An Analysis of the Spoken Aspects of Video Mediated Communication

Brid O'Conaill, Steve Whittaker
Interactive Technology Department
HP Laboratories Bristol
HPL-92-163 (R.1)
August, 1995

We present a method for comparing the spoken aspects of communication in two different types of video mediated meetings with spoken communication in face-to-face meetings. Based on an analysis of the media characteristics of face-to-face communication, we derive a set of predictions about how the spoken characteristics of communication will differ in these two systems, as compared with face-to-face communication. As predicted we found that adding video can *detract* from spoken conversation when networking bandwidths are limited. Contrary to our expectations, however, we found that even high quality video does not replicate all properties of face-to-face communication and we attempt to explain why.
ABSTRACT

We present a method for comparing the spoken aspects of communication in two different types of video mediated meetings with spoken communication in face-to-face meetings. Based on an analysis of the media characteristics of face-to-face communication, we derive a set of predictions about how the spoken characteristics of communication will differ in these two systems, as compared with face-to-face communication. As predicted, we found that adding video can *detract* from spoken conversation when networking bandwidths are limited. Contrary to our expectations, however, we found that even high quality video does not replicate all properties of face-to-face communication and we attempt to explain why. FOOTNOTE (This study first appeared as an article in Human Computer Interaction, and was also reported in INTERCHI93 Conference on Computer Human Interaction. Here, we concentrate on the method, its use and limitations. For a more complete report of the results please refer to the original article.)

1 INTRODUCTION

Many video communication technologies are premised on the hypothesis that by adding a visual channel to audio only technologies, such as the telephone, will improve communication, and simulate many of the important properties of face-to-face communication. Thus, the added benefits of having the visual information afforded by gaze, gesture and facial expressions will enrich the quality of the communication (ref. role of vision chapter). For organisations the promise of increased remote collaboration through improved mediated communication is very attractive. It means they can increase the number of potential co-workers while decreasing travel costs: fewer employees need to journey to meet their distant colleagues because they can interact effectively with their colleagues using videoconference or videophone (Johansen 1994). It is clear, however, that these promises have not been fulfilled (Egido, 1988, 1990, Noll,). Why then have video technologies failed to be the commercial success that was forecast as far back as their first introduction in 1927? Two answers have been put forward: inadequate analysis of user needs, (Johansen, 1984; Johansen & Bullion, 1984; 1990).
1984; Panko, 1992), and a failure to identify a strong task outcome benefit from the addition of a video channel (Chapanis, 1975; Reid, 1977; Short, Williams & Christie, 1976; Williams, 1977). Here, we explore another possibility - that adding video can compromise basic communication processes. A number of key processes in face to face communication are heavily dependent on two-way interaction and involve very precise timing. Our hypothesis is that adding a video channel to audio can disrupt both concurrency and timing and that current commercial systems may not support these key processes.

Unlike many laboratory prototype video systems, commercial systems are subject to bandwidth constraints in transmitting video data. Video data has to be compressed to travel over these commercial networks and this introduces delays in communication. The time taken to compress, transmit and decompress video data is greater than that for transmitting audio only. To provide synchronised video and audio, it is common practice to buffer the audio until the video is image is processed, and this introduces delays. Additionally, in the case of conferencing systems, there is a need to prevent feedback from the audio stream being retransmitted from the destination side. To alleviate this problem many systems are half-duplex, that is, the audio is transmitted from only one location at a given time. These channel properties are very different from the properties of face to face communication where the audio channel is both immediate and full duplex. In this chapter, we predict how these channel constraints in video mediated systems, make the spoken aspects of communication in such systems different from face to face communication.

Previous theoretical work has addressed the question of how mediated communication differs from face-to-face interaction. Prior attempts to characterise the differences have relied on notions such as "social presence" (Short et al., 1976), "cuelessness" (Rutter & Robinson, 1981), and "media richness" (Daft & Lengel, 1986). However, such terms are difficult to measure objectively and are open to interpretation. We wished to quantify differences in terms of *measurable* characteristics of speech processes that have independently been shown to be important in face to face interaction (see also Sellen, 1992, 1994, this volume). In understanding how these speech processes are affected in
video mediated communication, it should be possible to explain why video technologies have not succeeded and to suggest more limited applications for which it might be appropriate.

We compared interaction in two wide-area video conferencing systems with face-to-face conversation. The first system (ISDN) is representative of currently available commercial video conferencing systems. It has half duplex audio, transmission lags and poor picture quality. The second, (LIVENET), is a higher quality prototype system that is representative of future systems. It uses an analogue infrastructure with the properties of duplex audio, no transmission lags and full bandwidth video. After describing these two systems in detail we present some conversational features which are important to communication. For each of these features we derive predictions about they will differ across the three conditions as a result of the different channel properties of the systems.

1.1 The Systems

ISDN System (ISDN)

The system is located at Hewlett Packard Laboratories Bristol and the majority of the conferences held in Bristol are to the USA. Conferencing takes place over two ISDN lines each at 64kb/s. Rate adaptation must take place because US installations use a public switched 56kb/s digital network. Thus the available bandwidth is reduced from 128kb/s to 112kb/s. Of this 16kb/s is used for audio with an additional amount for communication between CODECs. The amount of bandwidth available for video transmission is approximately 90kb/s.

