HUMAN-COMPUTER INTERACTION, 1993, Volume 8, pp. 389-428

Conversations Over Video Conferences: An Evaluation of the Spoken Aspects of Video-Mediated Communication

Brid O'Conaill and **Steve Whittaker**
Hewlett Packard Research Laboratories, UK

Sylvia Wilbur
Queen Mary and Westfield College

ABSTRACT

Recent trends toward telecommuting, mobile work, and wider distribution of the work force, combined with reduced technology costs, have made video communications more attractive as a means of supporting informal remote interaction. In the past, however, video communications have never gained widespread acceptance. Here we identify possible reasons for this by examining how the spoken characteristics of video-mediated communication differ from face-to-face interaction, for a series of real meetings. We evaluate two wide-area systems. One uses readily available Integrated Services Digital Network (ISDN) lines but suffers the limitations of transmission lags, a half-duplex line, and poor quality video. The other uses optical transmission

Authors' present addresses: Brid O'Conaill and Steve Whittaker, Hewlett Packard Laboratories, Filton Road, Stoke Gifford, Bristol BS12 6QZ, United Kingdom; Sylvia Wilbur, Computer Science Department, Queen Mary and Westfield College, University of London, Mile End Road, London E1 4NS, United Kingdom. email for O'Conaill: boc@hplb.hpl.hp.com. email for Whittaker: sjw@hplb.hpl.hp.com. email for Wilbur: sylvia@dcs.qmw.ac.uk.

CONTENTS

and video-switching technology with negligible delays, full duplex audio, and broadcast quality video.

To analyze the effects of video systems on conversation, we begin with a series of conversational characteristics that have been shown to be important in face-to-face interaction. We identify properties of the communication channel in face-to-face interaction that are necessary to support these characteristics, namely, that it has low transmission lags, it is two way, and it uses multiple modalities. We compare these channel properties with those of the two video-conferencing systems and predict how their different channel properties will affect spoken conversation. As expected, when compared with face-to-face interaction, communication using the ISDN system was found to have longer conversational turns; fewer interruptions, overlaps, and backchannels; and increased formality when switching speakers. Communication over the system with broadcast quality audio and video was more

similar to face-to-face meetings, although it did not replicate face-to-face interaction. Contrary to our expectations, formal techniques were still used to achieve speaker switching. We suggest that these may be necessary because of the absence of certain speaker-switching cues. The results imply that the advent of high-speed multimedia networking will improve but not remove all the problems of video conferencing as an interpersonal communications tool, and we describe possible solutions to the outstanding problems.

1. INTRODUCTION

In most working environments, people collaborate in groups to undertake collective tasks. Face-to-face communication plays an important role in the development and maintenance of these collaborations, and it is also critical for certain classes of workplace communication task such as project definition, initiation, and planning (Finholt, Sproull, & Kiesler, 1990; Galegher & Kraut, 1990; Kraut, Egido, & Galegher, 1990). Changes in work practice mean, however, that physical proximity and hence informal face-to-face interaction may not always be possible for future work groups. There are three main trends here: telecommuting (Harkness, 1986; Kraut, 1989); mobile work, for example, from customer sites or "on the road" (Sproull & Kiesler, 1991); and concurrent engineering, with designers, suppliers, and manufacturers increasingly coordinating over widely dispersed geographical locations (Johansen, 1984).

Given these trends, it is imperative that some means be found to support informal interaction remotely. This has led to an increased interest in technologies that attempt to replicate face-to-face interaction, such as video conferencing and the video phone. These technologies are premised on the hypothesis that the more closely they mimic face-to-face communication, the more effective the communication that will take place. Current pervasive technologies for remote synchronous communication such as the telephone are limited to the auditory medium. The hypothesis is that, by adding a visual channel to the phone, the added benefits of gaze, gesture, and the ability to monitor people's reactions will improve the quality of the communication. In addition, by facilitating frequent high-quality interaction between distant sites, these technologies will increase the number of potential co-workers and, hence, improve the quality of remote collaboration. There should also be less need to travel because people can replace face-to-face meetings with video conferences.

However, an examination of the history of video conferencing and video phone reveals a lack of success. Despite promising past market forecasts, video technology has not gained widespread acceptance (Egido, 1988; Noll,

1992). In part this is due to inadequate analysis of user needs, in particular with regard to travel substitution (Johansen, 1984; Johansen & Bullen, 1984; Panko, 1992).

Other work has raised questions about the value of the video channel. Laboratory studies assessed the impact of different channels on the efficiency of solving various tasks. Results indicated little value to adding a visual channel for task-based communication such as information transmission or collaborative problem solving (Chapanis, 1975; Reid, 1977; Williams, 1977). The time taken to solve problems, and the quality of solution, is almost equivalent whether or not a visual channel is available. Other studies have found some tasks for which video does influence outcome, but this is highly dependent on task type (Short, Williams, & Christie, 1976; Williams, 1977).

Another approach has investigated video as a technology in the workplace, and, from this perspective, studies of desktop video and persistent video links have been undertaken (Abel, 1990; Dourish & Bly, 1992; Mantei et al., 1991). Although early reports of this work were mainly favorable, the systems have often been used by the developers or researchers themselves, and reports have sometimes been anecdotal. Recent evaluation work has also failed to find strong benefits for this technology. Evaluation of desktop video indicates that interaction is more like phone conversations than face-to-face meetings, with conversations tending to be brief and task focused (Fish, Kraut, Root, & Rice, 1992). Attempts to use video for opportunistic meetings ("social browsing") have similarly been unsuccessful when compared with face-to-face interaction: Unplanned video contacts were less likely to lead to lengthy conversations than similar face-to-face contacts (Fish, Kraut, & Chalfonte, 1990).

There are two problems with the research into workplace video. Most of it examines systems that link local work environments, and the technology is local-area networks or cable TV, which support high-quality audio and video. However, most commercially available systems are wide area, where networking contraints do not allow high-quality video and audio. We know from other work that reduced-quality audio and video have strong impacts on communication properties (K. Cohen, 1982; Krauss & Bricker, 1967; Krauss & Fussell, 1990; Rutter & Stephenson, 1977; Sellen, 1992). Also there may be less incentive to converse over a local video system or to hold certain types of conversations, when there is the alternative of engaging in face-to-face communication. For these reasons, we chose to examine real work meetings over two wide-area video-conferencing systems currently being used by organizations to support remote collaboration. We examine how the channel properties of the video-conferencing media affect the characteristics of the spoken conversations.

The aim of this research is therefore to identify possible reasons for the lack

of success of video-conferencing technology. Our claim is that the properties of the communication channels in those wide-area systems prevent the execution of certain basic communication processes that may be crucial for certain collaborative interaction tasks.

Previous work has addressed the relationship between the properties of different communication media and the conversational characteristics they can support. This work has shown that the more closely a set of media approximates to face-to-face interaction in their properties, the closer the conversational style is to face-to-face interaction. This has been demonstrated for a number of different conversation characteristics, such as number of turns, interruptions, overlapping speech, and pausing (Argyle, Lalljee, & Cook, 1968; K. Cohen, 1982; Jaffe & Feldstein, 1970; Rutter & Stephenson, 1977). A major problem with this research, however, has been to operationalize the theoretical constructs used for defining media traits, such as "social presence" (Short et al., 1976), "cuelessness" (Rutter & Robinson, 1981), and "media richness" (Daft & Lengel, 1984). In contrast, in this study, we select measurable variables based on channel properties: (a) half-duplex versus full-duplex audio and lags in audio and (b) broadcast versus low-quality video. We study the effects of these variables on a number of spoken conversational characteristics that have independently been shown to be important in face-to-face interaction. Our analysis mainly focuses on the spoken aspects of conversation, although we include a brief discussion of the impact of channel properties on visual behavior.

We address how these channel properties of the video-conferencing technology affect the nature of spoken conversation in real meetings by comparing interaction in two wide-area systems with face-to-face conversation. We outline the critical characteristics of conversation and then examine how these differ for the following interaction technologies: (a) a video-conferencing system with half-duplex audio, transmission lags, and poor picture quality; (b) a high-quality video-conferencing system with duplex audio, no transmission lags, and full bandwidth video; and (c) face-to-face communication. We first describe the two video-conferencing systems, motivate the conversational characteristics, and then derive predictions about how those characteristics differ across the respective media as a result of the different channel properties of these systems.

1.1. ISDN System

The system is located at Hewlett Packard Laboratories Bristol, and most of the conferences held in Bristol are to the United States. Conferencing takes place over two Integrated Services Digital Network (ISDN) lines each at 64 kb/sec. Rate adaptation must take place because U.S. installations use a

public-switched 56 kb/sec digital network. Thus the available bandwidth is reduced from 128 kb/sec to 112 kb/sec. Of this, 16 kb/sec is used for audio with an additional amount for communication between CODECs (coder decoder). The amount of bandwidth available for video transmission is approximately 90 kb/sec.

The video signal is compressed by removing both spatially and temporally redundant data using a Compression Laboratories, Inc. Rembrandt CODEC. This process takes about 120 msec, with an equivalent time required for decompression at the other site. The audio signal is also compressed but has to be buffered to synchronize it with the video. In addition, there is the propagation delay of sending the data. This delay depends on whether a terrestrial or satellite link is used. For a terrestrial link, a propagation delay of approximately 170 msec in each direction can be expected to the West Coast of the United States, although this will vary depending on the route taken. A satellite link is much slower. The time taken for the signal to travel from the earth station to the satellite is 135 msec, and an equivalent time is taken to transmit the signal to the next earth station. To connect to the West Coast of the United States, two satellite jumps are required, which means a delay of 540 msec in one direction. Thus, allowing for compression and transmission, the lag between a person on one site speaking and the signal arriving at the other site can vary between 410 msec and 780 msec, depending on the propagation route.

The audio channel is half duplex, so the voice of only one person can be transmitted at any time. This is necessary to eliminate problems caused by echo or feedback when the sound from the loudspeaker is picked up by the microphone and retransmitted across the line. There are also occasional transmission problems with the system, causing brief disruptions both to the audio channel and the video picture.

The conference room is a converted meeting room containing a table at which three people can sit comfortably. Sitting at the table, users can see two stacks in front of them (see Figure 1). The first is directly in front of the table at a distance of approximately 9 ft and contains a 26-in. color monitor above which two cameras are located. The monitor displays the live picture of the remote location. Mutual gaze is not possible given the offset camera, and the distance and the video quality make remote eye gaze and head movements unclear. The perceived distance of the remote participants is difficult to evaluate, but it seems to depend on actual distance from the screen and the nature of the image of the remote participants, namely, whether the shot is full face, head and shoulders, or full body.

A small desktop control panel enables users to switch between cameras, focus, pan, and zoom. Participants control their local camera, choosing the view they wish to transmit. The control panel allows users to switch between close-up shots of a speaker and a view of the participants seated around the

Figure 1. **The layout of the ISDN video-conferencing room.**

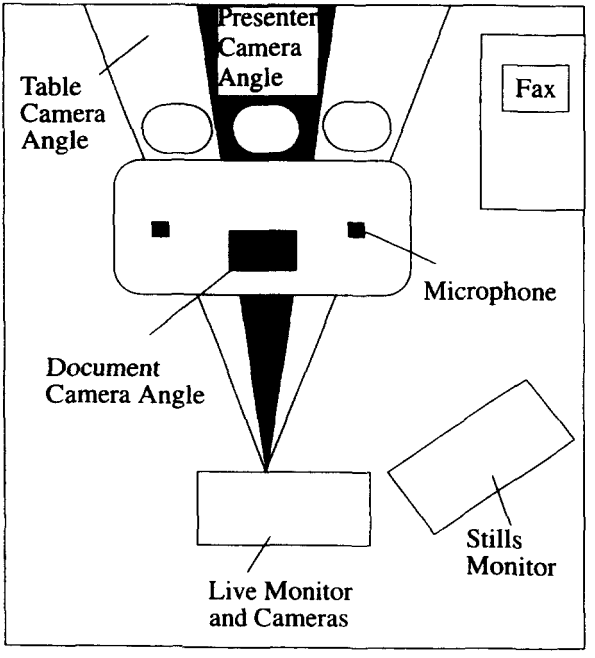


table. In practice, this alternating between views was rarely used. People tended to fix on a view showing all the remote participants seated at their table, displaying their head, arms, and upper bodies. On top and to the right of the cameras is a small 9-in. color monitor ("confidence monitor"), which displays the live picture that is transmitted to the remote site. This image has not been compressed and decompressed and is not, therefore, a true indication of what the other side sees. Thus, users are unaware of quality losses that may have occurred.

The second stack contains another large monitor, which is used to display stills from the remote site. *Stills* are individual static frames showing graphics and documents captured and transmitted using an overhead camera. It is not possible to gesture at these images. If gesturing is necessary, an alternative is to use the live channel for documents or graphics, but this means that the remote participants will be unable to view the images of the local participants.

1.2. LIVE-NET

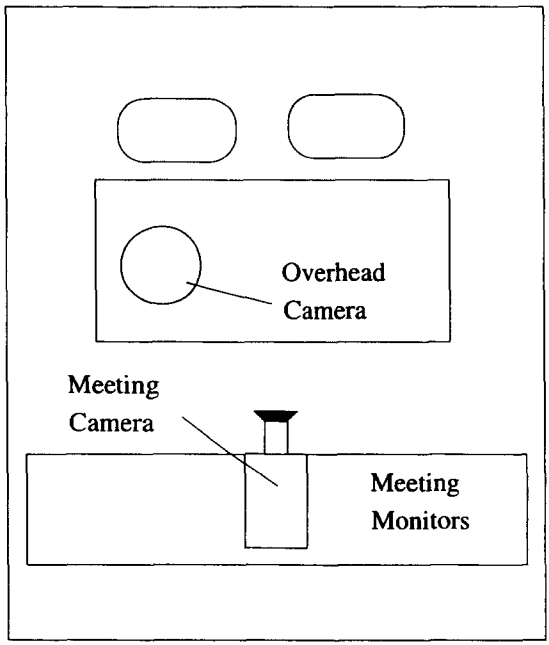
LIVE-NET is the London Interactive Video Education Network. The system has been in operation since 1987 and now connects eight sites to a

central video switch. The longest link is 42 km. It is used for intercollegiate lectures, seminars, and meetings. The colleges are dispersed over a large densely populated metropolitan area, making travel between the different colleges difficult and time consuming. LIVE-NET is an analogue system. Each site is connected by a pair of optical fibers, each carrying four full bandwidth video channels, with sound on a 6-MHz subcarrier, and a fifth lower bandwidth channel used for data up to 2 Mb/sec using a switched star topology. These five channels are frequency modulated onto a carrier, which is then converted into a 250-MHz multiplex and used to intensity modulate a laser diode. The result is a full motion picture with none of the frozen picture motion that is associated with some digital video systems.

As there is no video or audio processing, the time lag is simply the propagation time at the speed of light. Delays can therefore be measured in microseconds. The audio subsystem is full duplex. Several measures have been taken to eliminate feedback problems. First, there has been some acoustic treatment in the rooms to prevent loudspeaker sound being reflected back into the microphones. Second, a Shure AMS automatic microphone system is used, which has unidirectional microphones, does not pick up sound from the rear, and has a very fast switching system that ensures that only one or two microphones in the group are active at any time. Third, a frequency shifter (5 Hz) is used between the audio mixer and the network to limit howl reinforcement. The rooms used are typically lecture theatres or seminar rooms. An example layout is shown in Figure 2.

The participants sit at a table and face a set of four 20-in. monitors and a charge-coupled-device camera. A confidence monitor displays the outgoing picture. Figure 3 shows an example monitor set up. On the table is an overhead camera for the display of documents and a control panel for the cameras. The controls are used by the participants to select the camera to be used for output and to pan and tilt as necessary. As in ISDN, participants can directly select and control the images they transmit but not the images they receive. Where four or fewer sites are being connected, the sites are shown in full on the four monitors in front of the participants. If more sites wish to take part, a system called "chairman's control" is used. The sites are shown using "picture-in-picture" format in quadrants on the monitors as depicted by ABCD and EFGH in Figure 3. The chairman of the session chooses and displays the active speaking site on a full monitor. As the physical layouts for the sites vary, the perceived distances between participants also vary. In addition, as the number of participants increases at any single location, a wider angle of image is required, increasing perceived distance between participants. The broadcast quality video means that head movements are easily discernable. However, the offset camera means mutual gaze is difficult to achieve.

Figure 2. **Example layout for LIVE-NET.**



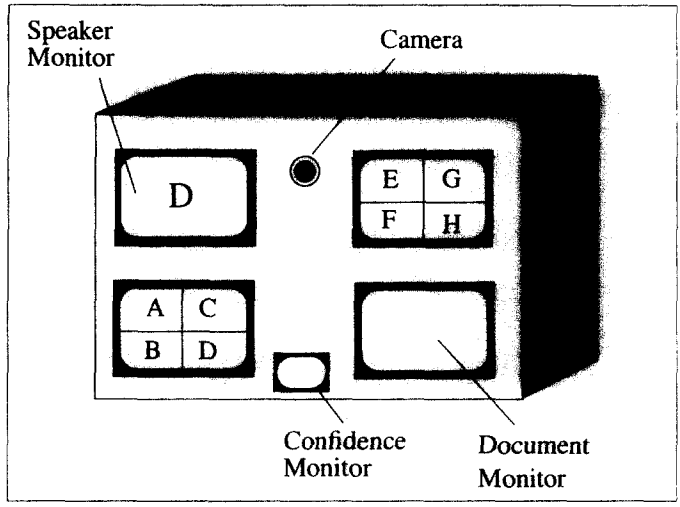
1.3. Face-to-Face Meetings

The face-to-face meetings took place in the conference rooms available on site at Hewlett Packard Laboratories Bristol. The room layouts were very similar to the one containing the ISDN system. Participants sat around tables approximately 6 ft long and 4 ft wide. Documents were shared by passing them around the table. An overhead projector was available but was not used in the meetings we observed.

2. FACE-TO-FACE COMMUNICATION AND PREDICTIONS ABOUT COMMUNICATION IN THE TWO VIDEO CONFERENCES

Communication is a joint activity that requires coordination of both process and content (Clark & Wilkes-Gibbs, 1986; Whittaker, 1992). To allow this coordination to take place, conversation is both incremental and interactive. A key aspect of interactivity is listener feedback. The speaker delivers utterances incrementally, while the listeners provide concurrent feedback that the conversation is on track, by giving both auditory backchannels (e.g., "mm," "uhu") and visual evidence in the form of head nods and eye gaze. This

Figure 3. **The LIVE-NET monitor set-up.**



positive concurrent feedback informs the speaker that he or she can rely on and build upon the listener's understanding (Clark & Schaefer, 1989; Duncan, 1972; Whittaker & Stenton, 1988; Yngve, 1970). Where this feedback is absent or even delayed, the speaker's ability to formulate efficient messages is reduced (Krauss & Bricker, 1967; Krauss & Fussell, 1990). Without feedback, speakers are unable to assume the message has been understood and may, therefore, attempt to clarify or reiterate points, sometimes unnecessarily, to ensure that the listener has not misunderstood. Absence or delay of feedback can therefore encourage the speaker to take long turns (Krauss & Bricker, 1967; Oviatt & P. Cohen, 1991). In normal face-to-face interaction, the flow of the speaker is not interrupted by backchannels because the audio channel is two way. There is a second sense in which communication is interactive: When a breakdown in understanding does arise, the listener can immediately interrupt the speaker for clarification or to register disagreement. Alternatively, by withholding verbal or visual feedback, the listener can indicate to the speaker that understanding is not guaranteed, and the speaker can then make the requisite modifications.

Central to the process of conversation is turn taking. In order to achieve the level of interactivity described earlier, speaker switches must be smooth and not disruptive to the overall flow of the conversation (Sacks, Schegloff, & Jefferson, 1974). How is this process achieved? There are a number of intonational, syntactic, pragmatic, and nonverbal devices that speakers use to indicate that they are about to finish their conversational turn. Listeners also use nonverbal devices to indicate that they wish to speak (e.g., leaning

forward or achieving mutual gaze). The fact that listeners are able to predict when speakers are about to finish means that there can be very low latencies, with speaker-switching pauses varying between 620 msec and 770 msec (Jaffe & Feldstein, 1970). In some cases, there is no pause, and overlaps occur between speaker transitions. Speakers also sometimes give overt cues to select the next speaker, such as naming an individual or directing a question to them. On other occasions, turn taking is not smooth. Speakers sometimes attempt to hold the conversational floor, and listeners make unsolicited bids for the floor (Levinson, 1983). These lead to overlapping speech, although whenever overlaps do occur they are usually resolved quickly with one speaker dropping out rapidly.

Verbal messages are often accompanied by multiple nonverbal cues such as gaze, facial expression, posture, and physical proximity. These serve a number of possible functions. They may help the listener to identify the meaning of the utterance (Argyle et al., 1968; Jaffe & Feldstein, 1970); they may also support smooth speaker transition by the use of eye gaze and posture change, although the importance of these cues is the subject of much debate (Williams, 1977); finally, listeners' nonverbal reactions may offer the speaker information about the effects his or her speech is having on the audience (Short et al., 1976).

In face-to-face interaction, all participants in principle have equal access to the conversational floor, although there are external factors, such as knowledge, that can influence participation levels (McGrath, 1984, 1990).

From what has been discussed so far, it can be seen that face-to-face conversation exhibits characteristics that depend on three properties of the communication channels: (a) low transmission lags—that is, messages are received almost instantaneously by listeners; (b) two way—for example, feedback can be produced at the same time as the speaker's utterances; and (c) multiple modalities—that is, both verbal and visual channels are used (Whittaker, 1992). How will the properties of spoken conversation be changed by communicating using technologies that do not have these channel properties.

The ISDN system introduces a transmission lag of between 410 msec and 780 msec for both audio and video, and a half-duplex (one-way) line for audio. In addition, the ISDN system allows only limited visual cues, because the picture quality is poor and subject to jitter and occasional frame loss. Figure 4 summarizes our expected findings for a number of spoken conversational characteristics. These predictions are based on the differences between the system channel properties and the properties of face-to-face interaction. We first define the characteristics and give explanations of our predictions.

An inherent limitation of our method is that we did not attempt to measure the effectiveness of the communication across different media as other

Figure 4. **Expected characteristics relative to face-to-face interaction for the two video-conferencing technologies.**

	ISDN	LIVE-NET
Backchannels	Fewer	Same
Interruptions	Fewer	Same
Overlaps	(?)	Same
Handovers	More	Same
Turn size	Larger	Same
Turn distribution	More	Same

laboratory studies have done (Chapanis, 1975; Morley & Stephenson, 1969, 1970, 1977; Wichman, 1970). With real-life meetings, it is not clear what is an appropriate objective measure of successful communication, and we cannot easily compare the success of the different meetings. Other work suggests, however, that the types of communication characteristics measured here have implications for task outcomes. Laboratory studies have shown that lack of support for interactive processes such as backchannels and interruptions has effects on outcome measures such as time to solution and participants' understanding (Kraut, Lewis, & Swezey, 1982; Oviatt & P. Cohen, 1991). The characteristics can therefore be seen as indirectly measuring communication effectiveness.

2.1. Backchannels

For the purpose of this study, only auditory backchannels were measured and not head nods or gaze behavior. Backchannels are short feedback utterances, produced by the listener to indicate functions such as attention, support, or acceptance of the speaker's message (Yngve, 1970). Examples of utterances serving as backchannels are "mm," "uhu," "right," "okay," and "yes," although these utterances can sometimes have other functions than those described. They are often delivered with split-second timing, for example:[1]

A: . . . in the absence of a of of a task for any particular set of users if we take the general task, =
B: =right=
A: =personal information management, =
B: =right=
A: =getting at documents whether they're faxes or . . .

[1] Transcription conventions are described in Section 3.3.

or concurrently, as follows:

A: if this is a register which has got too high a value [then =
B: [mm
A: =when I am testing this register [which should fail high
[mm

In face-to-face interaction, these backchannels are produced by the listeners concurrently with, or directly after, speaker input. However, in ISDN, the audio channel is half-duplex, and there is a substantial transmission lag. What is the impact of these factors? The lag means that at the remote location the backchannel will not be concurrent with or directly follow the material it is intended to reinforce. This serves to reduce its communicative impact and may disrupt the speaker at the remote location by its late arrival. In addition, the half-duplex line means either (a) the backchannel is suppressed altogether, or (b) it takes the audio channel from the remote speaker, so that information generated at the remote location is not received locally. All these factors should lead to fewer backchannels in ISDN.

By contrast, in the LIVE-NET system, the audio channel is full duplex, and transmission is nearly instantaneous. This allows both concurrent and timely backchannels to be delivered. This leads us to the following prediction: In ISDN, we expect fewer backchannels, but in LIVE-NET, backchannels should occur as frequently as in face-to-face interaction.

We then made predictions for other instances of simultaneous speech. Simultaneous speech can arise for a number of different reasons (Levinson, 1983). It is therefore important to distinguish between two different classes of simultaneous speech: (a) *overlaps,* that is, when there was a clearly identifiable reason why the next speaker should have broken into the current speaker's utterance, for example, if the speaker was having difficulty completing an utterance and (b) *interruptions,* which were defined, by exclusion, as instances of simultaneous speech that could not be classified as an overlap, when there was no reason for the next speaker to have broken into the current utterance. We provide detailed definitions of overlaps after discussion interruptions.

2.2. Interruptions

Interruptions are those instances of simultaneous speech when there is no indication by the first speaker that he or she is about to relinquish the floor. As such, they are deliberate attempts to gain the conversational floor without the prior consent of the current speaker, and they always occur in midturn.

A: my worry would [be my worry would be that
B: [No I don't I don't I'm not saying this person has to have

The same predictions hold here as for backchannels. In ISDN, with half duplex and transmission lags in audio, we should witness reduced attempts to interrupt the speaker. The half-duplex line means either that interruptions are highly disruptive, in that they take the channel and mask whatever the speaker is saying, or that they are suppressed and never transmitted to the remote location. In addition, the transmission lag may mean that, by the time the interruption arrives at the remote location, the speaker has already gone beyond the relevant material. This can lead to further disruptions, for example, if the interruption deletes material that then has to be repeated and turntaking re-established.

Again we expect that in LIVE-NET, with full-duplex audio and almost zero lag, the interruptions will be much easier to achieve successfully. They can be delivered in overlap with the speaker, and the absence of lag means that the conversation has not moved on by the time they are transmitted. This leads us to the following prediction: In ISDN, half-duplex audio combined with lags in audio should produce fewer interruptions. In LIVE-NET, interruptions should occur as frequently as in face-to-face interaction.

2.3. Overlaps

Overlaps are instances of simultaneous speech that follow signals speakers give indicating that they may relinquish the conversational floor (Levinson, 1983). We made predictions for three different types of overlaps.

1. *Projection/completion:* This type of overlap occurs when the next speaker anticipates that the current speaker is about to finish or tries to help the "forward movement" of an ongoing utterance (Clark & Wilkes-Gibbs, 1986). In projecting the possible finish by the current speaker, the next speaker may recognize that the message of the current speaker is complete, although the utterance has not finished.

A: initially that's true but I wonder how the market will shape up
[over time
B: [well you have to have a second punch behind it of course . . .

The next speaker may overlap in an attempt to complete the current speaker's utterance. This can occur when the next speaker perceives that the first is having some difficulty in completing his or her turn. Under these circumstances, the overlapping utterances usually contain very similar information.

A: ahm how the work [ho how how it works=
B: [pans out

A: =with [how the work pans out between the people
B: [pans out

We expected fewer projections/completions in ISDN, because deliberate attempts to complete or overlap the end of the current speaker's turn may either delete the relevant material or arrive at the remote location after the speaker has already finished the turn. In LIVE-NET, we expected equivalent numbers of projections as in face-to-face interaction, because low-lag full-duplex audio should support this type of intervention. Therefore, we predicted fewer projections in ISDN, with equal levels in LIVE-NET and face-to-face meetings.

2. *Floorholding:* This occurs where the next speaker tries to take the floor while the current speaker attempts to hold the floor while producing utterances that do not contain any information (Jefferson, 1984). Examples of this can range from self-repetitions to function word repetition ("so . . . so") or dysfluent speech ("er . . . er"):

A: that's not true you is it you could have an annotation which can either be a structured annotation or a free text annotation
[so [so
B: [some but [somebody has got to own the interface the top level interface=

We expected less floor holding in ISDN, because of adjustments by speakers. Speakers should be less likely to hold the floor because they want to avoid the disruptive effects of the half-duplex line on simultaneous speech, in deleting one speaker. There should be no such constraints with LIVE-NET, where floor holding should be possible without such disruptive effects. We predicted less floor holding in ISDN, with equal levels in LIVE-NET and face-to-face meetings.

3. *Simultaneous starts:* These are instances of simultaneous speech when two participants concurrently begin a new turn. These occur when two or more speakers compete for the floor when the previous speaker has just finished. In some instances, this may include an attempt by the original speaker to resume. This can happen when the original speaker yields the floor and, after some time has elapsed, believes there to be no contenders and so begins a new turn (Sacks et al., 1974).

A: well they'd be better be quick cos the nineteenth is next Wednesday
B: [next week isn't it
C: [That's right exactly

In ISDN, we should expect more simultaneous starts because of the problems that participants have in timing speaker switches. Because of the lag, and the desire not to overlap the end of the previous speaker's turn, listeners may deliberately wait to respond to ensure that the speaker has finished. Given the slow response, the original speaker may assume that no other person wants to speak and may then begin to speak again. Meanwhile, at the remote location, another participant already may have begun to speak. This situation conspires to produce simultaneous starts. The situation is different in LIVE-NET, where low lag times and full duplex should allow equivalent numbers of simultaneous starts as in face-to-face interaction. We predicted more simultaneous starts in ISDN, and LIVE-NET and face-to-face interaction should be equivalent.

2.4. Explicit Handovers

These occur when speakers signal that they intend to relinquish the floor using explicit verbal cues such as: (a) the use of questions; (b) tagging, using stereotyped questions such as "isn't it?", "arent they?", or statements such as "you know" or by the addition of redundant information on the end of a turn, for example;

> A: . . . ahm now I don't have I don't have a problem with that at all but it but it wouldn't it would not mean that we have at any one point one interface you know it would just be [you know

and finally (c) naming the next speaker (Levinson, 1983; Sacks et al., 1974).

We expected that speakers would try to alleviate the problem of speaker switching in ISDN by explicitly signaling that they had finished their turn. In ISDN, we therefore expected more instances of questions, tagging, and naming of the next speaker. This should not be true in LIVE-NET, where speaker switching should be unproblematic and explicit handovers unnecessary. We predicted more formal handovers in ISDN and that LIVE-NET should be equivalent to face-to-face interaction.

2.5. Total Number of Turns and Turn Length

Turns are defined as attempts by speakers to gain the conversation floor. We expected a number of factors to combine to increase turn length in ISDN. Given the problems of switching speakers, we expected that switches would occur less frequently and hence produce a greater number of longer turns. In addition, the difficulty of backchanneling and interrupting also would reduce the number of quick-fire interchanges serving to indicate or clarify under-

standing. The absence of feedback and clarifications may also lead speakers to overelaborate and supply redundant information. This should also increase turn length. We therefore expected the ISDN meetings to have more of the characteristics of formal presentations or lectures where speakers deliver large amounts of material as an uninterrupted monologue. In contrast, in LIVE-NET, there should be no problems with rapid speaker switching or quickfire exchanges, and turn number and length should be comparable with face-to-face interaction. We predicted (a) that ISDN should have large turns and infrequent changes of speaker and that (b) in LIVE-NET, turn length and frequency of switching should be equivalent to face-to-face interaction.

2.6. Turn Distribution

Finally, we expected that turns might be unequally distributed in the different technologies. In each video conference, it is possible for people to communicate with people at the local site (via standard face-to-face interaction) as well as with the remote site using the system. Given the difficulty of interacting over ISDN, our informal observations suggested that groups attempt to manage this problem by channeling their responses to the remote location through one specific individual at each location. We therefore expected that these local coordinates would dominate their group's contributions: The overall distribution of turns would be unequal, with these individuals having more turns than the average group member. In contrast, we expected turns to be more evenly distributed in the LIVE-NET meetings. We predicted (a) that turns should be unequally distributed in ISDN, with two speakers, one at each location, producing more turns than other group members but that (b) turns should be equally distributed in LIVE-NET and face-to-face interaction.

3. METHODOLOGY

3.1. Recording Method

The ISDN video conferences were recorded by placing a video camera next to the monitor and camera stack in the conference room. An additional monitor displaying the remote participants was placed beneath the table at which the participants sat. The video camera thus captured the local participants, with the remote participants visible on the monitor under the table. The stills screen was not monitored. The LIVE-NET meetings were recorded at the central video switch site. The picture on each of the two quadrant monitors was recorded onto videotape. The face-to-face meetings were audiotaped. An observer was present at each meeting who noted any events not picked up on tape.

Figure 5. **Primary and secondary functions and activities for each of the 14 recorded meetings.**

Functions/ Activities	ISDN	LIVE-NET	Face to Face
Primary	Information exchange	Information exchange	Information exchange
Secondary	Appraisal, generating ideas	Appraisal, problem solving	Task allocation, tactical decision making, problem solving
Primary	Information exchange	Information exchange	Information exchange
Secondary	Appraisal, generating ideas	Task allocation, decision making	Problem solving
Primary	Information exchange	Information exchange	Information exchange
Secondary	Task allocation, problem solving	Appraisal, problem solving	Appraisal, generating ideas
Primary	Information exchange	Information exchange	Information exchange
Secondary	Appraisal, generating ideas, decision making	Problem solving	Work-related gossip
Primary	Information exchange		Information exchange
Secondary	Appraisal, generating ideas		Appraisal

3.2. The Meetings

Five ISDN video conferences, four LIVE-NET meetings, and five face-to-face meetings were recorded and analyzed. All meetings were scheduled for work-related reasons and were not arranged for the study. We attempted to identify analogous groups and meetings for the three conditions. Details of the functions and activities of the meetings, based on the Description and Classification of Meetings classification of business meetings (Short et al., 1976), are shown in Figure 5. This shows that all the meetings were cooperative in nature, with their main function being to exchange information. Secondary functions and activities such as problem solving and idea generation also took place.

The face-to-face and ISDN meetings were to report progress where participants described the work that they had recently been doing. In some cases, this involved the demonstration of software. These meetings centered around project teams with one or two project managers being present. The LIVE-NET meetings were the coming together of representatives from different colleges. Participants from the various colleges gave updates on the developments and progress made at their site.

In most cases, the participants knew each other before the meetings, although in a few of the video conferences the people at either end of the link had not all previously met face to face. We could not control for certain

parameters of familiarity; for example, participants at either end of a video-conference link are likely to know each other better and have a greater understanding of local work. When possible, however, we tried to reduce this problem by our choice of face-to-face meetings: Two of the face-to-face meetings were between collaborators from the United States who were visiting the United Kingdom, and therefore had little day-to-day contact. All participants were familiar with using video conferences. As they already had experience with the systems, we did not expect participants' conversational strategies to alter significantly during the meeting. We therefore did not analyze whether conversational behaviors changed in the course of each meeting.

In the face-to-face and ISDN conditions, there was a mixture of agenda-based and non-agenda-based meetings. All the LIVE-NET meetings were agenda based.[2] Both the face-to-face and ISDN meetings had an average of six participants. For face to face and ISDN, the smallest meeting had four participants, and the largest had seven. LIVE-NET meetings were slightly larger. The largest had nine participants, one had eight, and the remainder seven. With the exception of one meeting, in which three sites took part, all the LIVE-NET meetings took place using four sites. Typically, all meetings lasted between 1 and 2 hr.

3.3. Analyzing the Results

A 20-min segment from the middle of each meeting was transcribed in detail. Each segment was taken 20 min from the start of the meeting so that differences in the opening sequences would not bias the results. For the same reason, the closing sequences were not analyzed. From these transcripts, measures were taken of the number of utterances, the number of words per utterance, backchannels, interruptions, overlapping speech, and handovers according to the previously given definitions.

The data were transcribed using a simplified version of the system developed by Jefferson for conversational analysis (Atkinson & Heritage, 1984). Our aim was to capture those characteristics of conversation associated with speaker transition, for example, the positioning of overlapping speech. We did not code phonetic information (e.g., prosodic turn-switching cues) such as question intonation. Sentences were transcribed as they were spoken, including any syntactical errors. The following extract shows some of the conventions in use and is followed by a glossary of the symbols.[3]

A: So it's it's moving to Italy ahm and ah we're not ah we I got a thing

[2] The participants of course chose whether or not the meeting was agenda based.

[3] The glossary is a shortened version of that given in Jefferson (1987).

B: (bet they) get lost on the way (.) [( )
A: [we've got some market stuff which eh
B: Oh [yes
A: [Tim I [ga I gave to Timmy oh it's circulating is it yeah [it seemed it was=
B: [it's circulating
B: [( )
A: =quite interesting ah

[A single left square bracket indicates the point of overlap.
= Equal signs, at the end of one line and the beginning of the next, indicate no gap between the two lines.
(.) A dot in parentheses indicates a tiny gap within or between utterances.
( ) Empty parentheses indicate the transcriber's inability to hear what was said.
(word) Parenthesized words are especially dubious hearings.
(( )) Double parentheses contain transcribers' descriptions.

We recorded at one location only so that assessments of simultaneous speech were analyzed only with respect to that location, although, as we will see, it is sometimes possible to determine the points at which simultaneous speech occur at the remote location.

The transcripts did not replace the tapes for scoring purposes but were used in conjunction with the tapes. We also conducted a reliability analysis, with two judges independently scoring two meetings in each condition, a total of 1,054 turns in total. Both judges tried to identify every instance of backchannels, interruptions, overlaps, and formal handovers. Reliability scores were measured as:

$$\frac{\text{(Number of Agreements} - \text{Number of Disagreements)}}{\text{(Number of Agreements} + \text{Disagreements)}}$$

These were as follows: Backchannels (0.91), overlaps (0.74), interruptions (0.62), and handovers (0.92). We also compared reliability of coding across the three conditions and found coding was most reliable for ISDN (0.89) and LIVE-NET (0.88) but slightly less reliable for face-to-face interaction (0.79). To evaluate the success of our coding, we computed kappa (J. Cohen, 1960) for each condition. The respective kappas for each condition across the four categories were as follows: ISDN = .92, $p < .001$; LIVE-NET = .93, $p < .001$; face to face = .86, $p < .001$. This indicates the reliability of our coding scheme.

Figure 6. **Mean number (and mean standard error) of backchannels and interruptions per meeting, showing levels of statistical difference for the three conditions.**

	ISDN	LIVE-NET	Face to Face	p
Backchannels	7.00[a,b]	30.50[a]	60.80	$<.001$
	(2.1)	(6.1)	(9.1)	
Total interruptions	1.40[a,b]	13.00	18.60	$<.01$
	(0.9)	(2.2)	(3.7)	
Interruptions excluding channel breaks	0.20[a,b]	11.75	18.60	$<.01$
	(0.2)	(3.3)	(3.7)	

[a]Significantly different from face to face in post hoc analysis of variance test.
[b]Significantly different from LIVE-NET in post hoc analysis of variance test.

4. RESULTS

4.1. Overview

In what follows, we present statistical analyses for each prediction followed by representative examples from the interactions to illustrate our claims. All analyses apply to the 20-min segment we analyzed for each meeting and not to the whole meeting.

4.2. Backchannels and Interruptions

Figure 6 shows the distribution of backchannels and interruptions in the three conditions. Mean levels of backchanneling were low in ISDN compared with face-to-face meetings (7.00 vs. 60.80), confirming our prediction that people in ISDN would avoid backchannels. The finding that backchannels were also reduced in LIVE-NET compared with face-to-face meetings (30.50 vs. 60.80) was not predicted, and we discuss reasons for this in the Conclusions section. The differences were analyzed in a one-way analysis of variance. The overall difference was significant, $F(2, 11) = 18.16, p < .001$, with backchannels being more frequent in face-to-face than in LIVE-NET, $F(1, 7) = 15.82, p < .01$, which are in turn more frequent than in ISDN, $F(1, 7) = 6.77, p < .05$).[4]

The example that follows indicates why backchannels were reduced in ISDN as shown in Figure 6. Where backchannels do occur, they can lead to a disruption of the flow of the speaker. In this instance, B responded with a backchannel to A's comment, "it would be interesting to see if ah we could marry that." Locally, the backchannel was placed after the suggestion

[4] Post hoc analysis of variance tests have been administered to make pairwise comparisons between the conditions, as recommended by Kirk (1982).

overlapping A's "because." However, because of the lag, A does not receive the backchannel until some words later leading him to hesitate ("ahh").

A: . . . portion of the interface that's been put there it would be interesting to see if ah we could marry that [because that was the intent of the **ahh** an original interrogation =

B: [mm

Again, as predicted, interruptions were also significantly less frequent in ISDN, as shown in Figure 6. In the face-to-face meetings, almost 10% of turns were interruptions compared with less than 2% in ISDN. This difference is statistically significant, $F(2, 11) = 12.22$, $p < .002$, with interruptions being more frequent in both face to face and LIVE-NET than in ISDN: face to face versus ISDN, $F(1, 8) = 20.66$, $p < .05$; LIVE-NET versus ISDN, $F(1, 7) = 27.83$, $p < .01$; face to face versus LIVE-NET, $F(1, 7) = 1.49$, $p > .05$.

This occurs despite occasional technical problems in ISDN, producing brief losses of audio and video for several hundred milliseconds. Many of the interruptions in ISDN followed such problems and represented requests for a repetition of information lost during the break in the channel. There were fewer of these problems in LIVE-NET and none in face to face, where the majority of interruptions are to clarify what the speaker has said. A second analysis removed interruptions following line breaks and showed that there were large differences between the media, $F(2, 11) = 11.95$, $p < .002$, with both face to face and LIVE-NET having more interruptions. All other results were equivalent for this second analysis.

The following examples illustrate the problem with line breaks in ISDN. An interruption follows a line break in ISDN, with B interrupting A to ask him to repeat what was lost. This example also shows the problem of resuming turn taking following interruptions: Because transmission of information is not instantaneous, A is unaware of what information has been lost. It is therefore necessary for B to indicate to A what portion of the sentence must be repeated.

A: . . . software components would be for a database and eh at what level in the system are those delivered I think there's clearly ah a need for a certain amount of communication ((break-up)) for the stations to talk with each other [and ( )

B: [Sorry we missed that from communication

A: okay for the the communication protocol that be . . .

In contrast, an interruption during LIVE-NET causes no problems for the speakers. No information is lost, so this does not need to be repeated and A simply drops out leaving B to take the floor.

Figure 7. **Mean percentage (and mean standard error) of turns occurring in overlap, showing levels of statistical significance for the three conditions.**

	ISDN	LIVE-NET	Face to Face	p
Total overlaps per turn	9.6%	12.3%	10.1%	ns
	(1.5)	(1.5)	(0.7)	
Overlaps resulting from projection/ completion per turn	2.9%[a,b]	9.2%	7.3%	$<.01$
	(1.1)	(0.7)	(0.9)	
Overlaps occurring during floor holding per turn	0.0%[a]	0.6%[a]	1.8%	$<.01$
	(0.0)	(0.3)	(0.3)	
Overlaps from simultaneous start	6.7%[a,b]	2.5%	1.0%	$<.01$
	(0.9)	(1.1)	(0.5)	

[a]Significantly different from face to face in post hoc analysis of variance test.
[b]Significantly different from LIVE-NET in post hoc analysis of variance test.

A: because we have people actively using Omega we have Beta both of which we would lose [(and we)
B: [that's a lot of money just to pay for those packages

4.3. Overlapping Speech

Overlaps were analyzed in terms of their frequency per turn. This was to allow for the fact that there were many fewer turns and speaker switches in ISDN, and the chance of generating an overlap is clearly dependent on the number of speaker transitions. Figure 7 shows that the overall number of overlaps per turn did not differ substantially, $F(2, 11) = 1.22$, $p > .05$. However, the different types of overlaps showed different distributions in the three conditions.

For projections, we found as we predicted that there were differences between the conditions, $F(2, 11) = 11.90$, $p < .002$, with more overlaps following projections in the face-to-face and LIVE-NET media. The combination of a half duplex line and lags seems to combine to reduce projections in ISDN, with listeners avoiding overlapping speech even when this could assist the speaker in composing his or her message. Projections were reduced in ISDN relative to face to face and LIVE-NET, as predicted: face to face versus ISDN, $F(1, 8) = 10.27$, $p < .05$; face to face versus LIVE-NET, $F(1, 7) = 2.56$, $p > .05$; ISDN versus LIVE-NET, $F(1, 7) = 20.59$, $p < .0001$.

We found that floor holding was much more frequent in face-to-face meetings than in both LIVE-NET and ISDN, $F(2, 11) = 12.19$, $p < .002$: face to face versus ISDN, $F(1, 8) = 33.15$, $p < .0001$; face to face versus LIVE-NET, $F(1, 7) = 5.56$, $p = .05$; ISDN versus LIVE-NET, $F(1, 7) = 3.58$, $p > .05$. Strikingly, there were no examples of floor holding in the

ISDN meetings. Contrary to our expectations, however, we found that floor holding was also reduced in LIVE-NET compared with face-to-face meetings, and we discuss this in the Conclusions section.

The picture was different for simultaneous starts. These can be regarded as breakdowns in the process of speaker switching brought about by ISDN lags. As predicted, they were much more likely in the ISDN medium than in both LIVE-NET and face to face, $F(2, 11) = 13.41, p < .001$: face to face versus ISDN, $F(1, 8) = 29.83, p < .0001$; face to face versus LIVE-NET, $F(1, 7) = 2.02, p > .05$; ISDN versus LIVE-NET, $F(1, 7) = 8.79, p < .05$.

Examples serve to illustrate these effects. In the face-to-face and LIVE-NET conditions, overlaps occurred mainly when the listener completed the speaker's utterance or projected a turn end as in the following LIVE-NET example. Here B understands the question after A's "anything" and so begins his turn overlapping the end of A's turn.

A: have you got an anonymous FTP or anything [that we can use
B: [I'm I'm not set up for that but I'll send it I mean cos I haven't got my act together yet . . .

In contrast, overlaps in ISDN are most likely to result from two speakers starting simultaneously. In the following example, A and B are at the local site and C is at the remote site. When he hears A finish, C assumes he can take the channel. However, because of the transmission lag, he is unaware that B has already begun to speak. Both B and C then drop out to allow the other to speak. It is then necessary for B to indicate to C that he should continue, using a formal handover.

A: . . . the icon will be ungreyed from all the displays so that other people may open it
((pause))
B: He doesn't have [the
C: [ju just to
B: go ahead

Another problem with ISDN is that, once turn taking is disrupted, it is difficult to re-establish it because of the role of split-second timing in this process. The result is that simultaneous starts tend to occur in batches. Given that the normal mechanisms for repair are disrupted over a half-duplex, lagged line in ISDN, how are these clashes resolved? Unlike the face-to-face situation where one speaker drops out, it is usual in ISDN for both speakers to stop and then for one to be granted the floor, either verbally by being told "go ahead" or visually by using hand gestures. If this does not occur, a second

Figure 8. **Mean percentage (and mean standard error) of turns ending in an explicit handover, showing levels of statistical significance for the three conditions.**

	ISDN	LIVE-NET	Face to Face	p
Total handovers	30.8%[a]	21.2%[a]	8.8%	$<.01$
	(5.7)	(2.0)	(0.8)	
Handovers by question	23.8%[a]	18.2%[a]	7.7%	$<.01$
	(3.4)	(1.7)	(1.0)	
Handovers by tagging	4.3%	1.9%	0.8%	ns
	(1.6)	(0.6)	(0.3)	
Handovers by name	2.7%[a]	1.1%	0.4%	$<.05$
	(0.9)	(0.4)	(0.2)	

[a]Significantly different from face to face in post hoc analysis of variance test.

or third clash can follow. In the following ISDN example, both speakers stop then start again. This is finally resolved by a third party telling the remote speaker to go ahead.

A: the visual [appearance
B: [uh just out of curiosity wh
A: the appearance of [that
B: [just out of curiosity what differ- ence
( )
C: go ahead

To summarize, there are no differences in the combined number of overlaps, but the subtypes of overlaps are differently distributed in the three conditions. Overlaps occur face to face mainly because of projections and floor holding, in LIVE-NET because of projections, and in ISDN because of simultaneous starts.

4.4. Explicit Handovers

We predicted that speakers would try to remedy the problem of speaker transition in ISDN by explicitly handing over the floor. Figure 8 shows turns ending in questions, tagging, and naming of the next speaker. Again this was measured in terms of frequency per turn because of the different numbers of turns across conditions. As we predicted, there was a greater number of each of these formal handovers in ISDN compared with face to face, because of the need to explicitly manage speaker transitions, $F(2, 11) = 9.46$, $p < .004$. Contrary to our expectations,, however, we found the same overall pattern of formal handovers in LIVE-NET as in ISDN: face to face versus ISDN, $F(1,$

8) = 14.70, $p < .01$; face to face versus LIVE-NET, $F(1, 7) = 38.22$, $p < .001$; ISDN versus LIVE-NET, $F(1, 7) = 2.06$, $p > .05$. Again we discuss this unexpected result in the Conclusions section.

Further analysis of the different classes of handover indicated that handovers using direct questions were more frequent in both video conferences, $F(2, 11) = 13.14$, $p < .001$: face to face versus ISDN, $F(1, 8) = 21.42$, $p < .001$; face to face versus LIVE-NET, $F(1, 7) = 30.47$, $p < .001$; ISDN versus LIVE-NET, $F(1, 7) = 0.21$, $p > .05$. Tagging was equal in all three conditions. Handovers by naming the next speaker were more frequent in ISDN than in face-to-face meetings, $F(2, 11) = 4.09$, $p < .05$: face to face versus ISDN, $F(1, 8) = 6.57$, $p < .05$; face to face versus LIVE-NET, $F(1, 7) = 2.21$, $p > .05$; ISDN versus LIVE-NET, $F(1, 7) = 2.39$, $p > .05$.

In the ISDN condition, participants used questions at the end of long turns to encourage speaker transition, for example:

A: . . . there are only two possible choices either there is an inpu an input file or there is none or rather either it is empty or not. If it is if there is data in it then the job runs correctly otherwise all the subsequent steps test the condition code and if it is different from zero then they don't run as simple as that any ah ((pause)) any counter indication on your end?

In some instances, names were used to address the question to a particular individual as in the following two LIVE-NET examples:

A: How much does that cost Mike?
B: Are we still on state of play Alan?

Tagging such as "is that okay" or "you know" or redundant information was equally frequent in all media. Here a participant in an ISDN conference ends a turn with a tag question that both facilitates speaker transition and acts as a check for understanding:

A: . . . The only thing you have to change is ahh the the step card and thats it its one line in this JCL. Have I made myself clear?

Participants had other methods of explicitly handing over control in video conferences. They were observed raising their hands as an indication of a desire to speak. In one ISDN conference, participants agreed to use their hands throughout the meeting to indicate they wished to speak.

Figure 9. **Mean number (and mean standard error) and size of turns, showing levels of statistical difference for the three conditions.**

	ISDN	LIVE-NET	Face to Face	p
Turns per meeting	74.2[a,b]	180.0	199.2	$<.01$
	(13.6)	(31.7)	(17.3)	
Words per meeting	3212.0	3529.5	3386.8	ns
	(251.2)	(165.6)	(283.5)	
Turns by participant	12.37[a,b]	23.77	34.82	$<.01$
	(2.01)	(4.29)	(6.07)	
Words by participant	535.3	455.5	603.2	ns
	(109.0)	(98.1)	(114.0)	
Words per turn	43.61[a,b]	19.23	17.08	$<.001$
	(3.42)	(1.34)	(0.98)	
Words per turn, not including turns of less than five words	62.19[a,b]	30.00	31.30	$<.001$
	(4.47)	(2.02)	(1.66)	

[a]Significantly different from face to face in post hoc analysis of variance test.
[b]Significantly different from LIVE-NET in post hoc analysis of variance test.

4.5. Turn Size

We predicted that the problems encountered in speaker transition, coupled with listeners' reluctance to interrupt or provide backchannels, would result in longer turns in ISDN. Figure 9 shows the number of turns taken and average word length. We analyzed the total number of turns on a meeting-by-meeting basis and also for each participant. Typically, the meetings held over ISDN were characterized by fewer turns of greater length. There were significantly fewer turns per participant in ISDN compared with LIVE-NET and face to face, $F(2, 86) = 6.48$, $p < .002$: ISDN versus face to face, $F(1, 8) = 13.02$, $p < .001$; ISDN versus LIVE-NET, $F(1, 7) = 5.68$, $p < .05$; face to face versus LIVE-NET, $F(1, 7) = 2.27$, $p > .05$. The complementary result was that the number of words per turn was significantly greater in ISDN than in the other two media, $F(2, 21) = 60.15$, $p < .001$: ISDN versus face to face, $F(1, 8) = 101.66$, $p < .0001$; ISDN versus LIVE-NET, $F(1, 7) = 62.82$, $p < .0001$; face to face versus LIVE-NET, $F(1, 7) = 1.76$, $p > .05$. It is possible that these effects are due to the reduction of brief turns in ISDN. To investigate this, we repeated the analysis excluding all turns of fewer than five words, but both effects were still present.

These differences in turn size were observed despite the fact that there were no overall differences in the total number of words per meeting in each condition, $F(2, 11) = 0.39$, $p > .05$. Although the total number of words remained constant across conditions, the differences between the conditions lay in how the words were distributed across turns. These results strongly support our prediction that ISDN would produce a "lecture-like" interaction,

with speakers holding the floor for lengthy uninterrupted monologues. In contrast, on both LIVE-NET and face to face, we observed many more short turns with a higher frequency of interruptions and backchannels.

The following examples show typical interactions for ISDN, LIVE-NET, and face to face. The first clearly shows the lecture-like style in ISDN. Here speakers supply large amounts of uninterrupted information, with transitions often being accompanied by pauses, and there is little evidence of incremental checking of listener understanding.

A: . . . and ah essentially what they are doing is they're ah comparing preoperative waves with with the actual interoperative ones they're looking at what the guy was like before they did anything to him to what he's like now ahm and its kind of you know they sort of look at this thing and they sort of say its its a bit different isn't it type of thing and your thinking yeah it is I suppose and and then they sort of say well actually I think I'll tell him but I I don't quite I don't I haven't quite got a grip on what the algorithm was they were sort of saying well it looks similar and look its sort of kind of moved that that way a bit ahm and that's how they were doing delays it was it was very approximate. ((pause))

B: Yeah I mean the two things that they seem to be looking at predominantly are latency over the preoperative signal and also some characteristics which we couldn't fathom which were like the shape of the waves you know something to do with peaks and you know like when they hit or you know how their characteristics changed and you know in some way that related to ahm you know the particular nerve that was being tested but . . .

In contrast, LIVE-NET has many more short turns, with conversational exchanges being incremental and interactive.

A: Is there any significant difference?

B: ahm there was a problem there was a mouse problem on two point one which occurred intermittently

A: It's a bug fix

B: yes yes

A: Not a new functionality

B: I don't think so no There's also a new version of of meta software (Etches) available

A: yes I know

The pattern is similar in face-to-face meetings.

A: then this point that's the order
B: no [because some of the points are are implied
[you only give it two
A: [ah okay
B: [cos cos you know you're drawing a rectangle obviously you only give [the two=
[ahh right=
B: =corners so you don't give all four corners
A: =okay
A: so so it's because, yea, so something like a circle

Both face-to-face and LIVE-NET conversations have a quick-fire character, with clarifications taking place (LIVE-NET, Turns 3 and 5) and also disagreements (face-to-face, Turn 2), showing that participants are able to react quickly to incoming information when they do not understand or when they disagree.

4.6. Turn Distribution

Finally, we expected that the different conditions would lead to unequal distribution of turns between participants. We expected that, in ISDN, participants would rely on two people, one at either end of the link, to manage interactions across the connecting link, and they would channel their responses through these people. However, when we examined the data for dominance by two speakers, this was not the case (see Figure 10). We measured the number of turns produced by the two most frequent speakers in the three conditions.[5] There was no overall difference either in the percentage of turns taken by these people or in the number of words that they spoke: turns, $F(2, 11) = 0.59$, $p > .05$; words, $F(2, 11) = 0.55$, $p > .05$. We also investigated whether ISDN served to exclude certain speakers: The fact that they are less able to interrupt might prevent participants who are not "chairpeople" from having the opportunity to speak. Again this hypothesis was not borne out by our results. We looked at the number of words and the number of turns for the two people who spoke least. Again there were no differences: turns, $F(2, 11) = 0.75$, $p > .05$, words, $F(2, 11) = 1.32$, $p > .05$. This result is interesting because it runs contrary to the perceptions of the

[5] These scores were normalized to allow for the fact that there were different numbers of people in each meeting.

Figure 10. **Normalized mean percentage of turns (and mean standard error) and words spoken by most and least frequent speakers, showing levels of statistical difference for the three conditions.** p = ns throughout.

	IDSN	LIVE-NET	Face to Face
Turns taken by two most frequent speakers/total turns	66.84% (6.87)	58.48% (3.66)	73.45% (4.54)
Turns taken by two least frequent speakers/total turns	8.28% (2.93)	2.29% (0.84)	10.56% (6.80)
Words spoken by two most frequent speakers/total words	78.10% (5.65)	70.76% (4.89)	76.58% (3.64)
Words spoken by two least dominant speakers/total words	5.24% (2.42)	1.38% (0.56)	11.6% (6.46)

people using ISDN and LIVE-NET. They report feeling that certain participants are able to dominate the meeting and that others are less able to contribute to it.

4.7. Users Comments and Informal Observations

No objective measures were taken of the use and effectiveness of nonverbal behavior such as gaze and gesture in ISDN and LIVE-NET, although there do appear to be differences from face-to-face meetings. Our informal observations and comments made by the users show several apparent effects. First, mutual gaze is difficult to achieve in both video systems because participants look at the image of the remote participants and not directly into the camera. Furthermore, gaze behavior in both video conferences differs from face-to-face meetings in extremely obvious ways. Participants tend to stare fixedly at the screen displaying the remote participants even when the speaker is local, and they therefore show none of the normal modulation of gaze behavior and local speaker monitoring that is characteristic of face-to-face interaction (Duncan, 1972). Similar effects of "monitor capture" or "TV watching" are reported in Abel (1990). The result is that the speakers are presented with an array of remote people staring relentlessly almost directly at them. Speakers report finding this situation confrontational. It also means that speakers receive little local attentional feedback, because local listeners are staring only at the remote site.

Further problems relate to the control of the cameras. In both video-conferencing systems, camera control is local. This led to a number of interchanges resulting from attempted changes to camera angles, because the local participants were unable to tell whether the displayed image they were presenting was adequate for the remote participants. This was a particular problem with the document camera, when the issue concerned the resolution

of the presented document. Here participants had discussions along the lines of "can you see it yet?", "back a bit", "is that okay?", and "back a bit more." This type of fine tuning of the image was further hindered by the lags in ISDN, which meant that feedback about the acceptability of the image was not timely.

People stated that video conferences involved more effort. For example, participants complained about the difficulty of assuming control of the conversation in both video conferences. One participant reported for ISDN, "I have the feeling that I want to say something, but there's no opportunity to speak. Then when the opportunity does arise, I don't take it because my comment often isn't relevant anymore." In contrast, one LIVE-NET user acknowledged the greater formality of LIVE-NET meetings compared with face-to-face meetings but said that she sometimes exploited this to hold the channel for longer periods.

On the other hand, despite the problems with the video-conferencing systems, people preferred these to audio conferencing. The main stated advantages of video conferencing were knowing who was at the remote location and knowing who was speaking, although users' behavior suggested this information was not always available in ISDN due to poor image quality. Another stated advantage was the feeling of "not talking into a void." Our users also commented that they found video conferencing appropriate for only certain types of meeting such as information exchange or project updates.

There were other problems that were specific to ISDN alone. It was at times difficult to identify the speaker at the remote location in ISDN. The quality of the visual information was poor, and this seemed to reduce the impact of visual speaker cues such as leaning forward, increased gesturing, and posture changes. Informal observations suggest that speaker identification often took several seconds, although there was one ISDN meeting in which a participant spoke for several minutes and was still misidentified. Some groups in ISDN attempted to resolve this problem by panning the camera and focusing exclusively on the local speaker, which solved the identification problem for the remote participants. Unfortunately, this produced awkward transitions and further panning and focusing when someone at the same location spoke next. It also meant a narrowing of the visual field for the remote participants, with the result that they only had visual information about the current speaker and not the other remote participants.

People also complained that in ISDN small movements are not picked up and that sudden movements appear jerky and blurred. The movements were described as "puppet-like" by one user. Some participants also reported attempting to compensate for poor image quality by using exaggerated gestures such as nodding and the shaking of heads to substitute for their inability to provide verbal feedback.

5. CONCLUSIONS

Many reasons have been put forward for the failure of video conferencing to gain widespread acceptance, including cost, incorrect marketing, and the questionable value of a video channel. There have been few detailed empirical studies of the actual communication that occurs over real implementations of wide-area video-conferencing systems. By examining how the characteristics of two such systems affect the nature of spoken conversations, we aimed to identify possible reasons for the lack of success of video-conferencing technology. We also sought to explain why channel properties affected conversational processes in the way they did.

Our results showed that, compared with face-to-face meetings, spoken conversation patterns are disrupted over ISDN with its half-duplex line, transmission lags, and poor image quality.

- Listeners produced fewer backchannels and interrupted less often.
- Listeners were also less likely to anticipate turn endings.
- Speakers also alter their behavior, being more likely to hand over turns formally using a question or naming the next speaker. They were also less likely to hold the floor with redundant phrases.
- The result of listeners reducing interruptions and speaker feedback, combined with the general difficulty of switching speakers, was a formal lecture style of interaction, with long turns, handed over by a very deliberate process.

In LIVE-NET, even when there is a full-duplex line, immediate transmission, and a broadcast-quality image, the properties of the spoken communication still differ from face-to-face interaction.

- Although listeners interrupt as frequently as in face to face, they are less likely to give backchannels.
- Speakers use questions to formally hand over the floor more frequently, and they are also less likely to hold the conversation floor with redundant information.

Thus, although LIVE-NET was similar to face to face, it was still characterized by highly formal conversational behaviors.

How can we explain these findings? Our initial claim was that certain key channel properties of ISDN disrupt basic communication processes. Face-to-face interaction has full duplex, almost instantaneous transmission of audio as well as high-quality visual information. As expected, when we changed these

Figure 11. **Observed differences in characteristics and channel properties responsible. ISDN = Integrated/Services/Digital/Network, LN = LIVE-NET, FTF = face to face.**

Characteristic	Result	Channel Properties
Backchannel	ISDN < LN < FTF	Lag, half duplex, picture quality, and directionality
Interruption	ISDN < LN, FTF	Lag, half duplex, picture quality
Overlaps		
Project/complete	ISDN < LN, FTF	Lag, half duplex, picture quality
Floor hold	ISDN, LN < FTF	Directionality
Simultaneous starts	ISDN > LN, FTF	Lag, half duplex, picture quality
Handovers		
Questions	ISDN, LN > FTF	Directionality
Tags	No differences	None
Naming	ISDN < FTF	Directionality
Turns		
Number	ISDN < LN, FTF	Lag, half duplex, picture quality
Size	ISDN > LN, FTF	Lag, half duplex, picture quality
Turn distribution	No differences	None

channel properties to those of the ISDN system, we produced a lecture-like style of interaction that lacked spontaneity. However, the argument that these channel properties are solely responsible for communication disruption must be re-examined in the light of the LIVE-NET data, because for several conversation characteristics, LIVE-NET was more like ISDN than face to face. This suggests that other channel properties are also critical here and that the account should be extended to include these properties and to determine which conversation characteristics they impact.

Figure 11 addresses this question. It depicts where differences occurred between ISDN, LIVE-NET, and face to face and which channel properties might be responsible. Our predictions about differences were only met for certain characteristics (viz., interruptions, projections, simultaneous starts, turn size, and turn frequency). Here we observed differences between ISDN and both LIVE-NET and face to face, with LIVE-NET and face to face being equivalent. The implication is therefore that differences in these characteristics are attributable to lags, half-duplex audio, and poor video quality. In contrast, other characteristics such as floor holding and handovers showed equivalence between LIVE-NET and ISDN.

How can we explain these similarities between ISDN and LIVE-NET and what channel properties are responsible? ISDN and LIVE-NET are similar because they both have nondirectional sound and vision from a restricted number of sources (i.e., one or two monitors and loudspeakers). Both systems contrast with face to face, where sound and visual behavior are directional, because they emanate from the different participants.

Other research has shown that head turning and eye gaze play an important role in speaker switching (Duncan, 1972) and that both these behaviors are reliant on directionality. Its absence in ISDN and LIVE-NET may lead to changes in speaker behavior, with speakers having to use the verbal channel to signal turn transitions explicitly and carefully manage speaker switches. This may explain the increased incidence of questions in ISDN and LIVE-NET and the reductions in floor holding. Sellen (1992) directly addressed the impact of directionality on conversations in video conferences. She found few objective effects for directionality, but this may have been due to the small image size employed, so more work should be done to test this.

Regarding other characteristics, ISDN was different from both LIVE-NET and face to face, as we expected. It therefore seems that changes in interruptions, projections, simultaneous starts, turn size, and length resulted from lagged half-duplex audio and poor picture quality. These channel properties seemed to lead to changes in listener behaviors. Being conscious of the disruptive effects of lag and half-duplex audio, listeners wait for the previous speaker to finish before taking the floor. The effect of reduced listener participation is to decrease the number of speaker switches and, hence, produce fewer but longer turns.

Backchannels seem to be affected by all the channel properties, with face to face being different from LIVE-NET, which in turn differs from ISDN. It may be that listeners are aware of the disruptive effects of backchannels with half-duplex, lagged audio, and this may account for differences between ISDN and LIVE-NET. The difference between LIVE-NET and face to face may arise because speakers rely on directional gaze in face to face to elicit backchannels, and this cue is removed in both LIVE-NET and ISDN.

Finally, there were conversational characteristics that seemed to be unaffected by channel properties, such as tagging and turn distribution. It may be that these are reflexive conversational behaviors produced independently of the communication situation.

We could not, however, isolate whether audio lag, half-duplex audio, or video quality was mainly responsible for the disruptions in ISDN. This was because we attempted to gather data for real systems for which these properties were not independent. Other laboratory work should be done to confirm which of the channel properties of the ISDN system was most disruptive of these conversation characteristics. Currently, we cannot rule out any of these channel properties, and other research has independently shown

that poor-quality visual information, lags, and half-duplex audio can all independently produce these types of effects (K. Cohen, 1982; Krauss & Bricker, 1967; D. Rutter & Stephenson, 1977).

A study of low-lag, full-duplex video conferencing by Sellen (1992) is similar to our LIVE-NET and face-to-face comparisons. Sellen stressed the effects of video mediation on listener behavior and concluded: "It is as if conversants in video-mediated conversations were more opportunistic or polite, waiting for a pause or for a speaker to finish before attempting to take the floor" (p. 57). We found some changes in listener behavior in LIVE-NET: The number of backchannels was reduced, although there were equal numbers of interruptions in face to face and LIVE-NET. We found in our study, however, that the main differences between LIVE-NET and face to face seemed to be attributable to changes in speaker behavior, with greater use of handovers and reduced floor holding.

The current work shows that certain basic communication processes are disrupted by the channel properties of the two mediated communication systems. Due to the fact that we observed real meetings with naturalistic data, we were unable to measure directly the overall effectiveness of communication in the different conditions. Other research has shown that characteristics such as backchannels and interruptions are related to task outcome (Kraut et al., 1982; Oviatt & P. Cohen, 1991). Laboratory studies are needed to measure the effects on task outcome of disrupting the different conversation characteristics under more controlled conditions.

Although we cannot judge overall quality, there may be implications about the kinds of tasks for which the current ISDN quality is appropriate. The lecture-like character and the inability to support quick-fire exchanges could mean that ISDN is unsuitable for tasks such as conflict resolution, planning, or negotiation, where the ambiguity of the information and the requirement for rapid clarification and feedback are critical for the success of the interaction (Daft & Lengel, 1984; Whittaker, 1992).

If ISDN cannot effectively support these tasks, this may contribute to the lack of success of this quality of video conferencing. It may be that future remote collaborators have to choose appropriate communication technologies for the task at hand and ensure that certain types of tasks (e.g., conflict resolution and negotiation) are resolved in face-to-face situations. Naturalistic studies or remote collaborators who are using multiple technologies should be conducted to determine how people currently allocate technologies to communication tasks, and more theoretical work is needed to specify the relationship between communication task requirements and the basic communication processes that are needed to support them.

What are the practical implications of these results? First, it seems that introducing low-lag, full-duplex channels leads to improvements in communication, as evidenced by the superiority of LIVE-NET over ISDN. This

suggests that we should continue to work on high-speed wide-area networks and compression technology to reduce the disruptions to communication described earlier. However, the LIVE-NET results also suggest that improving these properties alone will not exactly reproduce face-to-face interaction.

How might we improve video systems, in addition to improving networks and compression? One possibility is the implementation of directional audio and video, which might address the outstanding differences between LIVE-NET and face to face (Sellen, 1992). A second strategy would be to modify existing video-conferencing systems by acting on our users' comments. They suggested providing remote audio and video controls, so that remote participants are able to chose what they want to see and hear rather than have these choices made for them. They also suggested the use of several monitors. One monitor could be used to provide a high-quality image of the speaker or object of interest, and other monitors could then present lower quality panoramic images of the remaining remote participants for visual context.

Another approach to improving video systems is to examine the use of the video image for things other than pictures of participants' head and shoulders. Elsewhere, we suggest an alternative novel application of video in the notion of "video as data" for remote microsurgery (Nardi et al., 1993). We claim that there may be many situations in which video is best suited to transmit images of the work itself, rather than of the participants who are carrying out the work. Another application might be the "Open Distributed Office" in which video is used to give people in distributed teams monitoring information about whether remote collaborators are present or absent (Dourish & Bly, 1992; Fish et al., 1992; Mantei et al., 1991). This contrasts with other applications of video, because it stresses the benefits of video for background awareness instead of solely for direct communication.

Another short-term way to improve communication is to allow different trade-offs between audio and video in the limited bandwidth. We might improve communication by relaxing the requirement for synchronized audio and video and by reducing the bandwidth allocated to video. Studies have reported the importance of audio as compared with visual information in this type of application (Chapanis, 1975; Reid, 1977; Williams, 1977). Consequently, removing the requirement for synchronization would lead to the reduction of audio lag, because there is less audio than video data to be compressed. User studies are required to determine just what quality of audio and video is needed for this type of solution, most specifically to identify the audio requirements and the effects of lack of synchronization (Shah, Staddon, Rubin, & Ratkovic, 1992).

Finally, this work contributes to a developing theory of mediated communication. Other work has shown that the organization of mediated communication is critically dependent on the properties of the communication

channels (Whittaker, 1992; Whittaker, Brennan, & Clark, 1991). Previous research has explained this type of result in terms of concepts that are difficult to operationalize, such as "social presence" (Short et al., 1976), "media richness" (Daft & Lengel, 1984), or "cuelessness" (Rutter & Robinson, 1981). Here we were able to test predictions derived from an analysis of face-to-face interaction, about how certain channel properties influenced specific characteristics of speaker and listener behavior. Although our results were not entirely consistent with our initial hypotheses, this helped us both to identify a further potentially important channel property and to refine our understanding of the relationship between channel properties and communication characteristics. The new channel property is that of directionality, which we argue mainly impacts speaker behaviors. In its absence, speakers tend to show increased formality and explicitness in managing turn switching. A combination of our initial channel properties of lags, half-duplex audio, and poor-quality video contributed to what seemed to be reductions in the spontaneity of listener behavior, with lower levels of listener participation and more lecture-like speech as a result. Further work should test these more specific hypotheses about the relationships between these channel properties and communication characteristics.

Text

Acknowledgments. Thanks to Erik Geelhoed for assistance in statistical analysis; to Hewlett Packard Laboratories Bristol for providing facilities and a grant to support this work; to Richard Beckwith of University College, London, for providing a grant; and to the people who participated in the meetings and kindly allowed themselves to be recorded and interviewed. Thanks also to Lyn Walker, Herb Clark, and David Frohlich for discussions of these ideas.

REFERENCES

Abel, M. (1990). Experiences in an exploratory distributed organization. In J. Galegher, R. Kraut, & C. Egido (Eds.), *Intellectual teamwork* (pp. 489-510). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Argyle, M., Lalljee, M., & Cook, M. (1968). The effects of visibility on interaction in the dyad. *Human Relations, 28,* 289-304.

Atkinson, J. M., & Heritage, J. C. (1984). *Structures of social interactions: Studies in conversation analysis.* Cambridge, UK: Cambridge University Press.

Chapanis, A. (1975, March). Interactive human communication. *Scientific American, 232,* 36-42.

Clark, H., & Schaefer, E. (1989). Contributing to discourse. *Cognitive Science, 13,* 259-294.

Clark, H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22,* 1-39.

Cohen, J. (1960). A co-efficient of agreement for nominal scales. *Educational and Psychological Measurement, 20,* 37-46.

Cohen, K. M. (1982). Speaker interaction: Video teleconferences versus face-to-face meetings. *Proceedings of Teleconferencing and Electronic Communications,* 189–199. Madison: University of Wisconsin.

Daft, R., & Lengel, R. (1984). Information richness: A new approach to managerial behavior and organizational design. In B. Straw & L. Cummings (Eds.), *Research in organizational behavior* (pp. 191–223). Greenwich, CT: JAI Press.

Dourish, P., & Bly, S. (1992). Portholes: Supporting awareness in distributed group work. *Proceedings of the Conference on Computer Human Interaction,* 541–547. New York: ACM.

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). Videoconferencing as a technology to support group work: A review of its failures. *Proceedings of the Conference on Computer Supported Co-Operative Work,* 13–24. New York: ACM.

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* (pp. 291–326). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Fish, R., Kraut, R., & Chalfonte, B. L. (1990). The VideoWindow system in informal communication. *Proceedings of the Conference on Computer Supported Co-Operative Work,* 1–12. New York: ACM.

Fish, R., Kraut, R., Root, R., & Rice, R. (1992). Evaluating video as a technology for informal communication. *Proceedings of the Conference on Human Computer Interaction,* 37–48. New York: ACM.

Galegher, J., & Kraut, R. (1990). Computer-mediated communication for intellectual teamwork: A field experiment in group writing. *Proceedings of the Conference on Computer Supported Co-Operative Work,* 65–78. New York: ACM.

Harkness, R. (1986). Videoconferencing. In T. C. Bartee (Ed.), *Digital communications* (pp. 337–392). Indianapolis, IN: Howard W. Sams.

Jaffe, J., & Feldstein, S. (1970). *Rhythms of dialogue.* New York: Academic.

Jefferson, G. (1984). On stepwide transition from talk about a trouble to inappropriately next-positioned matters. In J. Atkinson & J. Heritage (Eds.), *Structures of social action* (pp. 191–222). Cambridge, UK: Cambridge University Press.

Johansen, R. (1984). *Teleconferencing and beyond: Communications in the office of the future.* New York: McGraw-Hill.

Johansen, R., & Bullen, C. (1984). Thinking ahead: What to expect from teleconferencing. *Harvard Business Review, 62,* 164–174.

Kirk, R. (1982). *Experimental design: Procedures for the experimental sciences.* Belmont, CA: Brooks/Cole.

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.

Krauss, R., & Fussell, S. (1990). Mutual knowledge and communication effectiveness. In J. Galegher, R. Kraut, & C. Egido (Eds.), *Intellectual teamwork* (pp. 111–146). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Kraut, R. (1989). Telecommuting: The trade-offs of home work. *Journal of Communication, 39,* 19–47.

Kraut, R., Egido, C., & Galegher, J. (1990). Patterns of contact and communication

in scientific research collaboration. In J. Galegher, R. Kraut, & C. Egido (Eds.), *Intellectual teamwork* (pp. 149-173). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Kraut, R., Lewis, S., & Swezey, L. (1982). Listener responsiveness and the co-ordination of conversation. *Journal of Personality and Social Psychology, 43,* 718-731.

Levinson, S. (1983). *Pragmatics.* Cambridge, UK: Cambridge University Press.

Mantei, M., Baecker, R., Sellen, A., Buxton, W., Milligan, T., & Wellman, B. (1991). Experiences in the use of a media space. *Proceedings of the Conference on Computer Human Interaction,* 203-208. New York: ACM.

McGrath, J. (1984). *Groups: Interaction and performance.* Englewood Cliffs, NJ: Prentice-Hall.

McGrath, J. (1990). Time matters in groups. In J. Galegher, R. Kraut, & C. Egido (Eds.), *Intellectual teamwork* (pp. 23-62). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Morley, I., & Stephenson, G. (1969). Interpersonal and interparty exchange: A laboratory simulation of an industrial negotiation at plant level. *British Journal of Psychology, 60,* 543-545.

Morley, I., & Stephenson, G. (1970). Formality in experimental negotiations: A validation study. *British Journal of Psychology, 61,* 383-384.

Morley, I., & Stephenson, G. (1977). *The social psychology of bargaining.* London: Allen & Unwin.

Nardi, B., Schwarz, H., Kuchinsky, A., Leichner, R., Whittaker, S., & Sclabassi, R. (1993). Turning away from talking heads: An analysis of video-as-data. *Proceedings of the Conference on Human Computer Interaction,* 337-334. New York: ACM.

Noll, M. (1992, May/June). Anatomy of a failure: Picturephone revisited. *Telecommunications Policy,* 307-316.

Oviatt, S., & Cohen, P. (1991). Discourse structure and performance efficiency in interactive and non-interactive speech modalities. *Computer Speech and Language, 5,* 297-326.

Panko, R. (1992). Managerial communication patterns. *Journal of Organisational Computing, 2,* 95-122.

Reid, A. (1977). Comparing the telephone with face-to-face interaction. In I. Pool (Ed.), *The social impact of the telephone* (pp. 386-414). Cambridge, MA: MIT Press.

Rutter, D. R., & Stephenson, G. (1977). The role of visual information in synchronizing conversation. *European Journal of Social Psychology, 2,* 29-37.

Rutter, D. R., & Robinson, R. (1981). An experimental analysis of teaching by telephone. In G. Stephenson & J. Davies (Eds.), *Progress in applied social psychology* (pp. 143-178). London: Wiley.

Sacks, H., Schegloff, E., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking in conversation. *Language, 50,* 696-753.

Sellen, A. (1992). Speech patterns in video-mediated conversations. *Proceedings of the Conference on Computer Human interaction,* 49-59. New York: ACM.

Shah, A., Staddon, D., Rubin, I., & Ratkovic, A. (1992). Multimedia over FDDI. *Proceeding of the Conference on Local Computer Networks,* 110-120. New York: IEEE.

Short, J., Williams, E., & Christie, B. (1976). *The social psychology of telecommunications.* London: Wiley.

Sproull, L., & Kiesler, S. (1991). *Connections.* Cambridge, MA: MIT Press.

Whittaker, S. (1992). *Towards a theory of mediated communication.* Unpublished manuscript.

Whittaker, S., Brennan, S., & Clark, H. (1991). Co-ordinating activity: An analysis of computer supported co-operative work. *Proceedings of the Conference on Computer Human Interaction,* 361-367. New York: ACM.

Whittaker, S., & Stenton, P. (1988). Cues and control in expert-client dialogues. *Proceedings of the Conference of the Association of Computational Linguistics,* 123-130. Cambridge, MA: MIT Press.

Wichman, H. (1970). Effects of isolation and communication on cooperation in a two-person game. *Journal of Personality and Social Psychology, 16,* 114-120.

Williams, E. (1977). Experimental comparisons of face-to-face and mediated communication. *Psychological Bulletin, 84,* 963-976.

Yngve, V. (1970). On getting a word in edgewise. *Proceedings of the Sixth Meeting of the Chicago Linguistic Society* (pp. 567-577). Chicago: Chicago Linguistic Society.

HCI Editorial Record. First manuscript received October 26, 1992. Revision received June 5, 1993. Accepted by Judith Olson. Final manuscript received August 26, 1993. — *Editor*
