related technologies such as video telephones and video-conferencing systems. One such overriding assumption appears to be that a face-to-face, head and shoulders view is the most important for interpersonal communication and collaboration.

Yet, as we have discussed, face-to-face interaction is only one amongst a variety of cooperative activities which take place. Indeed, in many settings, face-to-face interaction constitutes a relatively small part of working together, and is one amongst a diverse configuration of spatial and bodily arrangements through which personnel participate in each others' activities and accomplish the "business at hand". In working together, individuals are continually and seamlessly shaping their participation and involvement in each other's activities as the demands of the task(s) and the interaction emerge. Individuals not only coordinate their actions with each other through various artefacts such as documents (whether on paper or screen) but continually adjust their access to each other and the tasks in which they are engaged. Perhaps the most important element of this interactional work, is the ways in which individuals monitor each other's involvement in, or alignment to, an object or artefact. It is not simply a case of seeing what another is seeing, but rather seeing the other in relation to what he or she is looking at and doing. We need to consider ways of expanding access to the remote participant's activities, taking into account the flexible ways in which people accomplish collaborative tasks.

To explore some of these issues and in particular to consider the consequences of various ways of providing users with increased access to the remote space, we decided to undertake a series of experiments. The ultimate aim of the experiments was to inform the design of a system in which people could flexibly vary their access to their co-participants and their co-participant's activity and working environment. To do this, in collaboration with our colleague, William Gaver, we began by constructing two simple systems offering variable, expanded access. The purpose of experimenting with these systems was to explore some of the interactional consequences of providing users with variable accessibility: to find out what the possible advantages and disadvantages of different design solutions might be.

The Multiple Target Video (MTV) Studies

One of the obvious ways of expanding access into another's domain is simply to increase the number of views a participant has of the remote environment. Thus, it should be possible to enhance the capabilities of media spaces by adding cameras and positioning these so that an individual has more than just a single face-to-face view of his or her co-participants. Several commercial multi-media systems provide such possibilities, often by having an additional "document camera" pointed down onto the desk. However, the research we have outlined has suggested that having two views may still be limiting, being too restrictive and not providing a sense of colleagues' orientation to particular activities.

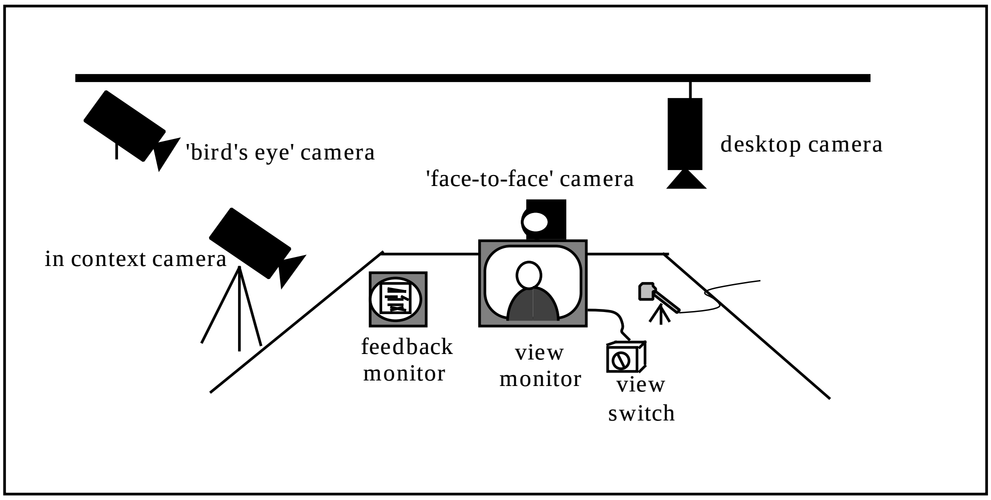


Figure 5. Schematic diagram of the configuration of the MTV I system using multiple cameras and a single monitor for each participant.

We therefore began by designing a system which offered more visual access: the Multiple Target Video (MTV) system. In the first of these experiments (MTV I), we offered different perspectives via four cameras: a conventional face-to-face view; a "desktop" camera to focus on the details of any activities on the work surface; a wider "in-context" view providing an image of the co-participant in relation to their work; and a "bird's eye" view giving access to the periphery of a colleague's environment. The difficulty for the design of such a system was how to display these different views to an individual.

