

RETHINKING VIDEO AS A TECHNOLOGY FOR INTERPERSONAL COMMUNICATION: THEORY AND DESIGN IMPLICATIONS

Steve Whittaker¹,

Lotus Development Corporation

One Rogers St

Cambridge, MA.

02142, USA

Keywords: Interpersonal communications, computer mediated communication, task management, audio, video, conversation management, lightweight interaction

¹Current address: ATT Labs, 180 Park Ave, Florham Park, NJ, 07932, USA, email: steve@research.att.com

Abstract

This paper re-assesses the role of real-time video as a technology to support interpersonal communications at distance. We review three distinct hypotheses about the role of video in the co-ordination of conversational content and process. For each hypothesis, we identify design implications and outstanding research questions derived from current findings. We first evaluate the *non-verbal communication hypothesis*, namely the prevailing assumption that the role of video is to *supplement speech*, and embodied in applications such as videoconferencing and videophone. We conclude that previous work has overestimated the importance of video at the expense of audio. This finding has strong implications for the implementation of such systems, and we make recommendations about both synchronisation and bandwidth allocation. Furthermore our own recent studies of workplace interactions point to other communicative functions of video. Current systems have neglected another potentially vital role of visual information in supporting the *process* of achieving *opportunistic connection*. Rather than providing a supplement to audio information, video is used to *assess the communication availability of others*. Visual information therefore promotes the types of remote opportunistic communications that are prevalent in face-to-face settings. We discuss early experiments with such connection applications and identify outstanding design and implementation issues. Finally we discuss another novel application of video: "*video-as-data*". Here the video image is used to transmit information about the work objects themselves, rather than information about interactants, creating a dynamic shared workspace, and simulating a shared physical environment. In conclusion we suggest that research move away from an exclusive focus on *non-verbal communication*, and begin to investigate these other uses of real-time video.

Introduction

Interpersonal face-to-face communication is prevalent in workplace settings. For most office workers, interpersonal communication occurs often, and for many people such as managers it represents their *most frequent* workplace activity. Questionnaire data and observational data produce estimates of between 35% and 75% of time being spent in face-to-face interaction, with these figures depending on job type (Kraut, Fish, Rice & Chalfonte, 1993, Panko, 1992, Sproull, 1984, Whittaker, Frohlich & Daly-Jones, 1994a). While these studies document the frequency of interpersonal communication, they do not show its precise benefits or function. The importance of interpersonal communication is demonstrated by research into scientific collaboration, which reveals that physical distance between scientists' offices is a strong predictor of whether they will co-publish. The argument runs as follows: people who are physically collocated are more likely to communicate frequently and informally, and it is this which promotes effective collaboration. There is also evidence that interpersonal communications are crucial for specific types of workplace tasks. Questionnaire studies indicate that frequent, opportunistic, face-to-face conversations are vital to the planning and definitional phases of projects (Kraut, Egido & Galegher, 1990). Other work shows that even when people have a choice between different communication technologies, such as email, phone and face-to-face, they still choose face-to-face meetings for planning and definitional tasks. This suggests unique benefits of this style of interaction for certain classes of task² (Finholt, Sproull & Kiesler, 1990). Finally, questionnaire studies support the effects of proximity (and hence interpersonal communications), on social and organisational knowledge: Researchers are more likely to be familiar with, and to respect the work of colleagues who sit close to them (Kraut *et al.*, 1993).

This research shows the importance of interpersonal communication, but trends towards telework, mobile work and the globalisation of business are geographically separating workers. For coworkers separated by large distances, and in different time-zones, the potential for opportunistic face-to-face interpersonal communications is massively reduced. We therefore need technologies that can support interpersonal communication between geographically remote coworkers.

To support interpersonal communication at distance, however, we first need to understand the key components of face-to-face interaction. Face-to-face interpersonal communication requires speakers and listeners to co-ordinate both conversational *content* and *process* (Clark & Brennan, 1991, Clark & Schaefer, 1989, Grosz & Sidner, 1986, Walker, 1993, Whittaker, Brennan & Clark, 1991). *Co-ordinating content* involves the construction and maintenance of shared beliefs. Speakers and listeners therefore have to infer and monitor other participants' understanding (Clark & Brennan, 1991, Clark & Schaefer, 1989, Whittaker *et al.*, 1991), as well as their interpersonal attitudes or motivation (Mehrabian, 1971, Short, Williams & Christie, 1976). A key problem here is that individual utterances radically underspecify speakers' beliefs and intentions (Allen & Perrault, 1986, Searle, 1990). How then do listeners restrict their interpretations to the set of meanings that the speaker intended? A key constraint on listener inference is the *shared context*, which can take the form of (a) *linguistic context*, ie. the set of entities and events that are talked about as the conversation unfolds (Grosz & Sidner, 1986); and (b) *physical context* arising from the fact that participants share the same physical space and have access to approximately the same set of objects and perceptual events (Clark & Marshall, 1981). The second component of conversation is *process co-ordination*, which is concerned with the mechanics and management of conversation. Face-

²This was not true for *all* types of task, for example participants generally chose email rather than face-to-face interaction for scheduling and task assignment.

to-face conversations adhere to the rule that only one person generally speaks at any given time, so techniques are required to manage the process of speaker switching. Process co-ordination also addresses how participants initiate and close conversations. (Sacks, Schegloff & Jefferson, 1974, Walker, 1992, Walker & Whittaker, 1990). A critical feature of *process co-ordination* is that it takes place on a moment-by-moment basis, and requires very precise timing (Clark & Brennan, 1991, Clark & Wilkes-Gibbs, 1987, O'Conaill, Whittaker & Wilbur, 1993, Whittaker *et al.*, 1991).

Face-to-face conversation relies on voice, visual behaviour and gesture and thus employs multiple sensory modalities. In contrast, current pervasive communication technologies such as the phone only transmit voice. This paper therefore addresses one central question. What benefits can visual technologies such as video play in supporting key features of remote interpersonal communication, that are not supported by speech only communication? Our claim is that technological work has focussed on one specific function of visual information, to support *non-verbal communication* and neglected functions such as using visual information to *initiate communication* or depict *shared work objects*. According to the *non-verbal communication hypothesis*, visual information is claimed to be crucial for co-ordinating *communication content*: listener feedback about understanding of the speaker's message is supplied by head nods and eye gaze (Clark & Brennan, 1991, Clark & Schaefer, 1989, Kahneman, 1973); and interpersonal attitude can be inferred from facial expressions and posture (Argyle, Lefebvre & Cook, 1974, Ekman & Friesen, 1975, Kleinke, 1986). Furthermore, *nonverbal communication* is claimed to aid *process co-ordination*: head nods and eye gaze can support speaker switching and turn-taking (Argyle, Lalljee & Cook, 1968, Duncan, 1972, Kendon, 1967). This has implications for technology design. According to this view, technologies such as the phone and audio conferencing are thought to be inadequate, because they do not support this form of *non-verbal communication*. In contrast, technologies such as videophone and videoconference, supply visual information about the facial expressions, posture and gaze of participants, making them more effective media for interpersonal communication.

Recent research prototypes and video product sales might lead us to question the importance of *non-verbal communication*. First there is the commercial failure of products such as the videophone. This was originally launched in 1971 by A T & T to highly optimistic market forecasts, but it proved to be unsuccessful (Noll, 1992). The story is similar for videoconferencing, where commercial growth is gradual (Egido 1988, 1990). There are also substantial technical barriers to making video systems viable. If the application operates between remote sites, then the high data rates required by video mean that compression has to be used. The time taken to compress and decompress the video introduces transmission lags which severely impact interpersonal communication (Cohen, 1982, O'Conaill *et al.*, 1993; Tang & Isaacs, 1993, Whittaker & O'Conaill, 1993). Finally although the cost of video and audio hardware is dropping rapidly, such systems are still expensive to build, which means that clear benefits must be demonstrated before systems will be widely installed.

We first evaluate evidence for the utility of providing non-verbal communication information using video. We conclude that previous work has overestimated the importance of supplying non-verbal information at the expense of speech. This finding has strong implications for the implementation of such systems, and we make design recommendations about both synchronisation and bandwidth allocation, based on the importance of audio relative to video.

Furthermore, we argue that an exclusive focus on the *non-verbal communication function* of video has led researchers to ignore other aspects of communication in which visual information plays a crucial role. Current video systems have

focussed on how visual information contributes to pre-established ongoing conversations, but neglected another important *process co-ordination* function of visual information - its role in *initiating opportunistic connection* (Goodwin, 1981, Heath & Luff, 1991, Kraut *et al.*, 1993, Whittaker *et al.*, 1994a). Studies of workplace communication show the prevalence of opportunistic communication compared with arranged meetings (Kraut *et al.*, 1993, Whittaker *et al.*, 1994a). These results raise an immediate question: if the predominant form of workplace communication is opportunistic rather than arranged meetings, how do participants co-ordinate and initiate such meetings given their unplanned nature? Other data show that without visual information about the *communication availability of others*, connection failure is high. The phone is the communication device most used for remote opportunistic communications, but it does not allow one to see in advance whether the potential call recipient is available. This may contribute to the fact that more than 60% of business phone calls fail to reach their intended recipient (Rice & Shook, 1990, Whittaker *et al.*, 1994a). In contrast, studies of workers who share the same work area show the crucial role of visual information in allowing people to determine the availability of others for communication (Festinger, Schacter & Back, 1950, Kraut *et al.*, 1993, Mintzberg, 1973, Whittaker *et al.*, 1994a). Providing *availability* information using video should therefore increase the chances of successfully initiating an opportunistic connection with a remote coworker. We discuss experiments to test the *connection hypothesis* and identify outstanding design and implementation issues.

Finally we evaluate a different communication function of video: "*video-as-data*", where video is used to create a *shared physical context* or *shared workspace* between remote interactants, and hence support the *co-ordination of communication content*. Here the video image is used to transmit information about the work objects themselves, rather than non-verbal information about interactants. Studies of face-to-face workplace interaction show the crucial role of shared artifacts such as documents or drawings (Luff, Heath & Greatbatch, 1992, Mosier & Tamaro, 1995, Suchman, 1992, Tang, 1991, Whittaker *et al.*, 1994a). Co-ordination of conversational content is achieved by using these shared objects: to mediate shared attention; to support common reference and understanding; to record agreements and progress; and as a target for gesture (Clark & Brennan, 1991, Cooper, 1974, Minneman & Bly, 1991, Whittaker *et al.*, 1991, Whittaker *et al.*, 1993). Again, we discuss early experiments to test this hypothesis about "*video-as-data*", and identify outstanding design and implementation issues.

In conclusion, we suggest that research move away from an exclusive focus on *non-verbal communication*, and begin to investigate these two other communicative functions of real-time video, for *initiating opportunistic connection* and supporting shared objects, as part of a *shared physical context*.

The non-verbal communication hypothesis

Evaluation of the *non-verbal communication* hypothesis is complicated by the fact that three different claims have been made about the mechanisms by which visual information supports different *non-verbal* aspects of remote interpersonal communication. These claims are that the visual channel supports the transmission of: (1) *cognitive cues* that are used to determine remote participants' understanding, such as head nods and visual attention (Clark & Brennan, 1991, Clark & Schaefer, 1989, Kahneman, 1973); (2) *turn-taking cues* afforded by head turning, posture and eye gaze which support conversation management processes, such as achieving smooth transitions when there are changes of speaker (Argyle *et al.*, 1968, Duncan, 1972, Kendon, 1967); (3) *social or affective cues* that reveal remote participants' emotional state or interpersonal attitudes which are manifested in facial expression, posture and eye gaze (Argyle *et al.*, 1974, Ekman &

Friesen, 1975, Mehrabian, 1971, Reid, 1977, Short *et al.*, 1976). Cognitive and social cues address the issue of co-ordinating conversational content, whereas turn-taking addresses process co-ordination.

How then do we evaluate the non-verbal communication hypothesis? One way to evaluate video as a technique for providing non-verbal information has been to conduct short-term laboratory studies comparing video-mediated with: (a) audio or (b) face-to-face interaction, in the context of a particular communication task. The comparison with audio reveals how and when video information enhances speech only communication, and the comparison with face-to-face communication about how effectively video/audio mimics face-to-face conversation. Another technique has involved longer term field studies, installing video systems to evaluate their use (Abel, 1990, Bly, Harrison & Irwin, 1993, Gaver, Moran, MacLean, Lovstrand, Dourish, Carter & Buxton, 1992, Fish, Kraut, Root & Rice, 1992, 1993, Mantei, Baecker, Sellen, Buxton, Milligan & Wellman, 1991, Tang, Isaacs & Rua, 1994).

There are major methodological difficulties in carrying out evaluation studies, particularly in field settings. Most of the work has investigated local area systems, where it is easier to build and maintain high performance systems, and to conduct experiments on intact workgroups. Local area applications possess fewer technical limitations, such as networking bandwidth restrictions. In principle this means that it is possible to examine what might be possible in future wide area systems, when these networking constraints have been removed. In contrast, many current wide area systems suffer performance limitations such as lag, half duplex audio and poor quality video, which can severely limit conversational process co-ordination (Cohen, 1982, Tang & Isaacs, 1993, O'Conaill *et al.*, 1993, Whittaker & O'Conaill, 1993). As against this, a problem for local area field study evaluations is that the benefits of video to the user may be reduced in the local setting, because it is often possible to talk to the other party face-to-face. Evaluations of wide area systems therefore offer potentially better information, given that collaborators cannot fall back on face-to-face interaction, but they suffer from technical limitations. The consequence is that we have data either: (a) from local workgroups using high quality video systems, where the communication benefits may be reduced given the availability of face-to-face communication; or (b) from distributed workgroups who have higher incentives to use the system, but are using inferior technology.

SYSTEM EVALUATIONS

The prototypical systems here are the videoconferencing suite or videophone. Table 1 shows the results of system evaluations, for both high and low quality systems. We review evaluations of each of the three subhypotheses about the role of *non-verbal communication*, first for high quality and then for low quality systems.

Table 1 about here

Cognitive Cueing

Chapanis and colleagues (Chapanis, 1975, Chapanis, Ochsman, Parrish & Weeks, 1972) conducted a series of laboratory experiments testing the *cognitive cueing* hypothesis, namely that visual cues such as head nods and gaze help speakers to evaluate listener's understanding and attention. They compared the effectiveness of a variety of different media combinations for different cognitive problem solving tasks, by looking at *task outcome* measures such as time to solution and quality of solution. The tasks involved complex instruction giving and route planning. In one task, subjects had to jointly construct a mechanical object where one person had the physical components and the other had the instructions. In another task, one person was given a map and the other given a copy of the Yellow Pages. They were asked to identify a

map location satisfying a number of criteria, eg. the nearest dentist to a given street address. The research compared two media conditions: audio only communication, and high quality video/audio. If video does indeed provide useful cognitive cues, then there should be benefits for providing visual information in these types of collaborative problem solving tasks, where it is important to track the understanding and attention of remote participants. However, the studies showed that adding visual information did not increase the efficiency of problem solving, or produce higher quality problem solving. Furthermore, other experiments comparing different combinations of media indicated that *speech* was the critical medium for interpersonal communication in collaborative problem solving: removing the speech channel had huge effects on the outcome of communication. If participants could use the speech channel, then the addition or removal of video, text or writing media had little effect on task outcome or quality of solution. These results showing little impact of visual information on cognitive problem solving have been replicated by several other laboratory studies (Reid, 1977, Short *et al.*, 1976, Williams, 1977). Most importantly, this is not an issue of video quality: even face-to-face interaction is no better than speech only communication for this class of task (Williams, 1977).

Similar negative results are suggested by field study research. A study of high quality local area videophone conducted over several months in a research laboratory showed few objective usage differences compared with the telephone (Fish *et al.*, 1992)³. Phone and videophone calls have similar durations, and are used for the same set of communication tasks. The researchers also administered a questionnaire asking people to state the tasks for which they felt that different communication techniques (eg. videophone, telephone, face-to-face) were appropriate. Multidimensional scaling techniques applied to people's answers indicated that videophone is viewed by users as more similar to the telephone than face-to-face communication.

Turn-taking

The results are more mixed for the *turn-taking* hypothesis. Sellen (1992, in press) investigated this in a series of laboratory studies of negotiation tasks, in which groups discussed contentious issues and tried to reach consensus. She compared high quality video/audio systems, with both face-to-face and speech only communication. There was little evidence to support the claim that high quality video information improves conversation management and turn-taking, when compared with audio-only conversations. For objective conversation process measures such as pausing, overlapping speech and interruption management, there were no process differences between the video/audio systems and speech only communication. Furthermore, none of the video/audio systems replicated face-to-face conversational processes. The video/audio systems showed reduced ability of listeners to spontaneously take the conversational floor, as measured by number of interruptions⁴. Video/audio systems led speakers to use more formal techniques for handing over conversational initiative, such as naming a possible next speaker or using "tag" questions⁵, when compared with face-to-face interaction. Similar data are reported by O'Conaill *et al* (1993) and Whittaker & O'Conaill (1993), who also found speakers holding real meetings using high quality videoconferencing used more formal turn-taking techniques than were

³Many other recent field trials have investigated videophones, open links and media spaces (Abel, 1990, Bly *et al.*, 1993, Gaver *et al.*, 1992, Mantei *et al.*, 1991, Tang *et al.*, 1994), but few of these studies have explicitly addressed the enhanced audio hypothesis. Instead their focus has either been on the technical feasibility of building distributed video systems or alternatively on discovering novel uses of video applications such as *video for connection*. We review these novel applications in the next section.

⁴The ability to interrupt the speaker at any point of the conversation, eg. to ask a clarifying question, is regarded as a *positive* aspect of conversation, indicating spontaneous speaker switching (O'Conaill *et al.*, 1994, Rutter & Robinson, 1981, Sellen, 1992, 1994, Walker & Whittaker, 1990, Whittaker & O'Conaill, 1993, Whittaker & Stenton, 1988).

⁵ Examples are "isn't it?", "aren't they?", "couldn't you?" and involve an auxiliary verb and question syntax, at the end of a sentence.

observed in face-to-face interaction. Our explanation of the failure of even high quality videoconferencing to replicate face-to-face communication processes was that most videoconferencing systems do not support *directional* sound or visual cues. They tend to present sound and picture from a single monitor and speaker which may compromise sound direction, head turning and gaze cues in group interactions. We return to this as an outstanding research and design issue.

However, there are some differences in *subjective* data about *turn-taking* gathered from questionnaires addressing subject's impressions of the impact of video on conversational processes (Sellen, 1992, in press). Video/audio is perceived to be better than speech in a number of ways. It is perceived to: support interruptions; lead to more natural conversations that are more interactive; increase the ability to listen selectively to particular speakers; allow one to determine whether one is being attended to; and to generally keep track of the conversation. People also believe that they are better able to track the attention of others, when they have video. Similar qualitative data are reported by Isaacs and Tang (1993), who found that video seemed to allow participants to manage pauses better than in speech only communication. Despite this, Tang and Isaacs (1993) found that high quality video was again not perceived as equivalent to face-to-face interaction: *subjective* data showed that video was not seen as being as effective in supporting interactivity, selective attention, and the ability to take initiative in the conversation.

Social cueing

Despite the current lack of support for cognitive cueing or turn-taking functions of video for non-verbal communication, there is stronger evidence for the claim that video supports the transmission of *social cues* and *affective* information. Adding video information to the speech channel changes the outcome and character of communication tasks that require access to affect or emotional factors. Example tasks here include: negotiation, bargaining, and conflict resolution. Participants focus more on the motives of others when they have access to visual information, and video/audio conversations are more personalised, less argumentative, more polite, and broader in focus. They are also less likely to end in deadlock than speech only communications (Reid, 1977, Short *et al.*, 1976, Williams, 1977). These results can be explained in terms of affective cues: providing visual access to facial expressions, posture and gesture allows people to make inferences about other participants' affective or emotional state. There are also *subjective* benefits to providing visual information: participants believe that video/audio and face-to-face interaction are better than audio only for tasks requiring affect, such as getting to know other people, or person perception tasks. In addition, groups conversing using audio and video tend to like each other more (Reid, 1977, Short *et al.*, 1976, Williams, 1977).

LOW QUALITY SYSTEM EVALUATIONS

The preceding evaluations all used *high quality* audio and video. With the exception of *affect*, these studies reveal few objective advantages of adding high quality video to audio information. Current technology limitations and restricted networking bandwidth mean, however, that high quality systems will not be available for some years. It is therefore crucial for design and implementation, that we understand the utility of low quality video. One key finding from studies of low quality video systems, is that in certain circumstances adding visual information can *detract* from the interaction *processes*, if the video is implemented in a way that interferes with audio. There are two ways that audio can be affected in low bandwidth systems. First, certain commercial systems *delay* audio transmission, to allow time for video compression

and decompression over wide area networks, in order to present synchronised audio and video⁶. Second, some videoconferencing systems enforce *half-duplex*⁷ audio to preserve bandwidth for video.

There is evidence that reducing audio quality to incorporate video is highly disruptive of *turn-taking processes*. In a naturalistic study, O'Conaill *et al.* (1993) compared face-to-face interaction with a low quality wide-area ISDN videoconferencing system, operating over 128 kilobits/second bandwidth. Because of bandwidth limitations, and the synchronisation of audio and video, the ISDN system had one-way audio lags of between 410 and 780 milliseconds. In addition, audio was half-duplex and picture quality poor, because only 90 kilobits/second of bandwidth were available for the video stream. The study measured a number of characteristics of conversation *processes*. Interactive aspects of conversation that required precise timing such as giving feedback, switching speakers and asking clarifying questions were much reduced in the ISDN system compared with face-to-face interaction. Given the half duplex audio and lags, speakers were unable to time their conversational contributions, with the result that backchannels or interruptions arrived too late, or at inappropriate points in the conversation. As a consequence, people had to explicitly manage speaker switches and there was increased formality in handing over the conversational floor, using devices such as selecting the next speaker by name. The result of both decreased interactivity and increased formality was a "lecture-like" style of interaction, with conversational turns in the video conference being three times as long as face-to-face ones, making the system only suitable for certain types of conversational task, such as information exchange, which do not require quickfire exchanges.

Similar results showing the impact of audio lags on *conversational processes* are reported elsewhere. Cohen (1982) compared communication processes in face-to-face communication, with low quality videoconferencing, for a series of laboratory tasks. The system she investigated had a 705 milliseconds lag in both video and audio to simulate the performance of the A T & T Picturephone. Participants found it hard to switch speakers and also to ask clarifying questions in videoconferences. There were twice as many speaker switches in face-to-face communication compared with the videoconferencing system, and many more interruptions. Tang and Isaacs also evaluated low quality videophone and videoconferencing systems (Tang & Isaacs, 1993, Isaacs & Tang, 1994). They found that lagged audio is highly disruptive of turn-taking, producing many fewer, longer turns. The study also provides strong subjective support for the importance of low lag audio. Participants preferred to use a separate half-duplex speakerphone to reduce delays in audio, even though it meant that synchronisation between audio and video information was lost.

SUMMARY

These studies provide only weak evidence for the *non-verbal communication* hypothesis. First there is little evidence for *cognitive cueing*. Neither face-to-face communication nor high quality video/audio systems show *objective* benefits over audio only communication for problem solving tasks (Chapanis *et al.*, 1972; Chapanis, 1975). The results are more complex for *turn-taking processes*: even high quality video/audio is no different from speech only interaction, but there are differences between face-to-face and speech only. This suggests that visual information can potentially have an impact

⁶ Exact lags depend on the system and network, but typical figures for one way lags are 705 milliseconds for the Picturephone system (Cohen, 1982), between 410 and 780 milliseconds for an ISDN system operating between the US and the UK (O'Conaill *et al.*, 1993, Whittaker & O'Conaill, 1993), and 570 milliseconds for an ISDN system operating from coast to coast in the US (Tang & Isaacs, 1993, Isaacs & Tang, 1994).

⁷ Half-duplex audio only allows unidirectional transmission of audio. This prevents certain key conversational processes that depend on multiple participants at different ends of an audio link from being able to speak simultaneously, for example backchannels to provide feedback to the speaker, or interruptive clarifying questions.

on *conversation processes*, but that current video systems do not support this. Finally there is strong evidence for *affective* cues being transmitted by video (Reid, 1977; Short *et al.*, 1976; Williams, 1977). *Subjective* measures more consistently show effects of visual information: high quality video conversations are preferred to audio because they support turn-taking and affect (Isaacs & Tang, 1994, O'Conaill *et al.*, 1993, Tang & Isaacs, 1993, Sellen, 1992, in press), although the questionnaire data of Fish *et al.* (1992) are more ambiguous about the subjective benefits of video.

The data for low quality systems reveal a different problem with attempts to provide non-verbal information using video. Here, adding video to audio is often accompanied by reduced audio quality, and this impairs *communication processes*. Thus, while the addition of video does not always bring objective benefits when there is high bandwidth available, when bandwidths are restricted, it may indirectly *detract* from audio quality, despite the fact that audio quality has been documented to be a crucial determinant of interaction effectiveness (Chapanis *et al.*, 1972, Cohen, 1982, O'Conaill *et al.*, 1993; Tang & Isaacs, 1993, Whittaker & O'Conaill, 1993).

We now turn to the research and design issues arising from these results.

Research and design implications for non-verbal communication

ALLOCATING BANDWIDTH APPROPRIATELY IN LIMITED BANDWIDTH SYSTEMS

In the short term, most networking bandwidths will not support high quality audio and video. We should not therefore sacrifice audio quality in limited bandwidth systems, given the weak evidence for the benefits of adding video, combined with overwhelming demonstrations of the importance of high quality audio. In limited bandwidth systems, we therefore recommend that systems deliver high quality audio with minimal delays, even at the cost of video quality, and loss of synchronised audio and video, unless the task specifically requires access to *affective* information.

One set of outstanding design questions is therefore to determine the minimum acceptable quality audio in limited bandwidth systems. We can then allocate the remaining bandwidth to video. There is some preliminary research into critical properties of audio, but we need to extend this. There is strong evidence that half-duplex audio is highly disruptive of communication (CCITT, 1988, Krauss & Bricker, 1967, O'Conaill *et al.*, 1993), but results are inconsistent about the exact point at which lags begin to interfere with communication: Tang and Isaacs (1993) report that 320 to 420 millisecond lags are acceptable but other data suggest that delays of 200 milliseconds are disruptive (Reiz & Klemmer, 1963, Wolf, 1982)

Once we have determined acceptable audio quality, we also need to know how to deploy network bandwidth for video. Is this best used for high frame rate low resolution video, or low frame rate high resolution video? The only published data available indicate that 5 frames per second over a 500 kilobits/second dedicated network is viewed as being "tolerable" by users (Tang & Isaacs, 1993). It is also vital that these studies of video quality are conducted in the presence of the audio channel, given that other research has shown that the evaluation of video quality is influenced by the quality of the accompanying audio (Johansen, 1984). One complicating factor here is that requirements for audio and video will be task-dependent.

PRESENTING AND CONTROLLING IMAGE AND SOUND

Although visual information influences *turn-taking processes* as evidenced by differences between face-to-face and speech only communication, even high quality audio and video do not replicate face-to-face processes. We need to understand this from both theoretical and design perspectives. One argument is that the design of current systems fails to achieve key *presentational* aspects of sound and vision (Gaver, 1992, O'Conaill *et al.*, 1993, Sellen, 1992). Thus in most videoconferences, sound and image come from a single monitor and hence a single spatial location. In contrast, in multiparty face-to-face interaction, both sound and visual information come from multiple separate sources, so that cues such as the *direction* of a sound cue can be used to determine who is speaking (Sellen, 1992). We therefore need to address issues concerning the impact of different presentations of the video image and sound (Gaver, 1992). Little work has been done here, but questions include: what is the impact of image size, and what should it show (Mantei *et al.*, 1991)? With systems that present a video image from a fixed camera position, should it show only head and shoulders, or the whole upper body to depict gesture (Heath & Luff, 1991; Tang *et al.*, 1994)? What is the role of *proxemics*, ie. perceived physical distance, which has been shown to influence face-to-face communication? How important is mutual gaze, which is not supported in many systems (Buxton & Moran, 1990)? How important is it to support side conversations when there are several participants, with the associated requirement for multiple audio channels (Sellen, 1992, in press)? How should groups be visually and auditorily presented? Using picture in picture⁸ or spatially separated? Early work suggests subjective advantages for spatial separation of image and sound (Sellen, 1992, in press). Users also complain about the lack of privacy of audio in desktop video applications, compared with the shielded mouthpiece provided by the phone, and this problem needs to be addressed (Tang *et al.*, 1994). Should systems also provide people with information about how they themselves appear, and if so how? One current technique is to use confidence monitors, but some people complain that these are off-putting (O'Conaill *et al.*, 1993; Sellen, 1992). Should there be a separate monitor for the video image to preserve screen real estate? There are also issues about the extent to which people control the images they can see. With most current systems, the remote camera is controlled by the people at the remote location, whereas there are clear benefits for participants being able to control the remote camera (O'Conaill *et al.*, 1993), but straightforward controls and hardware have yet to be designed to enable this (Gaver, 1992; O'Conaill *et al.*, 1993, Sellen, 1992, in press).

INTEGRATING VIDEO WITH OTHER COMMUNICATIONS APPLICATIONS

The above studies show that the benefits of video for *non-verbal communication* are both subtle and subjective. One design implication is that the most effective use of video may not be in a standalone application, eg. videophone, but combining it with other communications applications, including shared workspaces (Tang & Rua, 1994). Such integration also provides support for *shared objects*, as part of a joint *physical context* (Suchman, 1992, Tang, 1991, Whittaker *et al.*, 1993), and research shows that the perceived utility of videoconferencing is enhanced by the addition of a shared workspace (Fish *et al.*, 1992, Tang & Isaacs, 1993). There are other benefits that can result from combining workspaces and non-verbal information. A key issue with shared workspaces lies in providing effective information about the remote participant's attentional focus and this can be provided by video (Minneman & Bly, 1991; Tang, 1991, Whittaker *et al.*, 1991; Whittaker *et al.*, 1993). Indeed, a number of workspace systems have been designed on the premise that a major benefit of video is to supply awareness information about the set of objects in the shared workspace that the remote user is focussed on, rather than supporting non-verbal communication. We return to this point when we discuss "*video-as-data*".

EXPLORING THE INTERACTION BETWEEN TECHNOLOGY AND TASK

⁸ A technique for displaying multiple tiled images on a computer monitor.

Work is also needed to specify more precisely the set of tasks for which non-verbal information is useful. The only set of laboratory tasks where there is clear impact are those that require access to affect or emotion (Short *et al.*, 1976). This suggests that potential benefits for video may lie in the home, rather than the office market, because information about emotions may play a greater role in personal, rather than business-oriented communications. Early home-based commercial products in this area, such as the videophone were relatively unsuccessful however, although the exact reasons for this are currently unclear (Noll, 1992). Image and sound quality may be an issue: Current videophones transmit between 5 and 10 frames of video per second, with poor resolution and unsynchronised audio and video. Again research needs to be done to determine what video and audio quality is acceptable for such devices to provide the necessary affective information. Another suggested application which emphasises affect is teleworking. Studies of teleworkers have shown that social isolation is a problem (Olson, 1989). Remote teleworkers might engage in interpersonal communication via videophone to substitute for the opportunistic social communications they are missing by not sharing a physical office. As yet we know of no studies that have systematically tested this hypothesis.

Another potential set of communication tasks for low bandwidth systems may be those that involve predominantly *one-way communications*. Given the problems with current wide-area videoconferencing systems, in supporting *turn taking processes* (O'Conaill *et al.*, 1993; Sellen, 1992, in press, Tang & Isaacs, 1993, Whittaker & O'Conaill, 1993), one possibility is to use video for remote communications tasks where interactions are more formally structured, and there are fewer speaker switches. Examples of such tasks might be remote teaching or lectures, in which there are relatively long periods when only the lecturer is speaking. Isaacs, Morris and Rodriguez (1994) have successfully built a system supporting remote attendance at a lecture.

Better understanding of the relationship between communication technologies and tasks is also needed to enable distributed groups to make strategic use of video technology at particular points in extended collaborations. Thus if low quality videoconferencing is indeed inappropriate for quickfire conversational exchanges, then videoconferencing should not be used in the planning and negotiation phases of projects where high interactivity and rapid speaker switching is required (O'Conaill *et al.*, 1993). In contrast, for project updates and information exchange, wide area videoconferencing may be an appropriate technology.

Video for connection and opportunistic communication

We now turn to other uses of video for remote communication. The non-verbal communication hypothesis makes the assumption that connection between participants has already been established. Such work ignores a vital aspect of conversational *process* co-ordination, namely how participants *initiate* conversations. Recent research suggests a different role for visual information, where the unplanned nature of the majority of communications makes co-ordinating with others a major problem (Rice & Shook, 1990, Whittaker *et al.*, 1994a). In face-to-face settings people rely on visual information to determine the availability of others (Festinger, Schacter & Back, 1950, Kraut *et al.*, 1993, Mintzberg, 1973, Whittaker *et al.*, 1994a). Thus, instead of enhancing a *pre-established* audio connection, video can be used to establish remote opportunistic communications, by providing information about other participants' *availability* for communication.

Three separate classes of video application have been built to provide visual information to *facilitate connection* for unplanned interactions: (a) *glance* which enables a user to briefly "look into" the office of a co-worker to assess their communication availability; (b) *open links* in which persistent video/audio channels are maintained between two separate

physical locations, where these can either be links between private offices or public areas; (c) *awareness* applications in which video images of coworkers' offices are periodically sampled, so that "snapshots" of their office can show their recent movements and availability. The difference between awareness and glance is that awareness information is asynchronous: it may be a single frame updated periodically.

Table 2 summarises systems that have been built to support connection. This work is novel, so that systematic evaluations have only been conducted for "glance" and for some types of "open link".

Table 2 about here

There are methodological problems in drawing conclusions about the *video for connection* hypothesis from the currently available data. Again there are problems associated with networking bandwidth. In wide area connection applications, video quality is poor. There may therefore be less motivation for using video for achieving wide area connection, when participants are aware that the ensuing conversation will be over low quality video, and the studies reviewed in the prior section reveal that low quality conferencing is disruptive of conversation. In local area applications, there may also be reasons why *video for connection* may have reduced utility. Visual connection information about coworkers may already be available: as people move around their workplace, they pick up this type of information, without recourse to a video system. In both wide area and local area settings, these confounding factors may lead to reduced use of *video for connection*. Nevertheless, when people *do* choose to use the technology for assessing availability, we can still ask how successful the technology is in achieving connection, and we now turn to this data.

GLANCE

For a local area system, Fish *et al* (1992, 1993) tested the use of different types of glance, and their differential success in promoting opportunistic interactions. The results showed that a brief glance at a user selected recipient was the most frequently chosen type of glance: 81% of user initiated interactions were of this type, with 54% of these leading to an extended conversation. All other modes of glances were much less frequent and had much reduced likelihood of resulting in conversations. One type of glance was intended to simulate chance meetings such as "bumping into" another person in a hallway. In face-to-face settings neither participant normally intends such encounters, but they can promote extended work-related conversations. These types of chance encounters were implemented as a system-initiated connection between two arbitrary participants. These system-initiated connections showed very high failure rates, with 97% being terminated immediately without conversation. Overall, the glance options that callers chose indicate that they want direct control over *who* they connect to, and *when* they connect, rather than have the system do this. Furthermore, people wanted to use the "glance" as a preparation for communication, and not merely to know "who is around". Glances that allowed "looking into" another office without the option of communicating, were an infrequent user choice, accounting for only 12% of user selected glances.

The relationship between glances and opportunistic communication was also explored by Tang *et al.* (1994) for a system operating across multiple sites in a local area. Participants could first "look into" the office of a remote coworker, with the option of converting this into an extended conversation. Altogether, only 25% of glances were converted into conversations. This is no better than connection rates using only the phone. Why was successful connection so infrequent? A significant proportion of failures (38%), occurred when the recipient was out of their office, but the reasons for the

remainder are unclear: only 4% were when the recipient explicitly signalled that they were unavailable for communication. Many of the other failed connection attempts may occur when the recipient is in their office but busy with another activity, or another person. Tang *et al.* do not report this data, however.

CONTINUOUSLY OPEN LINKS

Video and audio can also be used to support "continuously open links" between the offices of remote collaborators (Adler & Henderson, 1994, Fish *et al.*, 1992, Heath & Luff, 1991, Mantei *et al.*, 1991). This is intended to approximate to sharing the same physical office, so that opportunistic communications can be started with minimal effort between connected participants, and visual and auditory information about communication availability is persistently available. One aspect of this is the ability to "waylay" a potential recipient who is out of their office, by monitoring the open link, and seeing when they return to it, ensuring that a vital communication takes place (Bly *et al.*, 1993, Fish *et al.*, 1992, Mantei *et al.*, 1991). There are also claims that working with a video link may be less intrusive than sharing a real office, but still offer many of the same benefits, in terms of access to other participants (Heath & Luff, 1991).

The available evaluation data suggest, however, that open links may be the exception rather than the rule. Fish *et al.* (1992) report that only 5% of connections lasted more than 30 minutes, and Tang and Rua (1994) report that only five interactions (out of a possible 233) lasted more than 30 minutes. In both cases, these data may *overestimate* the frequency of open links because they include extended continuous interactions as well as intermittent conversations over open links. Thus, both sets of usage data suggest brief interactions, rather than open links are the main uses of the system.

Open links can also be constructed between public areas, such as the systems built by Bellcore and Xerox PARC to link geographically separated sites (Abel, 1990, Bly *et al.*, 1993, Fish *et al.*, 1990, Root, 1988). Cameras were installed in common areas, transmitting images of people at remote sites, so that people can see, for example, who happens to be in the coffee area of a remote site. This was intended to promote opportunistic conversations of the type that can occur when people meet in public areas of the same site. The Bellcore system provided very high quality audio and video. The Xerox systems used much lower bandwidth connections (initially these were 56 kilobits/second, although exact system specifications changed in the course of the project). Both systems report the frequent use of open links for social greetings or "drop-ins" between remote sites, with Xerox reporting that 70% of open link usage was of this type (Abel, 1990, Bly *et al.*, 1993). Clearly, these brief interactions would have been unlikely to occur in the absence of the system. The use of the open link was mainly limited to these brief social exchanges, however, and the link was seen by the users as being ineffective in supporting work (Fish *et al.*, 1993). The Bellcore study also examined how often extended verbal communications resulted from sighting someone over the videolink. They compared this with the likelihood of interaction following face-to-face sightings, and found that sightings over a videolink were less likely to convert to extended conversation than face-to-face sightings (Fish *et al.*, 1990).

Despite these negative conclusions it may however, prove to be the case, that open links can be successful for particular workgroup settings and job types (Adler & Henderson, 1994). In addition, these results may have occurred because of the weakness of particular implementations. In several of the desktop systems, there was no facility for interrupting or overriding an existing open link. People may therefore be unwilling to maintain an open link knowing that this makes them unavailable for potentially important calls from other users. In support of this, other work indicates that 12% of

informal face-to-face workplace communications are terminated by the interruption of a third party (Whittaker *et al.*, 1994a).

Taken together these preliminary results on glance and open links indicate a lack of evidence for the utility of *video for connection*: (a) failure rates with glancing are as high as with phone alone; (b) open links are an infrequently chosen user option; (c) open links are less likely to promote conversation than face-to-face sightings; and (d) open links between common areas are not adequate to support work. These failures may, however, be due to confounding factors in the evaluations, such as the fact that some tests were conducted in local area settings where people may already have access to availability information, and the ability to engage in face-to-face interaction. In addition, connection implementations may have been implemented in a way that precludes interruptions in the case of open links between offices, or that fails to replicate the manner in which people initiate face-to-face conversations. We therefore need more evaluations in wide area settings with improved implementations. We now turn to other design and research issues for *video for connection* systems, which address some of the possible reasons behind the lack of evidence for this hypothesis.

Research and design issues in video for connection

PRIVACY: MINIMISING INTRUSIVENESS

A major problem with connection applications is that they are perceived by call recipients as being intrusive, and this is a potential barrier to their acceptance. A related issue here is that users fear they could be used for monitoring personal activities and hence compromise their *privacy* (Fish *et al.*, 1992; Gaver *et al.*, 1992; Mantei *et al.*, 1991). In what ways might this problem be addressed? Workplace studies of office workers who are collocated show the ways that workers determine the communication availability of others, negotiate privacy, and initiate conversation is subtle and complex. It also depends in part on the pre-existing relationship between the participants (Kraut *et al.*, 1992, 1993, Heath & Luff, 1991; Whittaker *et al.*, 1994a). These results suggest that methods of initiating conversations using video should be both flexible and task specific.

A number of systems include features to address the privacy problem, and control for intrusive interruptions. These allow users to configure "access" privileges, on a caller by caller basis to allow for filtering and blocking of both glances and calls (Gaver *et al.*, 1992; Mantei *et al.*, 1991). An alternative approach is to have recipients specify their general level of interruptibility to *any call*. Callers can then decide whether to interrupt given this availability information. Thus, if a person has a "do not disturb" setting, a caller has to decide whether their call is sufficiently urgent to merit interrupting the recipient (Harrison, Mantei, Beirne & Narine, 1994; Tang & Rua, 1994). Other systems have implemented different styles of initiation, either using audio cues to alert recipients to the fact that they are being glanced at (Gaver *et al.*, 1992), or by "fading in" images of the remote participant during a glance (Tang & Rua, 1994, Tang *et al.*, 1994). There are also questions about call uptake, can either party opt to continue with full audio and video, or is this the recipient's prerogative? Again we need evaluations of these different designs, to determine their impact on privacy and initiation.

Finally, there are major issues about what the remote participant should be able to see and hear. Most systems have implemented "video only" glances, but research on naturally occurring conversations has shown that access to *audio* information is more important in determining when and whether a caller interrupts (Whittaker *et al.*, 1994a). Clearly there are issues of privacy associated with allowing "electronic eavesdrops", and these may make providing this form of

information unacceptable to users (Nardi, Schwarz, Kuchinsky, Leichner, Whittaker & Sciabassi, 1993). There are also questions associated with what visual information to show in a glance. Should the camera be set up to present "head and shoulders" shots of the person if they are at their machine, or should the glancer be able to "look around" the remote office? While a controllable camera may be more useful to the caller, it is also harder to implement and potentially more intrusive.

MAKING CONNECTIONS LIGHTWEIGHT

Studies of workplace interaction show that a crucial feature of opportunistic communications is their *brevity*, and many current systems do not support fast enough connections. Whittaker *et al* (1994a) found that opportunistic face-to-face interactions lasted on average about 1.89 minutes, with 50% lasting less than 38 seconds. Fish *et al* (1992) showed that system mediated conversations lasted on average 4 minutes, and Tang *et al.*, (1994) found that system mediated conversations lasted around 4 minutes 10 seconds. The design implication for this is that connections must be *fast*. If a large proportion of interactions last less than 38 seconds (Whittaker *et al.*, 1994a), then a connection time of 11 seconds is too long, and this may severely compromise the use of the system for impromptu conversations (Tang *et al.*, 1994). This is especially true of glances: if the requirement is to assess the communication availability of a co-worker, then this should not take more than a few seconds, and user studies should be conducted to find out precisely how long is acceptable⁹. One solution is to maintain open links, but there is a clear limitation on the number of these that can be open and managed at a given time, especially in wide area applications. For this reason, "awareness servers" may be preferred over continuously open video links.

ADDRESSING FAILED ATTEMPTS TO COMMUNICATE

Since the aim of these systems is to promote successful opportunistic interactions, there should be methods to facilitate meetings in the event of failure, eg. when a glance reveals the lack of availability of the recipient. This suggests that "glance" features should be accompanied by methods to "leave a message" with information to call the initiating party back. Thus Tang and Rua (1994) implemented a "stickup" note, which was a means to leave a message, eg. "*call me back*", in the event of a failed attempt to initiate communication. Another solution to communication failure may lie in integration with other technologies such as pagers or mobile devices which allow access to the recipient, when they are away from their desk.

PARAMETERISING AWARENESS

More research is needed into the value and character of awareness (Dourish & Bly, 1993). How frequently should awareness updates be made, how should they be presented, and how should recipients be informed when awareness data is collected? Update frequency will have a crucial bearing on the bandwidths needed to support this application. If data rates are sufficiently low, then the Internet may be used for awareness data (Dourish & Bly, 1993). We also need to understand differences between awareness and glancing applications. Awareness allows the caller to track the availability of multiple recipients, but given the intermittent refresh rate, the information supplied is not current. Glancing gives up-to-date information, but it may be more intrusive for the recipient.

PUBLIC VERSUS DESKTOP CONNECTION APPLICATIONS

⁹The requirement for fast connection also applies to the opening of audio and video links following a successful glance, and also to application sharing and shared workspaces. Users will not tolerate long delays in any of these processes, given that overall interaction time is so brief (Tang *et al.*, 1994).

There are also issues about *where* opportunistic communications applications should be located. Most face-to-face opportunistic communications take place in the offices of individual users rather than in public areas: 65% of workplace communications occur in people's offices, compared with 15% in public areas and 17% while on the move (Whittaker *et al.*, 1994a). This argues for desktop, rather than public area applications. Location may also have an impact on conversation initiation and hence system design with regard to intrusiveness: Backhouse and Drew (1992) showed that people are "more interruptible" in public areas than in their own private offices.

"Video-as-data"

Using video as a means to achieve connection does not exhaust novel applications of video however. An alternative hypothesis is that a major benefit of video lies in its ability to depict complex information about dynamic 3D shared work objects, rather than images of the participants themselves. Thus, the video image can be used to transmit real-time information about the work objects themselves, and this can then be used to co-ordinate conversational content among distributed teams, by creating a *shared physical context* (Clark & Brennan, 1991, Clark & Marshall, 1981, Whittaker *et al.*, 1991, Whittaker *et al.*, 1993). The example discussed here is remote surgery, but other tasks such as concurrent engineering, or training also have similar requirements (Egido, 1988, Nardi *et al.*, 1993).

Distributed images of complex, dynamically changing, physical objects are used in brain microsurgery conducted by dispersed surgical teams (Nardi *et al.*, 1993). In surgery, action and communication are focussed around performing actions that result in complex changes to a physical object, namely the patient's brain or spine. We studied the effects of distributing the magnified image of the surgeon's actions on the patient to other members of the operating team. Members of the team were either within the operating theatre, or, in the case of remote consultants, elsewhere in the hospital, up to a mile away. In most cases, the consultant was involved in several concurrent operations. The image was viewable through a microscope (for the surgeon and assisting nurse), on a large shared monitor (for everyone in the operating theatre), and over analogue networks (for remote consultants).

We found four different types of communicative use of the image.

First, the dynamic image of the surgeon's actions allowed detailed co-ordination of interleaved physical action between the assisting nurse and the surgeon in the operating theatre. By monitoring the surgeon's actions, via a shared video image viewed through the microscope, the nurse could anticipate the surgeon's requirements and provide the correct surgical instrument, often without it being directly requested. A second communicative function of the video image was that it served to disambiguate other types of surgical data that were supplied to remote consultants, such as neurophysiological monitoring data. The interpretation of these neurophysiological data depends critically on precise information about the physical actions that the surgeon is currently executing, such as the exact placement of a surgical clamp or the angle and direction of entry of a surgical instrument. Without the video image depicting these actions, the remote consultant had to rely on verbal reports from those who were present in the operating theatre, and the inadequacy of the descriptions meant that the consultant often had to resort to physically visiting the operating theatre to observe the actions directly. Thirdly the video image served as a physical embodiment of progress through the operation. Members of the team who were involved in multiple operations at different locations and also those within the Operating Theatre could *see* the current stage of each operation by inspecting the physical image, and observing what stage of the procedure the surgeon was at. The remote consultants could thus co-ordinate their visits to each operating theatre accordingly, so as to arrive at times

when their physical presence was critical¹⁰. Finally, the image was used for learning and education. The application was installed in a teaching hospital that undertook innovative surgical procedures. Academic visitors and trainees would often come to the operating theatre to observe the novel procedures on the large monitor in the operating theatre, as they occurred. Some surgeons also recorded these procedures, to use them as aids in teaching classes.

It is important to contrast *video-as-data* with alternative uses of video to depict objects, such as television or films. *Video-as-data* focusses on *real-time* depictions of shared work objects. The video is used to bring complex objects at one physical location into a virtual shared workspace to co-ordinate distributed teams. Participants can then discuss and modify these objects together. This contrasts with mass media transmission of images, in television or films, where video technology is used to play back previously recorded images, where there is little possibility of interaction with others viewing those images, and no possibility of modifying those objects together.

Similar arguments for this use of "*video-as-data*" are made by Gaver, Sellen, Heath, and Luff (1993). They looked at the use of images of 3D objects in design tasks. Users could choose between a number of different images, including between an image of the other participant, and various views of the object under study. Participants rarely chose facial views of their co-participant (11% of the time), and "mutual gaze" (where both participants were simultaneously viewing each other) occurred only 2% of the time. Instead, people were much more likely to choose an image of the object, spending 49% of their time with the object views. This shows that for this class of design application, information about gaze and gesture of the other conversational participant seems to be less important than information about the *shared physical context*.

An extensive research programme has been executed by Ishii, who has built a series of prototypes that use video to combine a semi-reflective writing surfaces with images of the participant's upper bodies. This enables the fusing of an image of another participant onto the work surface itself, making it possible to see both participant and object simultaneously and hence accurately track visual attention, while writing or manipulating the object (Ishii & Kobayashi, 1992). Again, a major focus of the work is that crucial collaborative information is embodied in the work object, although systematic evaluation of the benefits of adding this attentional information has not yet been conducted. Other "*video-as-data*" systems include "object-oriented" video, in which users can manipulate parts of the video as objects, so they can reference, overlay, highlight, and annotate them, as well as use them for control and browsing (Tani, Yamaashi, Tanikoshi, Futakawa & Tanifuji (1992). Milgram, Dracsic and Grodski (1990) also developed a system that combines stereoscopic video and computer graphics, so that users can point to, measure, and annotate objects in the video. Again, however, we need more data on the utility of these systems.

Research and design issues concerning "*video-as-data*"

One key problem is to identify the set of applications in which dynamic 3D visual information and a set of shared objects is important. We have seen that for microsurgery and 3D design, "*video-as-data*" can be useful, but where else might it be crucial to share a dynamic image of the task's central object? Other possibilities might include concurrent engineering

¹⁰This function is similar to using *video for connection*, in that video information is used to co-ordinate a communication episode, between people at remote locations.

where the dynamic image is crucial to the workgroup's actions, but we need more studies of applications involving this type of data to determine this.

Another issue is video quality. In the microsurgery application, it was vital to have high quality video images. The microsurgical operating team could share high quality video images because they all worked at the same hospital site, and hence could be connected by analogue cable links, but there are considerable problems in achieving the necessary quality over wide-area networks. We still do not know the image requirements for other tasks such as concurrent engineering, and how easily the requisite image quality can be achieved over such networks.

Privacy is also a potential problem in such applications, especially when the video is recorded. People feel that information derived from remote monitoring or recording could be used to evaluate their job performance and that the presence of video cameras may lead them to change the way that they conduct their work and hence make them less effective (Nardi *et al.*, 1993). When monitoring is conducted intermittently, participants also complain about lack of information about when it is being conducted. They also express concerns about who can observe or obtain access to the remote video images. This is especially true in remote consultancy applications, when all the workteam are not co-present, and remote participants are "invisible" to the team in the operating theatre. It is thus unclear to those in the operating theatre when the consultants are observing the operation. There is also the problem of having one's actions taken out of context by a remote observer who may only observe a subset of all one's workplace activity (Nardi *et al.*, 1993). All these issues must be addressed if the system is to be accepted. This means that care has to be taken to ensure that adequate security is provided to restrict access to relevant co-workers, and that feedback is provided to people to give them information about when they are being recorded or observed.

The main focus of this paper is on real-time rather than stored video, but there are retrieval issues if "*video-as-data*" is used for playback, and tools need to be developed to analyse, index and access large amounts of video data. For example, surgeons and neurophysiologists in the surgery application requested tools that would enable them to identify critical video events and to examine causal relations between those events and other types of events such as changes in heart rate or breathing. Tools are being developed to allow analysis, indexing and replay of complex multimedia data that are generated in these circumstances¹¹ (Davis 1994, Harrison 1991; Weber & Poon, 1994; Whittaker, Hyland & Wiley, 1994).

Summary

We have reviewed evidence for three different hypotheses about the role of video in interpersonal communications, suggested new directions that future applications might explore, and identified outstanding research and design issues. With the exception of tasks that require access to affective information, we found that evidence for the *non-verbal communication hypothesis* is not strong, with few task outcome and process differences being found between audio and video-enhanced communications. Furthermore, despite the absence of compelling evidence for the *non-verbal*

¹¹ Although much of this work is recent, two main approaches have been taken to this problem. The first makes the assumption that indexing will take place *after* the material has been recorded. For example, systems have been built that allow video sequences to be categorised according to the types of events, participants and actions they depict (Davis, 1994). One problem with this approach is that it forces users to learn a complex categorisation scheme, with consequent issues concerning reliability of indexing when data is coded and accessed by multiple individuals. The second approach has people generate indices *while the material is being recorded*. These can take the form of handwritten notes or other user actions which can be used as user-centred indices for retrieval (Harrison, 1991, Weber & Poon, 1994, Whittaker *et al.*, 1994b). Some of these systems also allow hybrid indices constructed at the during, as well as after, recording has taken place. As these tools are still under development, we lack evidence of their effectiveness.

communication hypothesis, certain current implementations may have compromised overall system utility by focussing on video at the expense of providing full duplex, low lag audio. Failing to provide this type of audio information disrupts conversational processes that require precise timing and bidirectionality.

Nevertheless methodological and theoretical questions remain about the *non-verbal communication* hypothesis. We therefore need to sharpen the way the hypothesis is framed, so that more specific predictions can be tested and better systems designed. At the outset we identified three possible ways in which visual information might enhance audio communications: by providing: (1) cognitive cues to enhance shared understanding; or (2) social cues for providing information about interpersonal communication and affect; (3) conversation process cues to facilitate speaker change. There is some evidence for the impact of video in tasks depending on affect or emotion, supporting the social cueing hypothesis, and hence certain types of conversational content co-ordination. Neither process nor cognitive cueing accounts are well supported, however. For cognitive cueing, Chapanis *et al.*, (1975) show that even face-to-face communication is no better than speech only, but there are differences in *conversational processes* between face-to-face communication and high quality video (O'Conaill *et al.*, 1993, Sellen, 1992, in press). We therefore need to understand why even high quality audio and video do not replicate face-to-face processes. One possibility is that current systems do not accurately simulate the *presentational* aspects of face-to-face interaction: spatial audio and video may therefore be needed to replicate conversational *processes* (O'Conaill *et al.*, 1993, Sellen, 1992). However, there are other possible explanations and this claim needs to be tested.

Another possible explanation of the results on the non-verbal communication hypothesis is that certain types of information can be transmitted by different conversational media, whereas others cannot. Thus in face-to-face communication, *cognitive* and *process* information is partially transmitted by the visual channel, eg. by head nods, eye gaze and head turning.. However data on the efficacy of speech only communication and studies of conversation, indicate that cognitive and process information can also be communicated effectively by other (non-visual) cues in speech only communication (Walker, 1992, Walker & Whittaker, 1990). This suggests that cognitive and process information can partially *substitute* across different media. In contrast, the removal of the visual channel changes the outcome of tasks that require access to affect. Part of the reason might be that *affective cues* are often not generated intentionally, so that although speech can signal affect, speakers omit the full range of *affective* cues, when using audio only communication. Future theoretical work should address this issue of the *substitutability* of different media and information types, and the role of intentional cueing.

Another unresolved problem is to explain why subjective and objective measures are not in accord for the *non-verbal communication* hypothesis: while outcome and process show few differences between audio and video conversations, people *reliably prefer* video mediated communications (Fish *et al.*, 1992, Tang and Isaacs, 1993, Sellen, 1992). One possibility is that subjective preferences are an aspect of social cueing, and hence provide evidence for this hypothesis, but the social cueing account must be clarified for this argument to be sustained. One promising line of work is the investigation of person perception in video-mediated settings, which may enable specific predictions about "social presence" to be tested, independently of competing explanations about cognitive or conversational process cues.

The second hypothesis, that video is useful for the *process* of initiating unplanned conversations, has yet to be systematically tested, so that the putative function of promoting *opportunistic connection* is undemonstrated. While

workplace studies show the importance of opportunistic communications, it is currently unclear how well video can support their initiation and hence support this aspect of conversational process. One reason for the lack of clear evidence may be the methodological limitations of current studies. Evaluation work needs to focus more on situations in which there is a critical mass of users who are geographically remote: early evaluations have suffered from only investigating small user populations who often share the same physical space. Other design factors such as long delays in initiating communication, style of initiation, and most importantly, privacy issues, also have to be addressed before we can make clear statements about the effectiveness of video for *connection*. Work should also be done to investigate whether alternative technologies, e.g. active badges (Pier, 1991), could also supply availability information and hence substitute for *visual* information. There is also the question of the extent to which other asynchronous technologies can partially substitute for opportunistic meetings. Can a brief email or voicemail message replace a short synchronous discussion and hence reduce the need for remote opportunistic meetings?

Finally, *video-as-data* is a promising area, where more applications should be built and evaluations conducted. Early work on video has neglected the importance of shared objects, as part of a *shared context*. Given the lack of clear support for *non-verbal communication*, *video-as-data* may be a more successful use of video, if we can identify tasks that are focussed on complex dynamic 3D objects. However, as with *opportunistic connection*, there are also outstanding social issues about privacy and access that have yet to be addressed.

Overall, this paper suggests that the successful application of video technology for interpersonal communications still requires extensive research. Rather than the single function of broadening communication bandwidth implied by the *non-verbal communication hypothesis*, we need to extend the set of hypotheses we entertain about video, to think about video for initiating opportunistic communication and representing shared objects. The work reviewed here also suggests that the benefits of video are task- and situation-specific. Future research must explain when and why this technology brings benefits to interpersonal communication.

References

- ABEL, M. (1990). Experiences in an exploratory distributed organization. In J. GALEGHER, R. KRAUT & C. EGIDO Eds. *Intellectual Teamwork*. Hillsdale, N.J.: Lawrence Erlbaum Press.
- ADLER, A., & HENDERSON, A. (1984). A room of our own: experiences from a direct office share. In *Proceedings of CHI'94 Human Factors in Computing Systems*, 138-144, New York: ACM Press.
- ALLEN, J & PERRAULT, R. (1986). Analyzing intentions in utterances. In B. GROSZ, K. SPARCK-JONES & B. WEBBER Eds. *Readings In Natural Language Processing*. Los Altos, California: Morgan Kaufmann.
- ARGYLE, M., LALLJEE, M., & COOK, M. (1968). The effects of visibility on interaction in the dyad. *Human Relations*, **28**, 289-304.
- ARGYLE, M., LEFEBVRE, L., & COOK, M. (1974). The meaning of five patterns of gaze. *European Journal of Social Psychology*, **4**, 125-136.
- BACKHOUSE, A., & DREW, P. (1992). The design implications of social interaction in a workplace setting. *Environment and Planning*, **19**, 573-584.
- BLY, S., HARRISON, S., & IRWIN, S. (1993). Media spaces: Bringing people together in a video, audio and computing environment, *Communications of the ACM*, **36**, 28-45.

- BUXTON, W. & MORAN, T. (1990). EuroParc's integrated interactive intermedia feature (iiif): Early experiences. In *Proceedings of the IFIP WG8.4 Conference on multiuser interfaces and applications*, 11-34.
- CLARK H. & BRENNAN, S. (1991). Grounding in communication. In L.B. RESNICK, J. LEVINE & S. TEASLEY, Eds. *Perspectives on socially shared cognition*. Washington DC., APA Press.
- CLARK H. & MARSHALL, C. (1981). Definite reference and mutual knowledge. In A. JOSHI, B. WEBBER & I. SAG Eds. *Elements of discourse understanding*. Cambridge, Cambridge University Press.
- CLARK, H., & SCHAEFER, E. (1989). Contributing to discourse. *Cognitive Science*, **13**, 259-292.
- CLARK, H. & WILKES-GIBBS, D. (1986). Referring as a collaborative process. *Cognition*, **22**, 1-39.
- CCITT. (1988). Contribution Com XII. *CCITT Guidelines*, 199.
- CHAPANIS, A. (1975). Interactive human communication. *Scientific American*, **232**, 34-42.
- CHAPANIS, A., OCHSMAN, R., PARRISH, R., & WEEKS, G. (1972). Studies in interactive communication: I The effects of four communication modes on the behavior of teams during cooperative problem solving. *Human Factors*, **14**, 487-509.
- COHEN, K. (1982). Speaker interaction: video teleconferences versus face-to-face meetings. In *Proceedings of teleconferencing and electronic communications*, 189-199. Madison: University of Wisconsin Press.
- COOPER, R. (1974). The control of eye fixation by the meaning of spoken language. *Cognitive Psychology*, **6**, 84-107.
- DAVIS, M. (1994). Knowledge representation for video. In *Proceedings of AAAI94*. New York, MIT Press.
- DOURISH, P, & BLY, S. (1993). Portholes: Supporting awareness in a distributed work group. In *Proceedings of CHI'93 Human Factors in Computing Systems*, 541-547, New York: ACM Press.
- DUNCAN, S. (1972). Some signals and rules for taking speaker turns in conversation. *Journal of Personal and Social Psychology*, **23**, 283-292.
- EGIDO, C. (1988). Video conferencing as a technology to support group work: a review of its failures. In *Proceedings of Conference on Computer Supported Co-operative Work*, 13-24, New York: ACM Press.
- EGIDO, C. (1990). Teleconferencing as a technology to support co-operative work: a review of its failures. In J. GALEGHER, R. KRAUT, & C. EGIDO, Eds. *Intellectual Teamwork*. Hillsdale, N.J.: Lawrence Erlbaum Press.
- EKMAN, P. & FRIESEN, W. (1975). *Unmasking the face*. New Jersey: Prentice Hall.
- FESTINGER, L., SCHACTER, S., & BACK, K. (1950). Social pressures in informal groups. A study of human factors in housing. Palo Alto, California: Stanford University Press.
- FINHOLT, T., SPROULL, L., & KIESLER, S. (1990). Communication and performance in ad-hoc task groups. In J. GALEGHER, R. KRAUT & C. EGIDO, Eds. *Intellectual Teamwork*. Hillsdale, N.J.: Lawrence Erlbaum Press.
- FISH, R., KRAUT, R., & CHALFONTE, B. (1990). The videowindow system in informal communications. In *Proceedings of Conference on Computer Supported Co-operative Work*, 1-12, New York: ACM Press.
- FISH, R., KRAUT, R., ROOT, R. & RICE, R. (1992). Evaluating video as a technology for informal communication, In *Proceedings of CHI'92 Human Factors in Computing Systems*, 37-48, New York: ACM Press.
- FISH, R., KRAUT, R., ROOT, R. & RICE, R. (1993). Video as a technology for informal communication. *Communications of the ACM*. **36**, 48-61.
- GAVER, W. (1992). The affordances of media spaces for collaboration. In *Proceedings of the Conference on Computer Supported Co-operative Work*, 17-24, New York: ACM Press.
- GAVER, W., MORAN, T., MACLEAN, A., LOVSTRAND, L., DOURISH, P., CARTER, K., & BUXTON, W. (1992). Realizing a video environment: EuroParc's RAVE system. In *Proceedings of CHI'92 Human Factors in Computing Systems*, 27-35, New York: ACM Press.

- GAVER, W., SELLEN, A, HEATH, C., & LUFF, P. (1993). One is not enough: multiple views in a media space. In *Proceedings of CHI'94 Human Factors in Computing Systems*, 335-341, New York: ACM Press.
- GOODWIN, C. (1981). *Conversational Organization: Interaction between speakers and hearers*. New York: Academic Press.
- GROSZ, B. & SIDNER, C. (1986). Attentions, intentions and the structure of discourse. *Computational Linguistics*, **12**, 175-204.
- HARRISON, B. (1991). Video annotation and multimedia artifacts: From theory to practice. In *Proceedings of the Human Factors Society 35th Annual Conference*, 319-323.
- HARRISON, B, MANTEI, M., BEIRNE, G, & NARINE, T. (1994). Communicating about communicating: cross disciplinary design of a media space interface. In *Proceedings of CHI'94 Human Factors in Computing Systems*, 124-130, New York: ACM Press.
- HEATH, C, & LUFF, P. (1991). Disembodied conduct: communication through video in a multimedia environment. In *Proceedings of CHI'91 Human Factors in Computing Systems*, 99-103, New York: ACM Press.
- ISAACS, E., MORRIS, J. & RODRIGUEZ, T. (1994). A forum for supporting interactive presentations to distributed audiences. In *Proceedings of Conference on Computer Supported Co-operative Work*, 405-416, New York: ACM Press.
- ISAACS, E., & TANG, J. (1993). What video can and can't do for collaboration: a case study. In *Proceedings of the ACM Multimedia 93 Conference*. Anaheim., CA.
- ISHII, H., & KOBAYASHI, M. (1992). Clearboard: a seamless medium for shared drawing and conversation with eye contact. In *Proceedings of CHI'92 Human Factors in Computing Systems*, 525-532, New York: ACM Press.
- JOHANSEN, R. (1984). *Teleconferencing and beyond*. New York: McGraw-Hill.
- KAHNEMAN, D. (1973). *Attention and effort*. New Jersey: Prentice Hall.
- KENDON, A. (1967). Some functions of gaze direction in social interaction. *Acta Psychologica*, **26**, 1-47.
- KLEINKE, C. (1986). Gaze and eye contact: a research review. *Psychological Bulletin*, **100**, 78-100.
- KRAUSS, R. & BRICKER, P. (1967). Effects of transmission delay and access delay on the efficiency of verbal communication. *Journal of the Acoustical Society of America*, **41**, 286-292.
- KRAUT, R., EGIDO, C & GALEGHER, J. (1990). Patterns of communication in scientific research collaboration. In J. GALEGHER, R. KRAUT & C. EGIDO Eds. *Intellectual Teamwork*. Hillsdale, N.J.: Lawrence Erlbaum Press.
- KRAUT, R., FISH, R., ROOT, B. & CHALFONTE, B. (1993). Informal communication in organizations. In R. Baecker Ed., *Groupware and Computer Supported Co-operative Work*. San Mateo, California: Morgan Kaufman.
- LUFF, P. HEATH, C. & GREATBATCH, D. (1992). Tasks-in-interaction. Paper and screen based documentation in collaborative activity. In *Proceedings of Conference on Computer Supported Co-operative Work*, 163-170, New York: ACM Press.
- MANTEI, M, BAECKER, R, SELLEN, A, BUXTON, W, MILLIGAN, T, & WELLMAN, B. (1991). Experiences in the use of a media space. In *Proceedings of CHI'91 Human Factors in Computing Systems*, 203-209, New York: ACM Press.
- MEHRABIAN, A. (1971). *Silent messages*. Belmont, N.J.: Wadsworth Press.
- MILGRAM, P., DRACSIC, D., & GRODSKI, J. (1991). A virtual stereoscopic pointer for a real three dimensional video world. In *Proceedings of Interact'90*, 695-700.
- MINNEMAN, S., & BLY, S. (1991). Managing a trois: A study of a multi-user drawing tool in distributed design work. In *Proceedings of CHI'91 Human Factors in Computing Systems*, 217-224, New York: ACM Press.
- MINTZBERG, H. (1973). *The nature of managerial work*. New York, Harper and Row.

- MOSIER, J., & TAMMARO, S. (1994). Videoteleconferencing among geographically dispersed work groups: a field investigation of usage patterns and user preferences. *Journal of Organizational Computing*, **4**, 343-366.
- NARDI, B, SCHWARZ, H, KUCHINSKY, A, LEICHTNER, R, WHITTAKER, S & SCLABASSI, R. (1993). Turning away from talking heads: an analysis of "video-as-data". In *Proceedings of CHI'93 Human Factors in Computing Systems*, 327-334, New York: ACM Press.
- NOLL, M. (1992). Anatomy of a failure: Picturephone revisited. *Telecommunications Policy*, (May/June), 307-316.
- O'CONNAILL, B., WHITTAKER, S., & WILBUR, S. (1993). Conversations over videoconferences: an evaluation of the spoken aspects of video mediated interaction. *Human Computer Interaction*, **8**, 389-428.
- OLSON, M. (1989). Work at home for computer professionals: current attitudes and future prospects. *ACM Transactions on Office Information Systems*, **7**, 317-338.
- PANKO, R. (1992). Managerial Communication patterns, *Journal of Organisational Computing*, **2**, 95-122.
- PIER, K. (1991). Active Badge Panel. In *Proceedings of Conference on Organizational Systems*. Atlanta, Georgia.
- REID, A. (1977). Comparing the telephone with face-to-face interaction. In I. POOL Ed., *THE SOCIAL IMPACT OF THE TELEPHONE*. Cambridge, MA: MIT Press.
- Reiz, R. & Klemmer, E. (1963). Subjective evaluation of delay and echo suppressors in telephone communication. *Bell System Technical Journal*, **4**, 2919-2942.
- ROOT, R. (1988). Design of a multi-media vehicle for social browsing. In *Proceedings of Conference on Computer Supported Co-operative Work*, 25-38, New York: ACM Press.
- RICE, R & SHOOK, D. (1990). Voice messaging, co-ordination and communication. In In J. GALEGHER, R. KRAUT & C. EGIDO Eds. *Intellectual Teamwork*. Hillsdale, N.J., Lawrence Erlbaum Press.
- RUTTER, D., & ROBINSON, R. (1981). An experimental analysis of teaching by telephone. In G. STEPHENSON & J. DAVIES Eds., *Progress in applied social psychology*, London: Wiley Press.
- SACKS, H., SCHEGLOFF, E. & JEFFERSON, G. (1974). A simplest systematics for the organization of turn-taking in conversation. *Language*, **50**, 696-735.
- SEARLE, J. (1990). Collective intentionality. In P. Cohen, J. Morgan & M. Pollack Eds. *Intentions in Communication*. Cambridge, MA.: MIT Press.
- SELLEN, A. (1992). Speech patterns in video-mediated communication. In *Proceedings of CHI'92 Human Factors in Computing Systems*, 49-59, New York: ACM Press.
- SELLEN, A. (in press). Remote conversations: the effects of mediating talk with technology. *Human Computer Interaction*.
- SHORT, J, WILLIAMS E, & CHRISTIE, B. (1976). *The social psychology of telecommunications*. London: Wiley Press.
- SPROULL, L. (1984). The nature of managerial attention. In L. SPROULL & J. LARKEY, Eds, *Advances in information processing in organisations*. Greenwich, Connecticut: JAI Press.
- SUCHMAN, L. (1992). Constituting shared workspaces. In D. MIDDLETON & Y. ENGESTROM EDS., *Cognition and Communication at work*. London: Sage Press.
- TANG, J. (1991). Findings from observational studies of collaborative work. *International Journal of Man-Machine Studies*, **34**, 143-160.
- TANG, J. & ISAACS, E. (1993). Why do users like video: studies of multimedia-supported collaboration. *Computer Supported Cooperative Work*, **1**, 163-196.
- TANG, J., ISAACS, E, & RUA, M. (1994). Supporting distributed groups with a Montage of lightweight interactions . In *Proceedings of Conference on Computer Supported Co-operative Work*, 23-34, New York: ACM Press.

- TANG, J. & RUA, M. (1994). Montage: Providing teleproximity for distributed groups. In *Proceedings of CHI'94 Human Factors in Computing Systems*, 37-43, New York: ACM Press.
- TANI, M., YAMAASHI, K., TANIKOSHI, K., FUTAKAWA, M., & TANIFUJI, S. (1992). Object-oriented video: Interaction with real-world objects through live video. In *Proceedings of CHI'92 Human Factors in Computing Systems*, 593-598, New York: ACM Press.
- WALKER, M. (1992). Redundancy in collaborative dialogue. In *Proceedings of the International Conference on Computational Linguistics*, 345-351.
- WALKER, M. (1993). *Information redundancy in dialogue*. Ph.D. thesis. University of Pennsylvania.
- WALKER, M. & WHITTAKER, S. (1990). Mixed initiative in dialogue. In *Proceedings of 28th Annual Meeting of the Conference on Computational Linguistics*, 70-78, Morristown N.J.: ACL Press.
- WEBER, K. & POON, A. (1994). Marquee: A tool for real-time video logging. In *Proceedings of CHI'94 Human Factors in Computing Systems*, 58-64, New York: ACM Press.
- WHITTAKER, S., BRENNAN, S., & CLARK, H.H. (1991). Co-ordinating activity: an analysis of computer supported cooperative work. In *Proceedings of CHI'91 Human Factors in Computing Systems*, 361-367, New York: ACM Press.
- WHITTAKER, S., FROHLICH., & DALY-JONES, O. (1994a). Informal workplace communication: what is it like and how might we support it? In *Proceedings of CHI'94 Human Factors in Computing Systems*, pp. 130-137, New York: ACM Press.
- WHITTAKER, S., GEELHOED, E., & ROBINSON, E. (1993). Shared workspaces: how do they work and when are they useful? *International Journal of Man-Machine Studies*, **39**, 813-842.
- WHITTAKER, S., HYLAND, P., & WILEY, M. (1994). Filochat: Handwritten notes provide access to recorded conversations. In *Proceedings of CHI'94 Human Factors in Computing Systems*, 271-277, New York: ACM Press.
- WHITTAKER, S. & O'CONNILL, B. (1993). Evaluating videoconferencing. In *Companion Proceedings of CHI'93 Human Factors in Computing Systems*, New York: ACM Press.
- WHITTAKER, S. & STENTON, P. (1988). Cues and control in expert client dialogues. In *Proceedings of the Conference for the Association for Computational Linguistics*, 123-130, Cambridge, MA.: MIT Press.
- WILLIAMS, E. (1977). Experimental comparisons of face-to-face and mediated communication. *Psychological Bulletin*, **16**, 963-976.
- WOLF, C. (1982). Videoconferencing: delay and transmission considerations. In L. PARKER & C. OLGREN (Eds.), *Teleconferencing and Electronic Communications*, Hillsdale, N.J.: Lawrence Erlbaum.

AUTHOR	SYSTEM QUALITY	METHOD	OUTCOME	PROCESS	SUBJECTIVE
Chapanis et al (1972, 1977)	HIGH: Cable for analogue video and audio	Lab studies, comparing video with ftf and audio	Neither ftf ¹² nor video is better than phone for cognitive tasks	No differences between media	
Reid, (1976), Short et al., (1976)	HIGH: Cable for analogue video and audio.	Lab studies, comparing video with ftf and audio. Questionnaire	Video differs from audio only for certain "affective" tasks.	Video and ftf conversations, more polite and personalised than audio.	Perception that ftf and video are better than audio for "social/affective" tasks.
O'Conaill et al., (1993)	HIGH: Broadcast quality video, low latency full duplex audio.	Naturalistic study comparing high quality video with ftf		Increased formality, of high quality video, less backchannel feedback compared with ftf	
Sellen (1992, in press)	HIGH: Cable network, five high quality audio and video systems	Lab study, comparing high quality video with ftf and audio		Decreased overlapping speech and interrupts, increased formality for video compared with ftf. No benefit of video over audio only	Ftf better than video for interactivity and control. Video better than audio for interruptions, naturalness, interactivity, feedback and attention.
Fish et al (1992, 1993)	HIGH: Cable for analogue audio and video	Field and questionnaire study of videophone compared with ftf and phone	Videophone used in preference to phone.	Duration of videophone calls was identical with phone.	Videophone perceived to be more like phone than ftf, more intrusive than phone, and ineffective for carrying out work.
Cohen (1982)	LOW: Simulation of Picturephone, lagged audio and video	Lab study, comparing low quality video with ftf		Fewer turns and less simultaneous speech with video	Ftf more enjoyable than video.
Tang & Isaacs (1993). Isaacs & Tang (1993).	LOW: Lagged, low frame rate video, lagged audio	Questionnaire and naturalistic study comparing low quality video with ftf. Shared tools were added.	More use of shared tools with videophone. Presence of video reduces use of email.	Reduced interactivity, ability to direct attention, compared with ftf.	Perceived reduction in face-to-face meetings and phone use when video is available. Audio and shared tools perceived to be more important than video.
O'Conaill et al (1993)	LOW: Lagged, half duplex audio, with 90kb/s video.	Naturalistic study comparing low quality video with ftf		Increased formality and reduced interactivity of low quality video compared with ftf.	

Table 1: System evaluations of the non-verbal communication hypothesis.

¹² FTF is an abbreviation for face-to-face communication

AUTHORS	SYSTEM TYPE	METHOD	OUTCOME	SUBJECTIVE
Fish et al (1992, 1993)	Local area videophone with variety of glance features: glance alone, glance to a named recipient, system initiated glances.	Field study with questionnaire	Glance to named recipient most frequent and successful option. Glances where system selects recipients are unsuccessful. Waylays and open links occasionally observed.	System perceived as intrusive by recipients.
Tang & Rua (1994), Tang et al (1994).	"Fade in" glance with local area videophone	Field study	Glances used frequently, but low connection rate. Little impact of glance on uses of other communication technologies. Infrequent use of open links.	More intrusive than phone, ftf ¹³ , but perceived as replacing some ftf meetings.
Abel (1990), Bly et al (1993)	Common area open link over wide area	Observational study	Increased social greetings promoted by open link	Sufficient to maintain social relations between remote sites.
Fish et al (1990)	Common area open link over wide area	Field study	Fewer casual "sightings" over video connections converted to conversations than in ftf	
Mantei et al. (1991), Adler & Henderson (1994).	Videophone and open links	Observational studies		
Gaver et al. (1992)	Glance, open link modes	Observational studies		
Dourish & Bly (1993)	Awareness server	Observational studies		
Heath & Luff, (1991)	Videophone and open links	Ethnomethodology study	Reduced impact of video can make initiation difficult	

Table 2: Evaluating video for connection.

¹³ FTF is an abbreviation for face-to-face communication.