The video signal is compressed by removing both spatially and temporally redundant data using a CLI Rembrandt CODEC. This process takes about 120ms with an equivalent time required for decompression at the other site. The audio signal is also compressed but has to be buffered to synchronise 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 170ms in each direction, can be expected to the West Coast of the US, 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 135ms and an equivalent time is taken to transmit the signal to the next earth station. To connect to the West Coast of the US two satellite jumps are required which means a delay of 540ms 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 410ms to 780ms, depending on the propagation route.

The audio channel is half duplex and 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 disruptions both to the audio channel and the video picture.

The conference room is a converted meeting room located next to the other meeting rooms. It contains a table at which three people can sit comfortably. Sitting at the table a user 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 nine feet and contains a 26 inch colour monitor above which two cameras are located. The monitor displays the live picture of the remote location. In addition to the distance of the local participants from the monitor, there is the distance of the remote participants from their camera. This increases the perceived distance between participants. This 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. If the user is looking at the monitor it appears at the remote location almost as though they are
looking into camera. However, the distance and the video quality make eye-gaze and head movements unclear.

Figure 1 about here

A small desktop control panel enables users to switch between cameras and to focus, pan and zoom. Participants control their local cameras, choosing the view which 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 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 inch colour monitor, (“the confidence monitor”), which displays the live picture which 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 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.

**LIVE-NET (LN)**

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 fibres each carrying four full bandwidth video channels, with sound on a 6 MHz subcarrier, and a fifth lower bandwidth video channel used for data up to 2 Mb/s 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. Firstly, there has been some acoustic treatment in the rooms to prevent loudspeaker sound being reflected back into the microphones. Secondly, a Shure AMS automatic microphone system is used which has unidirectional microphones, which do not pick up sound from the rear, and a very fast switching system ensures that only one or two microphones in the group are active at any time. Thirdly, a frequency shifter (5Hz) 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.

Figure 2 about here

The participants sit at a table and face a set of four 20 inch monitors and a CCD 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 “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 again 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 3 about here

**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 six feet long and four feet 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 Deriving predictions on how speech processes are affected in video-mediated communication from**
**conversational theory**

From the above descriptions we can see that the channel properties of the two video conferencing systems differ from the face to face condition. The ISDN system introduces a transmission lag of between 410 and 780ms for both audio and video, and a half duplex (one-way) line for audio. In addition, both the LN and ISDN systems allow only limited visual cues, reducing the multimodal quality of the conversation. Both have relatively fixed camera angles and in ISDN the picture quality is poor and subject to jitter and occasional frame loss. How might we expect these differences to impact the nature of communication over the two systems? Figure 4 summarizes our expected findings for a number of spoken conversational characteristics. For each, we will characterise face to face communication before deriving our prediction on how the individual speech characteristics will be affected by the two systems.

INSERT TABLE HERE

### 2.1 Backchannels

Communication is a joint activity which requires co-ordination of both process and content (Clark & Brennan, 1991, Clark & Wilkes-Gibbs, 1986; Whittaker, 1992). A critical aspect of this is that speakers and listeners attempt to co-ordinate and maintain shared knowledge and beliefs (Clark & Brennan, 1991, Clark & Schaefer, 1989). To allow this co-ordination 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 short feedback utterances and visual feedback in the form of head nods and eye-gaze. This positive concurrent feedback is called backchanneling and serves several functions including attention, support or acceptance of the speaker's message. For the purpose of this study only auditory backchannels were measured and not head nods or gaze behaviour. 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 inform the speaker that
s/he can rely on and build upon the listener's understanding (Clark & Schaefer, 1989; Duncan, 1972; Whittaker & Stenton, 1988; Yngve, 1970). Analysis of FTF data shows they are often delivered with split second timing, for example:

A: ...in the absence of a 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...

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
B: [mm

In face-to-face interaction backchannels are therefore 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. This 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, More significantly it may disrupt the speaker at the remote location by its late arrival.

2. Transcription conventions are described in Section 3.3
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. In contrast in the LN system the audio channel is full duplex and transmission nearly instantaneous. This allows both concurrent and timely backchannels to be delivered. In LN, therefore, we expect backchannels to occur as frequently as in face-to-face interaction.

2.2 Interruptions
Interruptions are often associated with simultaneous speech. However, simultaneous speech can arise for a number of different reasons (Levinson, 1983). We distinguish between two different classes of simultaneous speech: (a) Overlaps: those instances where there was a clearly identifiable reason why the next speaker should have broken into the current speaker's utterance for example, after a transmission failure leads to a line break; (b) Interruptions: Those instances of where there was no such reason for the next speaker to have broken into the current utterance. We begin by discussing Interruptions.

We defined interruptions as those instances of simultaneous speech where there is no indication by the first speaker that they are about to relinquish the conversational floor. As such they are deliberate attempts to gain the conversational floor without the prior consent of the current speaker. Here is a FTF example of an interruption:

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 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 which then has to be repeated, and turntaking re-established. In contrast, we expect that in LN 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. We therefore predict an equivalent number of interruptions in LN as in FTF.

2.3 Overlaps

Overlaps are instances of simultaneous speech which follow a signal the speaker gives indicating that they intend to relinquish the conversational floor (Levinson, 1983). We made predictions for three different types of overlaps: Projection/completions, floorholding, and simultaneous starts.