In the first experiment, participants were given a single monitor to view their co-participant, and could change their view on the remote site by turning a knob. Thus, each participant could select for display on their monitor only one view at a time, doing so by sometimes momentarily passing through other views. To provide further information, each participant was also given a "feedback monitor" showing which view their co-participant had currently selected (see Figure 5).

In this experiment, the participants were given two tasks in collaboration with another colleague: the first was to draw a plan of their co-participant's office; and the second was to carry out a simple design task. In the design task, one of the participants, in the "design office", had a 3-D model of a room, complete with miniature furniture. The two participants' task was to agree on a layout for the room, subject to certain design constraints. The other person, in the "remote office", had to draw the final design. So that the individual in the remote office could see the model, one camera was used to focus on this model. This meant that the two participants did not have an identical range of views; only the individual in the "design office" had access to the bird's eye view.

The experiment revealed some interesting results with regard to the use of the four different views.[1] For example, the face-to-face view was rarely used in the accomplishment of either task. Instead, participants mainly switched between the in-context, model, desk, and bird's eye views. Given the nature of the design task it is perhaps not surprising that the participants in the "remote office" focused on the view showing most details of the model.[2] It appears that the in-context and bird's eye cameras afforded similar possibilities for the subjects in the design office, allowing them to assess their colleague's orientation and to make sense of particular aspects of visual conduct, for example, gestures pointing to objects. Finally, some participants appeared to make use of the switch to track their colleagues as they moved around the room. It also seems that the intermediate views that appeared whilst the user switched between settings provided opportunities to monitor aspects of a colleague's conduct and see, momentarily, other features in their domain.

However, the participants in the experiment did have difficulties with the system. As might be expected given the previous analysis, the MTV I system did not appear to alleviate the asymmetries revealed by our prior study of video-mediated communication. Participants still had difficulties when establishing and re-establishing engagement, with perturbations in talk accompanying the beginning of turns, and with designing their gestures in the course of the interaction.

Moreover, the possibility of having different views of each other's domain appeared to exacerbate problems associated with these asymmetries. Because each person could select from multiple views, participants appeared to be even more uncertain about what view the other had chosen at any point in time. This fact, combined with the possibility that the other's perspective could be transformed at any moment, further undermined the presumption of a reciprocity of perspectives. It is perhaps not surprising that participants appeared to have

[1] See Gaver et al. (1993) for more details of the method and results of these experiments.

[2] Although one 'remote' subject mainly opted for the in-context view that provided access to the co-participant in relation to the model.

difficulties both in achieving a common orientation to focused tasks and in managing the disengagement from collaborative activities. Participants often had to make apparent through talk their orientation to objects in the local environment and to the technology. For example, they had to reformulate directions given by their colleague, comment on their own movements and orientations and explicitly attempt to establish what the other could see. It seems that variable access to the other and their respective domain can make it more difficult for participants to preserve a sense of the (shifting) perspective of the other and to thereby coordinate their actions and activities with the contributions of their colleagues.

Interestingly, ascertaining the other's perspective did not seem to be alleviated by the presence of the feedback monitor, through which it was possible to find out what the other person was viewing. Instead, participants appeared to make use of this to refer to objects in their own domain, pointing to objects on the screen (such as their own documents) rather than pointing to those same objects in their local environment. Of course, this strategy was prone to difficulties, as pointing to the image on the feedback monitor was difficult, if not impossible for the remote person to see.

In addition to interfering with the establishment of a common frame of reference within which to work, the need to switch between views also appeared to be problematic because it precluded the ability to make a smooth and natural transition amongst views. Having to think about and negotiate the switch usually meant a break in the ongoing stream of activity. This may have presented enough of a distraction to discourage full use of the various views.

Whilst maintaining the effort to provide variable accessibility to another's domain, one possible solution to these difficulties was to remove the need for sequential access via a switch. In a second study, we used a system (MTV II) where each participant had multiple monitors each connected to a different camera in their colleague's office (Figure 6). In order to provide a symmetrical range of views for both participants, each person had three monitors showing an in-context view, a desk-top view and a face-to-face view. The desktop view could be used either for viewing the model or a document. The wide angle previously provided by the bird's eye view was, in effect, provided by a wide angle, in-context view. The three monitors were arranged in both rooms in a similar fashion with the face-to-face view in the middle. This meant that a orientation towards the face-to-face view would also appear to a co-participant as a reorientation away from their in-context or desktop view, and *vice versa*. As both participants had access to all views simultaneously, there was no need for a feedback monitor. The tasks in this study were the same as those used in the first study.

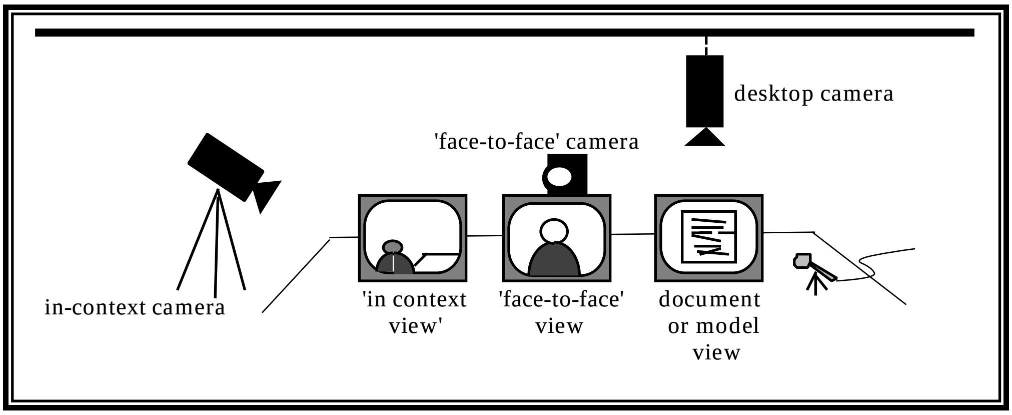


Figure 6. Schematic diagram of the configuration of the MTV system using multiple cameras and multiple monitors for each participant.

Preliminary observations reveal that participants used all the views, and their pattern of looking at the different monitors indicates that they "switched" views much more frequently than in MTV I. It may be that by removing the necessity of having to switch views manually, it allows participants to use the available views more fully. Taking advantage of a different view was less cumbersome and did not require a break in the activity at hand.

Although they used all three views, participants oriented mainly to the desktop monitor when engaged in focused collaborative work, either to view documents or the model. However, the face-to-face view was found to be used much more frequently than in MTV I. During the course of the interaction participants would often glance briefly at each other between more prolonged sequences of looking at the model or document. Thus, although the face-to-face view was not looked at for long periods of time, it appeared to be pivotal as subjects tended to return to it frequently. Often, co-participants utilised the view in order briefly to glance at their colleague. At other times, and especially after periods of prolonged silence, the face-to-face view was utilised with the in-context view to assess the other person's involvement in the activity.

Despite the advantages of having multiple views simultaneously available, the MTV II system still presented problems for the participants. Separate, fixed cameras still failed to offer complete access to the remote space. In addition, there still appeared to be difficulty in ascertaining the other person's perspective as evidenced by conversations in which participants attempted to make their orientation to the other explicit. The lack of access to a shared space for working, such as the inability to point to a shared document or artefact also presented a problem. These findings point out the issues which such simple experimental configurations fail to address, and clearly need to be considered in the design of future technologies.

Whilst the MTV experiments must be regarded as only preliminary attempts to build a more flexible environment for collaborative work, they are important in that they represent a break from the assumption that collaboration and communication is mainly face-to-face. This is an assumption that we believe has led to the impoverishment of much media space research. Multiple views of another's environment do appear to provide facilities for undertaking more complex and wide ranging collaborative work. However, the MTV studies have shown that design solutions which introduce variable access do this at the risk of imposing their own set of problems, as well as accentuating problems arising from the discontinuities and incongruencies that tend to be present in media space. It is important, however, to recognise these underlying problems, and to take them into account when envisaging how we might reconfigure media space. A more thorough understanding of the complex skills and competencies used by individuals in undertaking collaborative work can not only serve as a benchmark with which to evaluate our crude attempts to develop technologies, but provide an important resource in envisaging more innovative systems.