(i) 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, Clark & Schaefer, 1989). In predicting the possible finish by the current speaker, the next speaker may recognise that the message of the current speaker is semantically complete, although they may continue to speak. The next speaker may then begin to speak over the redundant part of the original speaker's message.

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 their turn.

A: ahm how the work [ho how how it works=
B: (pans out
A: =with [how the work pans out between the peo-
B: (pans out

We expected fewer projections overall 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 their turn. In LN we expected equivalent numbers of projections as in face-to-face interaction, because low-lag full duplex audio should support this type of intervention.

(ii) Floorholding: This occurs when the next speaker tries to take the floor while the current speaker attempts to hold the floor while producing utterances which do not contain any information (Jefferson, 1984). Examples of floorholding can range from self-repetitions to function word repetition ("so ... so"): 

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 
B: [some but [somebody has got to own the interface the top level interface= 

We expected less floorholding in ISDN, because of adjustments by both speakers and listeners. 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. Listeners should avoid trying to seize the floor for the same reason. There should be no such constraints with LN, where floorholding should be possible without such disruptive effects.

(iii) Simultaneous Starts: 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 may have already begun to speak. This situation conspires to produce simultaneous starts in ISDN. The situation is different in LN where low lag times and full duplex should allow equivalent numbers of simultaneous starts as in face-to-face interaction.

Overall we expect LN to result in the same number of overlapping speech acts given the similarity of the audio properties to FTF. For ISDN, we have different predictions for each of the subtypes of overlapping speech.

2.4 Explicit Handovers
Turn-taking is central to the process of conversation. Speaker switches must be smooth and not disruptive to the overall flow of the conversation to achieve a high level of interactivity (Kendon, 1972, Sacks, Schegloff, & Jefferson, 1974). How is this process achieved? There are a number of intonational, syntactic, pragmatic and non-verbal devices that speakers use to indicate that they are about to finish their conversational turn. Listeners also use non-verbal devices to indicate that they wish to speak e.g. leaning forward or achieving mutual gaze (Argyle & Cook, 19??, Kendon, 1972). 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 and 770ms (Jaffe & Feldstein, 1970). In some cases, there is no pause, and overlaps occur between speaker transitions. We identified three verbal devices used by speakers to indicate that they intend to relinquish the floor: (i) the use of questions; (ii) 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 (iii) naming the next speaker (Levinson, 1983, Sacks et al., 1974).

We expected that given the timing and concurrency problems in ISDN, speakers would be much more likely to explicitly signal that they had finished their turn. The poor quality of visual information may also have contributed to more formal handovers because eye-gaze and gestural cues are not reliable. In ISDN, we therefore expected more instances of questions, tagging and naming of the next speaker. This should not be true in LN where speaker switching should be unproblematic, and explicit handovers unnecessary.
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 conspire to increase turn length in ISDN. Where feedback, such as backchanneling is absent or even delayed, the speaker's ability to formulate efficient messages is reduced (Krauss & Bricker, 1967; Krauss & Fussell, 1990, Kraut, Lewis & Swezey, 1982?). Without feedback, speakers are unable to assume the message has been understood: they may therefore attempt to clarify or reiterate points, sometimes unnecessarily, to ensure that the listener has not misunderstood (Kraut et al., 1982?). Absence or delay of feedback can therefore encourage the speaker to take long turns (Krauss & Bricker, 1967; Oviatt & Cohen, 1991). In addition, the difficulty in interrupting the speaker in ISDN would reduce the number of "quickfire" interchanges. 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 LN 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.

2.6 Turn distribution

Finally we expected that turns would be unequally distributed in the different technologies. In face-to-face interaction, all participants in principle, have equal access to the conversational floor, although there are external factors such as knowledge and power which influence participation levels (McGrath, 1984, 1990). In each videoconference, 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 channelling their responses to the remote location through one specific individual at each location. We therefore expected that these local co-ordinators 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 LN meetings.

To summarise, the above analysis and predictions, it can be seen that certain critical features of face-to-face conversation depend on three properties of the communication channels: (1) low transmission lags, i.e. messages are received almost instantaneously by listeners; (2) two way transmission, e.g. feedback can be produced at the same time as the speaker's utterances, (3) multiple modalities, i.e. both verbal and visual channels are used in synergy (Whittaker, 1992). We have predicted how the channel limitations of the two systems will impact some conversational features.

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 on to video tape. The face-to-face, (FTF), meetings were audio taped. An observer was present at each meeting who noted any events not picked up on tape.

3.2 The meetings

Five ISDN video-conferences, four LIVE-NET meetings and five face-to-face meetings were recorded and analysed. 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 DACOM classification of business meetings (Short et al., 1976), are shown in Figure 5. This shows that all the meetings were co-operative in nature with their main function being to exchange information. Secondary functions and activities such as problem solving and idea generation also took place.

Table 5 about here

The FTF 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 centred around project teams with one or two project managers being present. The LN 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 the majority of 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, e.g. participants at either end of a video-conference link are likely to know each other better and have a greater understanding of local work. Where 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 U.S. who were visiting the U.K., and therefore had little day to day contact. All participants were familiar with using videoconferences. 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 analyse whether conversational behaviours changed in the course of each meeting.
In the FTF and ISDN conditions there was a mixture of agenda and non-agenda based meetings. All the LN meetings were agenda based. Both the FTF and ISDN meetings had an average of six participants. For FTF and ISDN, the smallest meeting had four participants, and the largest had seven. LN 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 LN meetings took place using four sites. Typically all meetings lasted between one and two hours.

3.3 Analysing the data

A twenty minute segment from the middle of each meeting was transcribed in detail. Each segment was taken twenty minutes 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 analysed. 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 definitions given above.

The data was transcribed using a simplified version of the system developed by Jefferson (Drew 1991; Jefferson 1987) for conversational analysis. Our aim was to capture those features of conversation associated with speaker transition, for example, the positioning of overlapping speech. We did not code the phonetic information e.g. prosodic turn switching cues. Sentences were transcribed as they were spoken including any syntactical errors. The extract below shows some of the conventions in use and is followed by a glossary of the symbols.

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

---

3. The participants of course chose whether or not the meeting was agenda based.

4. 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
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. Both tried to identify every instance of backchannels, interruptions, overlaps, and formal handovers. Reliability scores, 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 LN (0.88), but slightly less reliable for face-to-face interaction (0.79).

4 RESULTS
4.1 Overview
The following section summarises the findings from the study. FOOTNOTE(For a more complete analysis of the results please refer to O'Conaill, Whittaker and Wilbur, (1993)). For each prediction we present our findings and illustrate our claims with representative examples from the interactions. Where differences are discussed, these differences are statistically significant. Unless otherwise stated the differences were analysed in a one way ANOVA, and posthoc ANOVA tests were subsequently administered to make pairwise comparisons between the conditions following the recommendations of Kirk (1982). All analyses apply to the 20 minute segment we analysed for each meeting and not to the whole meeting.

4.2 Backchannels and Interruptions
Figure 6 about here
Figure 6 shows the distribution of backchannels and interruptions in the three conditions. Mean levels of backchanelling were low in ISDN compared with FTF (7.00 vs. 60.80), confirming our prediction that people in ISDN would avoid backchannels. The following example shows why backchannels were reduced in ISDN: Where backchannels do occur, their arrival at the remote location is delayed which 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 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

We did not predict however, that backchannels would also be reduced in LN compared with FTF (30.50 vs. 60.80). We suggest reasons for this in the Discussion section.

Again as predicted, interruptions were also significantly less frequent in ISDN. In the face-to-face meetings, almost 10% of turns were interruptions compared with less than 2% in ISDN. This occurred despite technical reasons for interruptions in ISDN, producing line breaks with consequent loss of audio and video for several hundred milliseconds. Many of the interruptions in ISDN followed line breaks and represented requests for a repetition of information lost during a break of the channel. There were fewer of these problems in LN and none in face-to-face, where the majority of interruptions are to clarify what the speaker has said and not requests for repetition.

A second analysis removed interruptions following line breaks, showed that there were large differences between the media, with both face-to-face and LN having more interruptions, than ISDN. All other results were equivalent for this interruptions 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 for a repetition of "lost" information. This example shows the added 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 data base 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 LN 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.

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 analysed 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. However, the different types of overlaps showed different distributions in the three conditions.

Figure 7 about here

For projections we found, as we predicted, there were differences between the conditions with more overlaps following projections in the face-to-face and LN media (7.3% FTF and 9.2% LN vs. 2.9% ISDN). The combination of half duplex and lags seem to combine to reduce projections in ISDN, with listeners avoiding overlapping speech even when this could assist the speaker in composing their message. Projections were reduced in ISDN relative to FTF and also LN, as predicted.

We found that floor-holding was much more frequent in face-to-face than in both LN and ISDN (1.8% vs. 0.6% and 0.0%). The difference between ISDN and FTF is explicable as a combination of speaker and listener adaptations to lag and half-duplex: Listeners try not to break in on current speakers, and speakers immediately stop talking when they have got nothing further to say. The extent of this adaptation was that strikingly, there were no examples of floorholding in the ISDN meetings. Contrary to our expectations, however, was the finding that floorholding was reduced in LN compared with face-to-face. We provide possible explanations for this in the Discussion 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 LN and face-to-face (6.7% vs. 2.5% and 1.0%).
Examples serve to illustrate these effects for the different types of overlap. In the face-to-face and LN conditions, overlaps occurred mainly when the listener completed the speaker's utterance or misprojected a turn end as in the following LN 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. On the completion of A's turn, C assumes he can take the channel. However because of the transmission lag he is unaware that B has continued. Both B and C 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 not possible over a half duplex 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. Where this does not occur a second or third clash can happen. 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 dif-
fer- ence ( )
C: go ahead

To summarise for overlapping speech, there are no differences in the combined number of overlaps but the subtypes of overlaps are differently distributed in the three conditions. Overlaps occur in face-to-face mainly because of projections and floorholding, in LN because of projections, and in ISDN because of simultaneous starts. This is consistent with our predictions, with the exception of the finding of reduced floorholding in LN.

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 FTF, (30.8% vs 8.8%), because of the need to explicitly manage speaker transitions. Contrary to our expectations, however, we found the same overall pattern of formal handovers in LN (21.2%) as in ISDN. Again we suggest possible explanations in the Discussion.
Further analysis of the different classes of handover indicated that handovers using direct questions were more frequent in both video conferences (23.8% ISDN, 18.2% LN and 7.7% FTF). 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 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?

Handovers by naming the next speaker were more frequent in ISDN than in face-to-face (2.7% vs 0.4%). LN (1.1%) was more like FTF. In some instances names were used to address the question to a particular individual as in the following two LN examples:

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

Tagging with questions such as "is that okay?" or "you know" or redundant information were equally frequent in all media (4.3% ISDN, 1.9% LN, 0.8% FTF). Here a participant in an ISDN conference ends a turn with a tag question which both facilitates speaker transition and acts as a check for understanding.

A: The only thing you have to change is ahh 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.

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 their average word length. We analysed both the total number of turns on a meeting by meeting basis and also for each participant. Typically, the meetings held over ISDN were characterised by fewer turns of greater length. There were significantly fewer turns per participant in ISDN compared with LN and face-to-face. The complementary result was that the number of words/turn was significantly greater in ISDN than in the other two media. It is possible that these effects are due to the reduction of brief turns (such as backchannels) in ISDN. To investigate this we repeated the analysis excluding all turns of less 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/meeting in each condition. While 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 in both LN and face-to-face we see many more short turns with higher frequency of interruptions and backchannels.
The following examples show typical interactions for ISDN, LN and FTF. 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.

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 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, LN 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 FTF.
Both FTF and LN conversations have a "quickfire" character with clarifications taking place (LN, lines 3 and 5) and also disagreements (FTF, line 2), showing that participants are able to react quickly to incoming information when they do not understand or when they disagree. This enables misunderstandings to be rapidly detected and addressed.

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, given the problems of managing the channel, participants would rely on two people ("chairpeople"), one at either end of the link, to manage interactions across the connecting link, and they would direct 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 that were produced by the two most frequent speakers in the three conditions. However, there was no overall difference either in the percentage of turns taken by these people or in the number of words that they spoke. We also investigated whether ISDN served to exclude certain speakers: the fact that they were less able to interrupt
might prevent participants who are not chairpeople from having the opportunity to speak. Again this hypothesis was not born 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. This result is interesting because it runs contrary to the perceptions of the people using ISDN and LN. They report feeling both that certain participants are able to dominate the meeting and that others are less able to contribute to it.

5 Discussion

Consistent with our predictions, our results showed that compared with FTF, spoken conversation patterns are disrupted over ISDN with its half duplex line, transmission lags and poor quality image:

+ Listeners produced fewer backchannels and interrupted less.
+ Listeners were also less likely to anticipate turn endings, and hence complete speaker utterances.
+ Speakers also altered their behaviour, 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.

Our theoretical claim is that the above findings result from the properties of the communication channels disrupting basic communication processes. Face-to-face interaction has full duplex, almost instantaneous transmission of audio as well as high quality visual information. When these channel properties are changed to those of the ISDN system, we produce a style of interaction that is lecture-like and lacks spontaneity. These findings are supported by other studies of systems with lagged audio (Cohen, 1982?, Isaacs & Tang, 1993). Listeners in ISDN seem to be more polite,
waiting for a pause or for a speaker to finish before making their conversational contribution.

However, a comparison of LN and FTF, contrary to our expectations, also demonstrated differences. Despite having a full duplex line, immediate transmission and broadcast quality image, the properties of the spoken communication still differed from face-to-face interaction in the following ways:

+ Although listeners interrupt as frequently as in FTF, 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 LN was similar to FTF it was still characterised by highly formal conversational behaviours. How can we explain these findings? The argument that lagged half-duplex audio and poor quality video, are solely responsible for communication disruption can no longer hold true. If these channel properties underlie all the communication disruptions we observed, then we should have seen no difference between LN and face-to-face. In fact we did observe differences between LN and face-to-face. This suggests that other channel properties are also critical here and the account should be extended to include these properties and the conversation features they impact.

What are these other channel properties and how could they affect conversation? ISDN and LN both have non-directional audio and video: in both, sound and vision originate from a restricted number of sources, i.e. one or two monitors and loudspeakers. This contrasts with FTF where sound and visual behaviour are directional, because they emanate from the different participants. In FTF, eye-gaze and localised sound may be used to support the speaker switching process (Kendon, 1967, Duncan, 1972). This might explain why ISDN and LN showed reduced floorholding and more explicit handovers. The absence of such spatial cues in ISDN and LN forces speakers to be more verbally explicit in managing intended speaker switches, possibly because they cannot use cues such as eye-gaze to indicate the next speaker. Recent research on directional
audio and video attempts to test this hypothesis, although results so far have not been positive (Mantei et al., 1991; Sellen, 1992).

We also need to refine our explanations of ISDN in the light of the unexpected differences between LN and FTF. Which disruptions of communication were exclusively due to ISDN channel properties? Comparison of ISDN and LN enables us to determine this. ISDN differed from LN for backchannels, interruptions, projections, simultaneous starts and turn size. One explanation here is that the channel properties of ISDN reduced listener participation. It prevented listeners from indicating assent or dissent and caused them to wait for the original speaker to finish before taking the floor. The effect of reduced listener participation is to decrease the total number of speaker switches and hence the overall number of turns. We did not, however, isolate whether audio lag, half-duplex audio or visual quality was mainly responsible for these 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 features. Currently we cannot rule out any of these channel properties, and other research has independently shown that removing visual cues, lags and half-duplex audio can all independently produce these types of effects (Cohen, 1982; Kreisz & ??, Krauss & Bricker, 1967; Rutter & Stephenson, 1977, Tang & Isaacs, 1993, Whittaker, 1995; in press, Wolf, 1982).

Turning to more subjective results, are participants aware of differences between the different media? Our interviews with users and informal observations suggest they are. When interviewed people stated that videoconferences involved more "effort". For example, participants complained about the difficulty of assuming control of the conversation in both videoconferencing media. 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..." Other people exploited the problems of speaker switching to their
advantage: one LN user acknowledged the greater formality of LN meetings compared with face-to-face 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, notwithstanding the poor image quality in ISDN. Another stated advantage was the feeling of "not talking into a void". Finally our users had clear ideas about the limitations of ISDN for certain tasks: they commented that they found it appropriate for only certain types of meeting such as information exchange or project updates.

6 CONCLUSIONS

The method we describe here aimed to predict how the nature of communication was affected by the channel properties in video-mediated communication systems. In the first place it contributes to our practical knowledge about the nature of video-mediated communication. Many potential explanations have been suggested for the failure of video-conferencing to gain widespread acceptance, including cost, incorrect marketing and the questionable value of a video channel. However, there have been few detailed empirical studies of the actual communication that occurs over real implementations of wide area video-conferencing systems. Previous research has explained the characteristics of videoconferencing in terms of concepts that are difficult to operationalise such as "social presence" (Short, et al., 1976), "media richness" (Daft & Lengel, 1984) or "cuelessness" (Rutter & Robinson, 1981). Other work has shown that the organisation of mediated communication is critically dependent on the properties of the communication channels (Whittaker, Brennan & Clark, 1991; Whittaker, 1992) and the predictions we generated were derived from a framework based on face-to-face interaction. Here we were able to test predictions based on an analysis of face-to-face
interaction, about how lags, half-duplex audio and poor quality visual information would disrupt specific aspects of communication. In doing so, the study contributes to a developing theory of mediated communication. We now examine the practical implications of these results before discussing the limitations of this method.

Firstly, we can draw implications about the kinds of tasks for which the current ISDN quality videoconferencing is appropriate. The "lecture-like" character and the inability to support quickfire exchanges may 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 and Lengel, 1984; Whittaker, 1992). If ISDN cannot effectively support these tasks, this may contribute to the lack of success of this quality of videoconferencing. It may be that future remote collaborators have to choose appropriate media to the task at hand and ensure that certain types of task, e.g. conflict resolution and negotiation, are resolved in face-to-face situations. Naturalistic studies of remote collaborators who are using multiple media should be conducted to determine how people currently allocate media 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.

From a technology perspective it would seem that introducing low lag, full duplex channels will lead to improvements in communication, as evidenced by the superiority of LN 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 above. However the LN 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 LN and FTF (Sellen, 1992). A second strategy would be to modify existing videoconferencing systems by acting upon our users' comments. They suggested providing remote audio and video controls,
so that remote participants are able to choose what they want to see and hear rather than have these choices made for them. They also suggested the use of several monitors to enable one monitor to 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 other remote participants for visual context.

Additionally, while we have been able to suggest technology improvements based on this study, we have not explored alternative uses of video technology. Another approach to improving video systems is to examine the use of the video image for things other than pictures of participant's head and shoulders. Elsewhere an alternative application of video in the form of "video as data" is presented (REF NARDI et al. CHAPTER, Nardi et al., CSCW article and Whittaker 1995, in press). This work suggests, that in some environments, it may be better to transmit pictures of the work itself rather than of the participants who are carrying out the work. Further applications can be seen in the area of "open distributed offices" where video is used to provide background awareness rather than as a direct communication channel, (REF BELCORE XEROX CHAPTERS). Methods which support these type of investigations are presented elsewhere in this book.

This method has provided a way of quantifying and illustrating some important differences between communication that has been video-mediated and that which occurs face-to-face. While this method served to objectively measure differences in the spoken characteristics of communication, an inherent problem, is that it does not measure the effectiveness of the communication across different media as laboratory studies have done (Chapanis 1975; Morley & Stephenson, 1969, 1970, 1977; Wichman 1970). One problem is that with real life meetings, it is not clear what is an appropriate objective measure of successful communication, nor how we can easily compare the success of the different meetings. However, other work suggests 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 and Cohen, 1991).

Another hypothesis which is not addressed by this method is whether media differences result in changes in the content of conversation across different media. It is not clear whether the topics discussed varied across media. It has been suggested that video-mediated meetings are more task focused and as a consequence save time for participants, (Johansen, 1984, Short et al., Rutter & Robinson, 1977?). It is not possible to support or refute such a hypothesis using this method.

Finally some practical difficulties in applying this method should be noted. Because of the detailed nature of the analysis the method is a time consuming one. Analysis must be carried out on each verbal utterance, determining its origin, function and timing. There have been some attempts to automate this process using recording systems that can time utterances and determine if they were uttered in overlap etc., (Cohen, Sellen CHI paper). However, as we have seen, a single measured feature of speech may arise for multiple conversational reasons: for example, speech delivered in overlap may arise for one of several reasons and gross measures will not capture these differences. It is likely even where automated analysis is used that additional manual measures will need to be undertaken.

Nevertheless, the method has useful practical and theoretical implications. By quantifying the different conversational characteristics of different mediated communication systems, we have been able to identify critical interaction processes that are compromised by system properties. We have therefore been able to suggest how system redesign might improve these processes. In addition, we have been able to refine theoretical terminology such as "cuelessness" and "media richness" to identify and predict measurable characteristics of conversation that are affected when conversation is
mediated. This allows replicability and comparison of results with other similar studies (eg Cohen, 19982, Isaacs & Tang, 1993, this volume, Sellen, 1992, in press, this volume).
REFERENCES


Figure 1: Layout of the ISDN video-conferencing room
Figure 2: Meeting area of a LIVE-NET Conferencing Room
Figure 3: The LIVE-NET monitor set-up
<table>
<thead>
<tr>
<th></th>
<th>ISDN</th>
<th>LN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Backchannels</td>
<td>fewer</td>
<td>same</td>
</tr>
<tr>
<td>Interruptions</td>
<td>fewer</td>
<td>same</td>
</tr>
<tr>
<td>Overlaps</td>
<td>(?)</td>
<td>same</td>
</tr>
<tr>
<td>Handovers</td>
<td>more</td>
<td>same</td>
</tr>
<tr>
<td>Turn size</td>
<td>larger</td>
<td>same</td>
</tr>
<tr>
<td>Dominance</td>
<td>more</td>
<td>same</td>
</tr>
</tbody>
</table>

Figure 4: Expected characteristics relative to Face-to-face interaction for the two videoconferencing technologies
<table>
<thead>
<tr>
<th>Functions/Activities</th>
<th>ISDN</th>
<th>LN</th>
<th>FTF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary</td>
<td>Information Exchange</td>
<td>Information Exchange</td>
<td>Information Exchange</td>
</tr>
<tr>
<td>Secondary</td>
<td>Appraisal, Generating Ideas</td>
<td>Appraisal, Problem Solving</td>
<td>Task Allocation, Tactical decision making, Problem solving</td>
</tr>
<tr>
<td>Primary</td>
<td>Information Exchange</td>
<td>Information Exchange</td>
<td>Information Exchange</td>
</tr>
<tr>
<td>Secondary</td>
<td>Appraisal, Generating Ideas</td>
<td>Task Allocation, Decision Making</td>
<td>Problem Solving</td>
</tr>
<tr>
<td>Primary</td>
<td>Information Exchange</td>
<td>Information Exchange</td>
<td>Information Exchange</td>
</tr>
<tr>
<td>Secondary</td>
<td>Task Allocation, Problem Solving</td>
<td>Appraisal, Problem Solving</td>
<td>Appraisal, Generating Ideas</td>
</tr>
<tr>
<td>Primary</td>
<td>Information Exchange</td>
<td>Information Exchange</td>
<td>Information Exchange</td>
</tr>
<tr>
<td>Secondary</td>
<td>Appraisal, Generating Ideas, Decision Making</td>
<td>Problem Solving</td>
<td>Work-related Gossip</td>
</tr>
<tr>
<td>Primary</td>
<td>Information Exchange</td>
<td>Information Exchange</td>
<td>Information Exchange</td>
</tr>
<tr>
<td>Secondary</td>
<td>Appraisal, Generating Ideas</td>
<td>Appraisal</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 5:** Primary and Secondary Functions and Activities for each of the fourteen recorded meetings
<table>
<thead>
<tr>
<th></th>
<th>ISDN</th>
<th>LN</th>
<th>FTF</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Backchannels</td>
<td>7.00</td>
<td>30.50</td>
<td>60.80</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td></td>
<td>(2.1)</td>
<td>(6.1)</td>
<td>(9.1)</td>
<td></td>
</tr>
<tr>
<td>Total Interruptions</td>
<td>1.40</td>
<td>13.00</td>
<td>18.60</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td></td>
<td>(0.9)</td>
<td>(2.2)</td>
<td>(3.7)</td>
<td></td>
</tr>
<tr>
<td>Interruptions excluding channel breaks</td>
<td>0.20</td>
<td>11.75</td>
<td>18.60</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td></td>
<td>(0.2)</td>
<td>(3.3)</td>
<td>(3.7)</td>
<td></td>
</tr>
</tbody>
</table>

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

‡. significantly different from FTF in post hoc ANOVA test.

†. significantly different from LN in post hoc ANOVA test.

*. shown in parentheses
<table>
<thead>
<tr>
<th></th>
<th>ISDN</th>
<th>LN</th>
<th>FTF</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Overlaps per turn</td>
<td>9.6%</td>
<td>12.3%</td>
<td>10.1%</td>
<td>n.s.</td>
</tr>
<tr>
<td></td>
<td>(1.5)</td>
<td>(1.5)</td>
<td>(0.7)</td>
<td></td>
</tr>
<tr>
<td>Overlaps resulting from projection / completion per turn</td>
<td>2.9%††</td>
<td>9.2%</td>
<td>7.3%</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td></td>
<td>(1.1)</td>
<td>(0.7)</td>
<td>(0.9)</td>
<td></td>
</tr>
<tr>
<td>Overlaps occurring during floor holding per turn</td>
<td>0.0%‡</td>
<td>0.6%‡</td>
<td>1.8%</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td></td>
<td>(0.0)</td>
<td>(0.3)</td>
<td>(0.3)</td>
<td></td>
</tr>
<tr>
<td>Overlaps from simultaneous start</td>
<td>6.7%††</td>
<td>2.5%</td>
<td>1.0%</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td></td>
<td>(0.9)</td>
<td>(1.1)</td>
<td>(0.5)</td>
<td></td>
</tr>
</tbody>
</table>

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

‡. significantly different from FTF in post hoc ANOVA test

†. significantly different from LN in post hoc ANOVA test

* shown in parentheses
<table>
<thead>
<tr>
<th></th>
<th>ISDN</th>
<th>LN</th>
<th>FTF</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total Handovers</strong></td>
<td>30.8%‡</td>
<td>21.2%‡</td>
<td>8.8%</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td></td>
<td>(5.7)</td>
<td>(2.0)</td>
<td>(0.8)</td>
<td></td>
</tr>
<tr>
<td><strong>Handovers by question</strong></td>
<td>23.8%‡</td>
<td>18.2%‡</td>
<td>7.7%</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td></td>
<td>(3.4)</td>
<td>(1.7)</td>
<td>(1.0)</td>
<td></td>
</tr>
<tr>
<td><strong>Handovers by tagging</strong></td>
<td>4.3%</td>
<td>1.9%</td>
<td>0.8%</td>
<td>n.s.</td>
</tr>
<tr>
<td></td>
<td>(1.6)</td>
<td>(0.6)</td>
<td>(0.3)</td>
<td></td>
</tr>
<tr>
<td><strong>Handovers by name</strong></td>
<td>2.7%‡</td>
<td>1.1%</td>
<td>0.4%</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td></td>
<td>(0.9)</td>
<td>(0.4)</td>
<td>(0.2)</td>
<td></td>
</tr>
</tbody>
</table>

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

‡. significantly different from FTF in post hoc ANOVA test

†. significantly different from LN in post hoc ANOVA test

*: shown in parentheses
<table>
<thead>
<tr>
<th></th>
<th>ISDN</th>
<th>LN</th>
<th>FTF</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of turns per meeting</td>
<td>74.2††</td>
<td>180.0</td>
<td>199.2</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td></td>
<td>(13.6)</td>
<td>(31.7)</td>
<td>(17.3)</td>
<td></td>
</tr>
<tr>
<td>No. of words per meeting</td>
<td>3212.0</td>
<td>3529.5</td>
<td>3386.8</td>
<td>n.s.</td>
</tr>
<tr>
<td></td>
<td>(251.2)</td>
<td>(165.6)</td>
<td>(283.5)</td>
<td></td>
</tr>
<tr>
<td>No. of turns by participant</td>
<td>12.37††</td>
<td>23.77</td>
<td>34.82</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td></td>
<td>(2.01)</td>
<td>(4.29)</td>
<td>(6.07)</td>
<td></td>
</tr>
<tr>
<td>No. of words by participant</td>
<td>535.3</td>
<td>455.5</td>
<td>603.2</td>
<td>n.s.</td>
</tr>
<tr>
<td></td>
<td>(109.0)</td>
<td>(98.1)</td>
<td>(114.0)</td>
<td></td>
</tr>
<tr>
<td>No. of words per turn</td>
<td>43.61††</td>
<td>19.23</td>
<td>17.08</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td></td>
<td>(3.42)</td>
<td>(1.34)</td>
<td>(0.98)</td>
<td></td>
</tr>
<tr>
<td>No. of words per turn not</td>
<td>62.19††</td>
<td>30.00</td>
<td>31.30</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>including turns of less than</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>five words</td>
<td>(4.47)</td>
<td>(2.02)</td>
<td>(1.66)</td>
<td></td>
</tr>
</tbody>
</table>

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

†. significantly different from FTF in post hoc ANOVA test

‡. significantly different from LN in post hoc ANOVA test

* shown in parentheses
<table>
<thead>
<tr>
<th></th>
<th>ISDN</th>
<th>LN</th>
<th>FTF</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turns taken by two most</td>
<td>66.84%</td>
<td>58.48%</td>
<td>73.45%</td>
<td>n.s.</td>
</tr>
<tr>
<td>frequent speakers / total</td>
<td>(6.87)</td>
<td>(3.66)</td>
<td>(4.54)</td>
<td></td>
</tr>
<tr>
<td>turns</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turns taken by two least</td>
<td>8.28%</td>
<td>2.29%</td>
<td>10.56%</td>
<td>n.s.</td>
</tr>
<tr>
<td>frequent speakers / total</td>
<td>(2.93)</td>
<td>(0.84)</td>
<td>(6.80)</td>
<td></td>
</tr>
<tr>
<td>turns</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Words spoken by two most</td>
<td>78.10%</td>
<td>70.76%</td>
<td>76.58%</td>
<td>n.s.</td>
</tr>
<tr>
<td>frequent speakers / total</td>
<td>(5.65)</td>
<td>(4.89)</td>
<td>(3.64)</td>
<td></td>
</tr>
<tr>
<td>words</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Words spoken by two least</td>
<td>5.24%</td>
<td>1.38%</td>
<td>11.6%</td>
<td>n.s.</td>
</tr>
<tr>
<td>dominant speakers / total</td>
<td>(2.42)</td>
<td>(0.56)</td>
<td>(6.46)</td>
<td></td>
</tr>
<tr>
<td>words</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 10: Normalised mean percentage and mean standard error* of turns and words spoken by most and least frequent speakers, showing levels of statistical difference for the three conditions

* shown in parentheses