SUMMARY

In the light of research concerned with interaction and collaboration both in a media space and more conventional working environments, it is not surprising that recent developments in video-mediated communication have not met with the success that we might have envisaged. Whilst we will undoubtedly find markets for the video-telephone and for video-conferencing facilities, as well as more sophisticated forms of connectivity, we may have to go some way before we develop a technology which can support more complex forms of collaboration both within and across organisational environments. The limitations of current audio-visual infrastructures do not simply derive from the asymmetries that they inadvertently introduce into interpersonal communication. Indeed, as we have suggested elsewhere (Heath and Luff 1992b), a certain insensitivity to another with whom one is intermittently working may have certain advantages in fulfilling one's own individual responsibilities. Rather, audio-visual connectivity, including the development of more sophisticated media spaces, has been primarily concerned with providing physically distributed personnel with face-to-face views of each other, whereas collaborative work recurrently involves variable and contingent access, not simply to each other's physical domain and artefacts, but to the emergent activities in which the participants are engaged. Many forms of audio-visual connectivity, from basic videophones through to advanced media spaces have attempted to support interpersonal communication rather than collaboration and in consequence, perhaps, built relatively impoverished resources for working environments.

In attempting to address some of the issues concerning the requirements for and development of a media space, we inevitably confront some of the underlying 'difficulties' which give rise to the peculiar forms of interpersonal conduct we find in communication mediated through audio-visual technologies. The relative insignificance of a look, the impotence of gestural activity, and their consequences for the articulation of talk, derive not only from the distorted presentation of the co-participants to each other, but also from the incongruent interactional environments provided by a media space. As Dourish, Adler, Bellotti & Henderson. (1994) have recently argued, some of the difficulties which derive from this incongruence may be dealt with by an informal culture emerging amongst frequent users to manage the problems which arise in the operation of the system. This is undoubtedly the case. And yet, the interactional asymmetries which arise by virtue of the incongruent environments provided within a media space remain, and become increasingly severe, as we attempt to develop more sophisticated technologies to support collaborative work. Indeed, as the MTV experiments demonstrate, the more we try to develop a technological infrastructure to support variable access to each other's activities and the environments this necessitates, the more we can generate difficulties for users. Rather than rely upon the emergence of *ad hoc* informal culture to manage these problems, we need to explore systematically ways in which we can provide personnel with the resources with which to collaborate and participate, where necessary, in each other's activities, whilst unobtrusively preserving the individual's sensitivity to his own and his colleague's environment. If we can begin to address these issues and build a technology which can support task-based interaction and collaboration, then we might well be surprised with the impact of media spaces and telecommunications on work and organisational life.

REFERENCES

Birdwhistell, R. L. (1970). Kinesics and Context: Essays on Body Motion Communication. London: Allen Lane.

Bly, S. A. (1988). A Use of Drawing Surfaces in Different Collaborative Settings. In CSCW '88, 26th-28th September, (pp. 250-256). Portland, Oregon:

Borning, A., & Travers, M. (1991). Two Approaches to Casual Interaction over Computer and Video Networks. In G. M. Olson, J. S. Olson, & S. P. Robertson (Ed.), CHI '91, (pp. 13-19). New Orleans, Louisiana: ACM Press.

Bull, P. (1983). Body Movement and Interpersonal Communication. Chichester: John Wiley and Son.

Dourish, P., Adler, A., Bellotti, V., & Henderson, H. (1994). Your Place or Mine? Learning from Long-Term Use of Video Communication (Working Paper, Rank Xerox EuroPARC Cambridge.

Egido, C. (1990). Teleconferencing as a technology to support cooperative work: Its possibilities and limitations. In R. E. Kraut, J. Galegher, & 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.

Gaver, W. W., Moran, T., Maclean, A., Lovstrand, L., Dourish, P., Carter, K. A., & Buxton, W. (1992). Realizing a video environment: EuroPARC's RAVE system. In G. Bennett, G. Lynch, & P. Bauersfeld (Ed.), CHI 92, (pp. 27-35). Monterey, CA: ACM Press.

Gaver, W. W., Sellen, A., Heath, C. C., & Luff, P. (1993). One is not enough: Multiple Views in a Media Space. In INTERCHI '93, (pp. 335-341). Amsterdam: ACM.

Goodwin, C. (1981). Conversational Organisation: Interaction between a Speaker and Hearer. London: Academic Press.

Greatbatch, D., Luff, P., Heath, C. C., & Campion, P. (1993). Interpersonal Communication and Human-Computer Interaction: an examination of the use of computers in medical consultations. Interacting With Computers, 5(2), 193-216.

Heath, C. C. (1986). Body movement and speech in medical interaction. Cambridge: Cambridge University Press.

Heath, C. C., Jirotka, M., Luff, P., & Hindmarsh, J. (1993). Unpacking Collaboration: the Interactional Organisation of Trading in a City Dealing Room. In G. de Michelis, C. Simone, & K. Schmidt (Ed.), ECSCW 1993, (pp. 155-170). Milan: Kluwer.

Heath, C. C., & Luff, P. (1991). Disembodied Conduct: Communication through Video in a Multi-Media Office Environment. In G. M. Olson, J. S. Olson, & S. P. Robertson (Ed.), CHI '91, (pp. 99-103). New Orleans: ACM Press.

Heath, C. C., & 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., & Luff, P. (1992b). Media Space and Communicative Asymmetries: Preliminary Observations of Video Mediated Interaction. Human-Computer Interaction, 7, 315-346.

Heath, C. C., & Luff, P. (in press). Converging Activities: Line Control and Passenger Information on London Underground. In Y. Engestrom & D. Middleton (Eds.), Cognition and Communication at Work: Distributed Cognition in the Workplace Cambridge: Cambridge University Press.

Heath, C. C., & Nicholls, G. M. (forthcoming). Animating Texts: Selective Readings of News Stories. In L. B. Resnick & R. Saljo (Eds.), *Discourse, Tools and Reasoning: Situated Cognition & Technologically Supported Environments*

Kendon, A. (1990). Conducting interaction: Studies in the Behaviour of Social Interaction. Cambridge: Cambridge University Press.

Luff, P., & Heath, C. C. (1993). System use and social organisation: observations on human computer interaction in an architectural practice. In G. Button (Eds.), Technology in Working Order (pp. 184-210). London: Routledge.

Olson, G. M., & Olson, J. S. (1991). User-Centered Design of Collaboration Technology. Journal of Organisational Computing, 1(1), 61-83.

Olson, J. S., Olson, G. M., Mack, L. A., & Wellner, P. (1990). Concurrent editing: the group interface. In D. Diaper, D. Gilmore, G. Cockton, & B. Shackel (Ed.), Interact '90 - Third IFIP Conference on Human-Computer Interaction, (pp. 835-840). Cambridge: North Holland.

Schegloff, E. A. (1972). Notes on a conversational practice: formulating place. In D. Sudnow (Eds.), Studies in Social Interaction (pp. 75-119). New York: Free Press.

Schegloff, E. A. (1984). On Some Gestures' Relation to Talk. In J. M. Atkinson & J. C. Heritage (Eds.), Structures Of Social Action: Studies In Conversation Analysis (pp. 266-296). Cambridge: Cambridge University Press.

Schutz, A. (1962). Collected Papers I: The Problem of Social Reality. The Hague: Martinus Nijhoff.

Sellen, A. (1992). Speech Patterns in video-mediated conversations. In G. Bennett, G. Lynch, & P. Bauersfeld (Ed.), CHI '92, (pp. 49-59). Monterey, Ca.: ACM Press.

Sellen, A. (in press). Remote conversations: The effects of mediating talk with technology. Human-Computer Interaction.

Smith, R. B., O' Shea, T., O' Malley, C., & Taylor, J. S. (1989). Preliminary experiments with a distributed, multi-media problem solving environment. In First European Conference on Computer Supported Cooperative Work, (pp. 19-35). London: