FINAL REVISION FOR THE JOURNAL OF COMPUTER SUPPORTED COOPERATIVE WORK
DRAFT OF OCTOBER 17, 2000

Co-Constructing Non-Mutual Realities: Delay-Generated Trouble in Distributed Interaction

KAREN RUHLEDER
Graduate School of Library and Information Science, University of Illinois, Urbana-Champaign, 501 East Daniel Street, Champaign, IL 61820, USA
E-mail: ruhleder@uiuc.edu

BRIGITTE JORDAN
Xerox Palo Alto Research Center, 3333 Coyote Hills Road, Palo Alto, CA 94303, USA
E-mail: jordan@akamail.com

Abstract

The use of remote communication technologies to carry out daily work is becoming increasingly common, and their use in certain settings is already commonplace. Yet, in spite of the fact that significant sums are being spent on the acquisition of technologies to support distributed work, we are only beginning to understand the intricacies of these interactions. This paper identifies and analyzes one particular limitation of video-based teleconferencing, the impact of an audio and video delay on distributed communication. It offers a detailed microanalysis of one distributed team's use of videoconferencing to support remote teamwork. We explore through this analysis the impact which technology-generated delays may have on shared meaning-making between remote participants. We draw conclusions about the significance of our findings for understanding talk, interaction and collaboration across remote links, and conclude with recommendations for designers, users and implementers.

Keywords

CSCW, remote collaboration, telework, videoconferencing, audio, conversation analysis, interaction analysis

1. Remote Communication

1.1 The Problem

Corporations, government agencies and academic institutions increasingly rely on remote communication to carry out their daily work. Audio, video and data communications between remote teams are becoming increasingly common, and their use in certain settings is already commonplace and unremarkable. Yet, in spite of the fact that significant sums are being spent on the acquisition of technologies to support distributed work, we are only beginning to understand the intricacies of these interactions. This paper identifies and analyzes one particular limitation of video-based teleconferencing, the impact of an audio and video delay on distributed communication. Our aim is to map out this obstacle through our analysis and to suggest ways in which designers and users can establish a synergy between new technologies and new work practices, thereby contributing to an on-going dialogue within the CSCW community.

We are engaged in long-term research to investigate how communication technologies affect interaction and collaboration across distributed sites. Our approach for the current phase of our research focuses both on the *types of interactions* carried out over remote links and on the *characteristics of the technologies* which support (or hinder) those exchanges. Specifically, we are interested in what interactions may be best suited for different kinds of remote communication and what work practices are required to support them.

The body of our paper is concerned with a detailed micro-analysis of some of the interactions the new video-, audio- and data-sharing technologies support. In particular, we explore in detail the impact which technology-generated delays may have on shared meaning-making between remote participants. In the final section, we draw conclusions about the significance of our findings for understanding talk, interaction and collaboration across remote links and provide a potential set of recommendations for designers, users and implementers who seek to integrate these new technologies into their worklife.

Remote communication has become of paramount importance in an increasingly globalizing world. The technologies that support linkages of geographically separated teams, such as video conferencing, internet-based email, fax, cell phones, pagers and the like, are being adopted by corporations to support collaboration between virtual team members. Institutions of higher education are adopting them to support learning where instructor and students are in different parts of the world. Under the pressure of intense competition and the need to cut costs, the possibility to solve problems with the help of the new communication technologies is often eagerly espoused by progressive decision-makers. They are bombarded with manufacturers' promises to cut corporate travel time and expense, or to increase access to education while lowering the cost of providing it.

Sometimes these promises work out to everybody's satisfaction. At other times, people may decide after a while to go back to their old set-up. Most frequently, new arrangements for communication require serious adjustments both to the technology and the supporting maintenance structure and to people's habitual ways of working and learning. In our studies, we have been struck by a particular phenomenon: while participants may be positive about the new

technologies at first, they often confess when we come to know them better that they really don't like it. They may say that it makes them feel uncomfortable or that it takes a face-to-face meeting to really get to know the other person, but have no real sense of the origin of these feelings. These reports are supported by laboratory and field studies that mention greater discomfort, lower levels of trust, and greater skepticism about others' competence in remote interaction (O'Conaill et al. 1993; Isaacs et al. 1995; Olson, et al. 1995; Sellen 1995; Storck and Sproull, 1995; see also early work on delay and telephony, e.g., Riesz and Klemmer, 1963). These studies, however, also fail to identify the mechanism through which these dynamics arise.

This set of often vague complaints and dissatisfactions motivated us to try to come to a better understanding of the dynamics of remote communication with an eye to providing guidelines to developers and implementers. To anticipate the results of our study, we found that the negative impact of audio and video transmission delay between geographically separated parties is pervasive but unrecognized. It is this technology-generated delay that may, at least in part, account for the discomfort people experience in videoconferencing. We hypothesize that the mechanisms through which transmission delay affects trust and confidence between communicants are *turntaking, sequence organization* and *repair.* It is through these mechanisms that participants in an interaction construct shared meaning and demonstrate social and subject matter competence.

1.2 Delay in Remote Communication

One feature of most remote communication technologies currently in use is that they generate transmission delays.[1] This affects the way in which communicants participate in the conversation on both sides of the link. For example, in comparing meetings using video-conferencing technologies with face-to-face meetings, Tang and Isaacs (1993) found that a .57 second one-way delay in audio transmission markedly disrupted turntaking. The audio delay led to difficulty in negotiating turntaking, with communicants less likely to engage in complex, subtle or hard-to-manage interactions. Earlier studies have documented similar disruptions in turntaking when delay was present (Cohen 1982; O'Conaill et al. 1993). Tang and Isaacs (1993) found that, overall, while users wanted video as a component of a conferencing system, they would tolerate a video delay far more easily than an audio delay. Again, length of delay is key, and video settings not characterized by discernible delay exhibit no effect on turn frequency, duration or distribution (Sellen 1995).

We are particularly interested in how delay affects communicants' *experience* of the conversation. Consider a hypothetical conversation between two remote collaborators. One person asks her collaborator a question, which he answers as soon as he hears; his response then travels back to her. She thus hears the response as coming after a gap determined by double the

[11] Some computer conferencing systems generate a delay of one second or more, influenced by factors such as bandwidth and network traffic. In one on-line distance education program we have studied, class chatroom comments suffered from up to 14 second delays. Regarding satellite communication, theorists tell us that there is a lower limit below which the best design can't carry us. Increasingly important for long-distance transmission, the minimal satellite circuit takes at least .5 seconds for the simple round trip (Tirro 1993; Gordon and Walter 1993). Humans perceive silences of .2 seconds or longer. Research has demonstrated that such gaps are significant within the context of human interaction (Rochester, 1973; Thompson and Hopper, 1992; Walker and Trimboli, 1982).

length of the delay inherent in the technology, a gap she can interpret in a number of ways. He, however, thinks he has answered promptly, but may now perceive a gap before receiving her acknowledgment. What is said and heard by users on each side of the communications link is thus different, but in such a way that neither side is aware of the discrepancy. To put it another way, communicants are not co-present to the communication in the same way. This has, as we shall see, far-reaching consequences.

Practically, what we see is that the interaction of the delay with what people say on either side leads to unusual phenomena, including unilaterally perceived gaps, swapped words, and unintended interruptions. Both the cause and consequence of these phenomena are invisible to participants as they engage in real time conversation. Only by detailed post-hoc analysis of the interaction can one see the lack of alignment occasioned by the delay, and the ways in which delay-occasioned phenomena contribute to potentially serious shifts in meaning. Our findings offer evidence for and illustration of these observations.

2. Data and Methods

2.1 Data

The field research that produced the data for our analysis took place at a now-defunct holding company that used groupware and communication technologies to support a distributed work environment. The ethnographic background has been documented in Ruhleder, et al. (1996) and Ruhleder and Jordan (1997). The company headquarters managed several business units distributed across the United States. Lotus Notes® and video conferencing were widely disseminated in order to facilitate interactions between headquarters and these geographically distributed holdings. Different technologies were freely combined to create different possibilities for remote group work, depending on the circumstances, the local availability of specific technologies, and the preferences of group members.

We carried out fieldwork over a period of four months, during which we collected data through participant observation at headquarters and several business units, unstructured interviews, review of on-line and paper materials and video taping of technology-mediated meetings. All headquarters staff and some members of the business units were interviewed at least once over the course of the four months. The resulting data set includes fieldnotes, interview write-ups and partial transcripts, photographs, paper and electronic documents, and a set of video tapes capturing both sides of various small- and large-group remote interactions.

In this paper, we draw examples from a typical 19-minute video segment including pre-meeting and meeting activities between three East Coast software developers and three West Coast accountants. The meeting was conducted using PictureTel®, a video-conferencing technology, and served as a planning session prior to a software prototype demonstration the following day. We set up our own video camera at each of the two sites to videotape participants as they interacted via PictureTel. The resulting analysis produced detailed video transcripts that we used in order to synchronize what people were doing at each site, what they said, and what they heard over the link. In our examples below, we have simplified the transcripts for greater clarity; a full transcript is available from the primary author.

Our analysis in this paper focuses on the audio component of the transmission.[2] Because of the delay of approximately one second, we noticed that what one side heard was different from what the other side heard. Silences were of different duration, cues came at the wrong times. We identified 32 episodes within this 19 minute segment that exhibited these characteristics. These characteristics were identified by multiple analysts without the aid of any special technological manipulation or assistance such as slowing the sounds down or using a metronome. The analytical approach is outlined in the following section.

2.2 Method: Interaction Analysis

We analyzed these video tapes using video-based Interaction Analysis (IA), as outlined in Jordan and Henderson (1995). This technique consists of an in-depth micro-analysis of how people interact with one another, their physical environment with its documents and artifacts, and their "virtual" or "distributed" environment with its remote participants and shared electronic artifacts. Like ethnography in general, IA looks for orderliness and patterns in people's routine interactions, but operates at a finer level of detail than conventional ethnographic observation. The roots of this technique lie in ethnography, sociolinguistics, kinesics, proxemics, and ethology, but it has been shaped most significantly by conversation analysis and ethnomethodology. Having emerged over the past 20 years as a distinct form of analysis, it has been extended to a wide variety of organizational settings.

Interaction Analysis involves several different types of activities on the part of the ethnographer or ethnographic team. Extensive ethnographic fieldwork enables the researcher to identify specific interactions for video taping (in our case, remote meetings) and furnishes a background against which the video taping is carried out. Content logs, which summarize events on a tape, provide an overview of the data corpus and are used for locating sequences for further analysis. They also serve as a basis for making transcripts of particularly interesting segments. Finally, collaborative tape analysis is carried out within a multi-disciplinary group of analysts. Analytic categories are allowed to emerge out of a deepening understanding of the taped participants' interaction. Emerging patterns are checked against other tape sequences and against other forms of ethnographic observations.

These activities are iterative, and frequently overlap. Content logs generate potential tape sequences for analysis; tape analysis suggests further content logging and transcribing with emergent categories in mind. This, in turn, identifies new sequences for analysis, or suggests new venues for video taping. The application of Interaction Analysis within this particular project is outlined in more detail in Ruhleder and Jordan (1997).

3. Analytic Framework

[2] We also collected video tapes of what people saw on the monitors in their respective meetings rooms. The video transmission was also delayed, and was of such poor quality that it was difficult to make out facial expressions and even gestures. For these reasons, we found it more useful to focus exclusively on the audio channel in our analysis.

To make our analysis intelligible, we draw on three concepts from the field of Conversation Analysis: Turntaking (Sacks et al. 1974), Sequence Organization (Sacks 1987 [1973]; Schegloff 1988, 1990, 1997b), and Repair (Schegloff, et al 1977).[3]

3.1 Turntaking

Humans are above all social creatures. To be social means to take turns. This is true on the conversational level, within systems of etiquette, and in social norms about gift exchanges, favors, and reciprocal invitations. Explicit and tacit turntaking systems have shaped human discourse throughout history, between generations, on conscious and unconscious levels.[4] As a matter of fact, turntaking is part of what it means to be human, and being able to enact a particular turntaking system is part of what it means to be a member of that particular social group.

Taking turns at talk is the basic mechanism for interaction, and is supported by both verbal and non-verbal cues (Kendon 1967; Jefferson 1973; Goffman, 1974; Sacks et al. 1974; Goffman 1981; Goodwin 1981). It forms a hidden, underlying foundation for order in human interaction. Conversational turntaking is critical to informal interaction between individuals; communicants expect that other participants in an interaction will be able to appropriately enact the rules that govern social intercourse. Situation-appropriate turntaking is the foundation for ascribing competence to others in face-to-face interaction. When disrupted, it can lead to frustration and misunderstanding (Jordan and Fuller 1975).

In verbal exchanges, speakers project and mark the end of a turn in a number of ways. In informal conversation, they may do this by pitch of voice, by body language, by asking a question, or by verbally letting people know that their story is done ("... and that was that!"). Additionally, they may select the next speaker through the recipient design of the turn itself. This can be done through gaze, for example, or by asking a question or making a statement to which only one person may properly respond ("... but I guess that's nothing compared to the ice storm that just hit Central Illinois," she said, looking at her colleague from Champaign-Urbana).

Who gets to speak next is governed by a basic set of rules for turntaking, here paraphrased from the seminal paper by Sacks, Schegloff and Jefferson (1974):

(1) When the current speaker reaches a point at which their turn may be done, one of three things happen:

---

[3] Our analytical work has been grounded in an extensive body of literature initiated almost 30 years ago by Harvey Sacks, Emanuel Schegloff, and Gail Jefferson, and sustained by them and their colleagues since then. This body of work is too extensive for us to cite individually, but nevertheless continues to inform our work and the work of others.

[4] The French anthropologist Levi-Strauss proposed that marriage rules among preliterate tribes can best be understood as turntaking systems where exogamy rules guarantee that in the course of generations wives are exchanged between groups. He in turn draws on Marcel Mauss who proposed in *The Gift* (1990) that the exchange of gifts in pre-literate societies involved mutual rights and obligations extending over generations.

(a) The current speaker can pass along the turn to another person by gaze or recipient design (e.g., finishing in a way that suggests a next speaker).

(b) If no particular person is indicated through linguistic or non-verbal cues, the first person that starts speaking gets the next turn.

(c) If no one else takes a turn, the original speaker can resume, often building on or adding to the prior turn.

(2) Whichever of these options has been taken up, the same set of options applies for the next turn.

Application of these rules is invisible, and routinely is accomplished with split second timing. This timing is finely coordinated between speaker and co-participant(s). A disruption of this system leads to anywhere from discomfort to breakdown or open rupture.

Current understanding of conversational rules derives largely from face-to-face settings and telephone conversation, but is being extended now through studies of other forms of technology-mediated talk.

3.2 Repair

In face-to-face conversation, when trouble occurs—if, for example somebody mishears what somebody else is saying—a repair is promptly initiated and carried out (Jordan and Fuller 1975; Schegloff et al. 1977; Schegloff 1979b; Schegloff 1997c). If people ask, "huh?" or don't take their turn when they should, or show in some other way that something isn't quite right, the original speaker may repair the problem by repeating, saying it another way, or explicitly indicating who should take the next turn. The repair is initiated by a "trouble flag" like the "huh?" Key here is the notion that the trouble flag is a *shared resource* for multiple participants in an exchange: all parties present understand that something has gone wrong and must be fixed. All participants must recognize some feature of the interaction as troublesome and must act upon it or mutually agree to leave it be.

Under certain circumstances in technology-mediated communication there may be no trouble flag comparable to that of face-to-face communication. The origin of a problem may not be available to either speaker or hearer, as a conversation may be heard differently on each side of the link. As a consequence, the need for repair is not mutually recognized and, though one party may initiate a repair, the other party may not recognize it as such. This may leave participants with a vague but pervasive feeling that something is "not quite right." Alternatively, one side may try to repair trouble that *they alone* perceive, creating confusion for the other side. In "normal" conversation, hitches in the sequencing of turns occur routinely. They are treated as evidence for trouble in the interaction and are remedied by repairs of various sorts. Repairs are initiated as trouble arises, with speakers preferring to self-correct within the next couple of turns after evidence for trouble

first occurs (Schegloff, et al., 1977; Schegloff 1987, 1992, 1997a). The disturbances in turntaking generated by transmission delays are particularly insidious because they raise no easily identifiable trouble flags, yet disrupt the mechanisms of turntaking and repair.

3.3 Sequence Organization

Participants in an interaction monitor and analyze a conversation turn by turn, as each utterance may have implications for what action should or might be taken in the next turn as a response to it. To investigate the implications which one utterance will have for another, let us consider the notion of adjacency pairs from conversation analysis. Adjacency pairs consist of two turns in which the second pair part is contingent on the first. Some commonly occurring pair types are greeting-greeting (farewell-farewell), question-answer, and offer-accept/decline:

Greeting-greeting
A: Hello.
B: Hi there!

Question-answer
A: Did you get my note?
B: Yes, I saw it this morning.

Offer-decline
A: Are you interested in collaborating on this project?
B: You know, I think I'm pretty bogged down at the moment, so I'd better not.

The occurrence of the first part of an adjacency pair strongly constrains what type of next turn is expected. Thus, a question requires an answer, a greeting requires a return greeting, an offer requires an acceptance or a decline. If the expected second pair part is not forthcoming, a "noticeable absence" is created ("he's ignoring me" or "she's in a bad mood").

Conversation analysis has identified as important those two-part sequences in which the second part of the pair embodies either some form of alignment with the first part (a preferred response), or some form of distancing from it (a dispreferred response). "Preferred" or "aligning" does not necessarily mean "affirmative" or "agreeing," but rather refers to consonance with the speaker's expectation about the response. For instance, consider the question, "You're not going to the seminar this afternoon, are you?" In this case, the preferred or aligning response would be a negative one, "no, I'm not," which is consonant with the first speaker's expectations.

Preferred responses are generally delivered right away, without a perceived pause or hesitation. In other words, they are placed contiguously with respect to the first part of the sequence (Pomerantz 1984, Sacks 1987 [1973], Schegloff 1988, Schegloff 1990). Dispreferred responses, however, are framed differently. The second speaker might hesitate for a moment before responding, or might delay the start of the response by hedging or saying "uhhh":

A: Did you remember my book?
B: (pause) Well, um…
I put it right by the door, and then I got a phone call ….

The gap suggests that the response will be a dispreferred one, and so the first speaker may weaken or negate the original statement. This is called a *backdown.*

A: That was a pretty good presentation.
(pause)
If you're into that kind of work.
B: Well, I suppose someone has to do it.

Here, the first speaker anticipated a dispreferred response, or backdown, because of the gap and amended their original statement.[5]

When two parties use a technology that introduces a delay, the first speaker may perceive a pause before a response even when the second speaker does not. This gap may be interpreted as indicative of an impending dispreferred response. In some cases, this leads the first speaker to somehow reframe or weaken their original statement in anticipation of this dispreferred response, *even when the response given (and eventually received) was a preferred or aligning one.* The examples we draw on below outline in detail how the phenomenon of delay can result in differently heard conversations on two sides of a link, and illustrates the impact which this difference may have on participants' perception of the interaction.

4. Trouble in Technology-Mediated Interaction

In face-to-face conversation, listeners hear the speaker right away as they are speaking. The incremental nature of this production— turn-so-far by turn-so-far, word-by-word, syllable-by-syllable, phoneme-by-phoneme— means that a listener hears each component sound at the unfolding cusp of time as it is produced.[6] It is this facility that allows somebody to interrupt at a key moment to finish off an utterance, as it enables them to analyze what they have been hearing bit by bit, and jump in appropriately. This process must work effectively in the production of synchronized utterances.

We are interested in what happens when this process is disrupted because of a delay in transmission. Consider the *hypothetical* conversation in Figure 1, in which no transmission delay occurs.

What Ann Says	What Ann Hears	time in sec	What Bill Hears	What Bill Says
Ann says: Did you get the report I sent you?		1.00	***Bill hears:*** *Did you get the report I sent you?*	

[5] The role of pauses and the meaning people ascribe to them is contextually determined. For instance, Jordan (1992) found that in complex settings such as airport operations, individuals might ask a question and not receive an answer until other tasks had been completed. In our setting, however, the expectation was for immediate response.

[6] Emanuel Schegloff drew our attention to the critical importance of this feature in a working session.

	Ann hears: *Yes, thanks.*	2.00		**Bill says:** Yes, thanks.

Figure 1: *Hypothetical interaction;* Face-to-face, no transmission delay

Ann and Bill are in the same room, speaking face-to-face. When Ann (far left column) asks a question, Bill hears it right away as she speaks it (second column from the right). When Bill replies (far right column), Ann hears the response as it is being produced (second column from the left). Ann asks a question and Bill offers a response that is prompt and appropriate to the situation. The speed with which an answer is delivered is implicative. As outlined above, an immediate response in face-to-face (and also telephone) conversation signifies alignment, while a brief pause or sounds indicating hesitation routinely foreshadow a negative response to a question. In other words, a pause at this point potentially indicates an upcoming lack of alignment between parties.

What happens if Ann and Bill are in different locations, and the medium they are using generates a one second transmission delay? Let's now consider the same hypothetical conversation and the subtle problems engendered by the delay, outlined in Figure 2.

What Ann Says	What Ann Hears	time in sec	What Bill Hears	What Bill Says
Ann says: Did you get the report I sent you?		1.00		
		2.00	***Bill hears:*** *Did you get the report I sent you?*	
		3.00		**Bill says:** Yes, thanks.
	Ann hears: *Yes, thanks.*	4.00		

Figure 2: *Hypothetical interaction;* Technology mediated, one second delay

Ann and Bill are not in the same room, but are speaking via a teleconferencing set-up. When Ann speaks (far left column), Bill hears the question about a second later. Ann, meanwhile, hears no response at all. When Bill replies (far right column), Ann still hears no response for a full second. Bill perceives his answer as having been given promptly, but Ann perceives a delay in the response. Ann asks a question and, while there may still be a shared perception about the appropriateness of the answer, there is no longer a shared perception about its promptness.

From Bill's perspective, this particular exchange is essentially the same as above— Bill heard Ann's question and responded. Yet from Ann's perspective the delayed response opens up multiple new interpretations. She may recognize it as a phenomenon of the technology. Human delay, however, may also be interpreted as hesitancy or doubt or calculation. Ann has framed her question in such a way that she anticipates a positive response. Preferred responses to such question are given immediately and are aligning in nature, that is, they align the interactants with each other. Delay suggests that a non-aligning response is forthcoming. In this case, the delay

may create the expectation that Bill doesn't know what report Ann means, that he can't remember receiving it, or that he is hesitant to discuss it for some reason.

In this hypothetical example, Ann waited and Bill's response eventually reached her. All sorts of other variations are possible: Ann could have rephrased the question during the delay, she could have decided to drop the topic altogether and move on, or she could have given further justification for why she asked her question. This would have sounded inappropriate or awkward to Bill, given that he perceived himself as having answered her question promptly. In any case, the conversation perceived on her side of the link is *different* from the one perceived on Bill's side. The origin of this difference can be found in the technology-generated delay, and the problems are those of alignment. Ann and Bill hear each other's words accurately, and understand their content. The essential difficulty lies in the fact that these words acquire different meaning because of the delay.

If we believe that appropriate turntaking is foundational to ordered human interaction, and that turntaking is the basis for social life, then two people having two fundamentally different conversations with each other raises serious questions about what it means to "share" a conversation in these distributed settings. We now turn to our data, drawing on a set of representative extracts in which transmission delay disrupts turntaking across the link in some way. Both parties, each on one side of a video link, carry out a conversation that appropriately follows a set of conversational rules, but experience the conversation differently because of the silences generated by the technology-induced delay.

4.1 Unintended Interruptions Due to Lack of Perceived Response

In settings where the audio signal is significantly delayed and visual cues are not helpful, the effectiveness of the turntaking system becomes problematic in subtle ways. The examples below highlight one particular phenomenon in which the first person finishes speaking and (because of the delay) is not aware that the second person has started a new turn. Because of the delay in hearing the second person's utterance, the first person begins speaking again.

4.1.1 Example 1: "Is there a LiveBoard?"

In the following example, Ann and Bill are waiting for the meeting to start, and Ann is joined by her co-worker, Amy. Ann and Amy are in one location, Bill at another. The two locations are connected via a teleconferencing system. Just prior to this excerpt, Ann asks if they need to preset a channel for the LiveBoard, a shared electronic whiteboard. Bill hears this question as Amy asks whether or not the LiveBoard will be used (the start of the excerpt below). There is a one second transmission delay.

Location #1			**Location #2**	
What Amy/Ann Say	**What Amy/Ann Hear**	**time/ sec**	**What Bill Hears**	**What Bill Says**
Amy says: Is there a		1.0	***Bill hears Ann:*** *Should we preset*	
LiveBoard?		1.5	*one for the LiveBoard?*	
Ann says: Oh, is there a		2.0	***Bill hears Amy:*** *Is there a*	
LiveBoard?		2.5	*LiveBoard?*	
We're gonna do		3:0	***Bill hears Ann:*** *Oh, is there a*	**Bill says**: Ah, no,
LiveBoard.		3.5	*LiveBoard?*	I don't think...

	***Amy/ Ann hear**: Ah, no,*	4.0	*We're gonna do*	
	I don't think...	4:5	*LiveBoard.*	
Ann says: No?		5.0		
		5:5		
Ann says: OK		6:0	***Bill hears Ann:** No?*	**Bill says**: No
		6:5		
	***Amy/Ann hears**: No*	7:0	***Bill hears Ann:** OK*	

Figure 3: "Is there a LiveBoard?"

Ann asks if the LiveBoard will be used, then states, "We're gonna do LiveBoard." As Ann makes her statement, she hears Bill start to say, "Ah, no, I don't think…," but trails off. Ann interprets this to mean that they will not be using the LiveBoard. She seeks to confirm this, "No?" and then acknowledges the lack of LiveBoard without waiting for a reply. Finally, she hears Bill confirm that there is no LiveBoard.

Bill hears a slightly different version of the conversation. He hears Amy's initial question and begins to answer it, interrupted by Ann's repetition of the question. He then hears Ann's assumption that the LiveBoard would be used, and her question seeking confirmation. Bill answers her and hears an acknowledgment from her.

These kinds of overlaps and mis-answered questions are typical of what happens in quick interchanges over a remote link. As in the example above, wrong questions get answered. In this case, Bill tries to answer Ann's question in a timely manner at 3.0, only to find himself in overlap with Ann. She not only interrupts him, but also contradicts what Bill has started to say by stating at 4.0 that they *are* going to use the LiveBoard.

From Ann's perspective, however, Bill starts to answer her question at an appropriate place at 4.0, but then trails off for no apparent reason. Ann infers his answer and settles the question, closing off the topics with her "OK" at 6.0. From her perspective, Bill adds a "no" that is inappropriate in that particular context. To Bill, however, topic closure appears to work differently. Bill hears Ann ask "No?" at 6.0. He confirms her inference that no LiveBoard will be used, then hears her acknowledgment in a contextually appropriate sequence.

4.1.2 Example 2: Waiting for the "OK"

In this example, an acknowledgement on the part of one side of the link disrupts a continuation on the part of the other side's presentation. This disruption in the flow of conversation again is experienced differently on both sides of the link.

Location #1			Location #2	
What Ann Says	**What Ann Hears**	**time in sec**	**What Bill Hears**	**What Bill Says**
	Ann hears: *... everything is basically there,*	1.0		**Bill says**: ...and then that will be the basis
	and then that will be the basis	2.0		for the working prototype.
	for the working prototype.	3.0		OK?
	OK?	4.0		
Ann says: OK.		5.0		
		6.0	***Bill hears****: OK.*	**Bill says**: and, (pause < 1 sec)
	Ann hears*: and, (pause < 1 sec)*	7.0		and according to our schedule,
	and according to our schedule, ...	8.0		it really depends on if we have…

Figure 4: Waiting for the "OK"

Ann hears Bill outline the procedure for the prototype demonstration and agree upon the aspects of the working prototype that will be shown. She then hears his alignment-seeking "OK?" and responds at once to show that she is still with him in the presentation. At that point, Ann hears Bill start to move on to the next part of the discussion. However, instead of continuing smoothly, he stops suddenly, pauses, and restarts, for no reason apparent to her. The reason for the pause is apparent on Bill's side, however. He has asked for acknowledgment, and received none. After waiting for two seconds, he continues on, only to be interrupted by Ann's "OK."

Here, again, each party acts appropriately, yet perceives the situation differently. From Bill's perspective, Ann doesn't respond at an appropriate moment (4.0), and then interrupts him (6.0). From Ann's perspective, on the other hand, she has responded to Bill's "OK?" and now hears Bill falters for no apparent reason. Ann cannot know that her own "OK" made Bill hesitate before continuing.

Bill and Ann have different information about the interaction that is taking place across the link. They literally hear different conversations in terms of how utterances are ordered and related in time. Since pauses and overlaps are consequential interactional events, each of them is acting in a different stream of events. As a result, they cannot recognize the same troubles or coordinate repair activities. Non-overlapping utterances on one side overlap and collide on the other, while an immediate answer becomes prefaced by a pause. Bill is seeking not just a response, but alignment, and the silence implies that this alignment might not be achieved. Ann's lack of uptake (from Bill's perspective) sets up an expectation of a dispreferred response (like, "no, I don't follow"; "no, this isn't what I expected"), which is then not delivered.

4.2 Rephrasings Due to Expectation of a Dispreferred Response

As discussed above, when a speaker asks a question and does not receive an immediate answer, the implication is that the response will be a dispreferred one. Given this indication, the speaker

may weaken, negate, or modify their original statement in some way. We see this in the examples presented below.

4.2.1 Example 1: "Nobody from Omega Group?"

In the following example, Ann and Bill are talking prior to the formal start of a project meeting. Bill lists off who is there, and Ann asks about the members of a workgroup within the organization, Omega Group, who had participated in previous meetings.

Location #1			Location #2	
What Ann Says	**What Ann Hears**	**time in sec**	**What Bill Hears**	**What Bill Says**
Ann says: Nobody from Omega Group?		1.0		
Ann says: ..er..* [*not heard by Bill]		2.0	***Bill hears:*** *Nobody from Omega Group?*	
Ann says: They're not in-		3.0		**Bill says:** No.
volved?	***Ann hears:*** *No.*	4.0	***Bill hears:*** *They're not involved?*	
Ann says: OK		5.0		
		6.0	***Bill hears:*** *OK.*	

Figure 5: "Nobody from Omega Group?"

Let's consider this exchange from Ann's perspective. At 1.0, Ann asks Bill for clarification, confirmation, or some form of account. She then hears silence, implying that a dispreferred answer is forthcoming. After a second has passed, Ann adds another unit to her turn, saying, "They're not involved?" Given the expectation set up by prior meetings, the first question, "Nobody from Omega Group?" could be answered in a number of ways: perhaps they weren't available, perhaps they'll be late, perhaps they've been reorganized. The delay-induced silence, however, generates a negative reformulation on Ann's part that implies the answer that she expects based on the Bill's lack of up-take.[7] Ann then hears what seems to be a prompt answer to her reformulation, "No," which she acknowledges with equal promptness, "OK."

Let's consider this now from Bill's perspective. Bill hears the original question about Omega Group and answers it promptly, "No." At this point, it is appropriate for Ann to take the next turn with a response or another question. However, its placement gives it a meaning different from the one intended. Bill has already answered the question when she asks, "They're not involved," and her "OK" is too late and out of sequence if it is meant to acknowledge his answer.

On the surface, an orderly interchange of words has occurred, yet it has not facilitated the orderly creation of shared meaning. Ann reformulates a question based on a seeming lack of response from Bill, while he is asked a question he thought he had already answered. What Ann adds as a result of the perceived silence on Bill's part ("They're not involved") is implicative. In face-to-face conversation, such additions project to both parties the imminent occurrence of a non-

[7] We have not pursued this systematically, but in this sequence it is particularly noticeable that Ann looks down at her hands and fidgets during the silence. We speculate that this may be a sign of expectations not met.

aligning response. In cases such as these, the silence is only observed by one party to the interaction, leading that person to react by adding another unit to their turn, thus leading their conversational partner to hear something unexpected. Multiple episodes of this kind open up the possibility for increased confusion or discomfort across the link.

4.2.2 Example 2: "Maybe I didn't communicate clearly…"

As we have seen above, the delay-induced silence heard by one party can generate unnecessary added installments to a turn. Another feature of the delay is that a timely answer may interrupt the speaker during that added installment and, in turn, the addition may interrupt an answer or continuation. The following example illustrates both situations. Bill is finishing up an answer to a previous question about the status of a particular component of the software being developed, with his co-worker Bob interjecting his assurance (noted under "What Ann Hears"). Bill's addition to his turn interrupts Ann's response, and her response interrupts his addition.

Location #1			**Location #2**	
What Ann Says	**What Ann Hears**	**time in sec**	**What Bill Hears**	**What Bill Says**
	Ann hears: *They're all part*	1.0		**Bill says:** and they'll be
	of the project (Bob: "they're in")	1:5		represented in the
	and they'll be	2.0		prototype that you get,
	represented in the	2:5		that we're talking about,
	prototype that you get,	3.0		through the report screens.
	that we're talking about,	3:5		
	through the report screens.	4.0		**Bill says**: Maybe I
Ann says: So you're		4:5		didn't communicate
	Ann hears: *Maybe I didn't communicate*	5.0	***Bill hears****: So you're*	that clearly.
Ann says: Yeah.	*that clearly.*	5:5		
You're farther along		6.0		
than we thought,		6:5	***Bill hears****: Yeah.*	
I think, which is good…		7.0	*You're farther along...*	

Figure 6: "Maybe I didn't communicate clearly…"

Ann hears Bill finish an explanation about the report screens that will be part of the prototype demo. She starts to respond. Just as she begins to speak, Ann hears Bill break in with a statement about not having communicated clearly. She acknowledges his statement, then returns to her original statement commenting on the status of the project.

Bill, on the other hand, experiences the episode differently. He finishes his explanation, but receives no acknowledgment or comment. About half a second later, he suggests one possibility, that his explanation wasn't clear enough. Before this addition to his turn is complete, Bill hears the start of Ann's next statement overlap with the end of his. There is a second's worth of silence on her part, following an affirmation potentially implying that he *didn't* communicate clearly. Then she continues.

The delay affects the order in which people hear the other party's utterances and the length of the silences that they perceive. This phenomenon causes different problems on each side of the link

in terms of each side's ability to understand why things are happening the way they are. Bill's comment at 4:0 about not having communicated clearly is made in reaction to a perceived lack of response from Ann, implying lack of understanding or alignment. But she did respond in a timely manner, introducing a follow-on to his explanation at 4:5, which she hears Bill cut off. At 6.0 she resumes her original statement about how far along the project is.

The phenomenon of delay engenders unnecessary additions and unintended interruptions from the perspective of only one side of the link at each occurrence. What is the impact on the broader interaction when these events occur repeatedly over the course of a conversation? Each time, one person thinks they have behaved appropriately while the other perceives the behavior as inappropriate or disruptive. We speculate that these situations, in which multiple problematic exchanges pile trouble upon trouble on both sides of the link, are at the root of the feelings of discomfort and mistrust reported by users of communication technologies characterized by delay.

4.2.3 Example 3: "I saw your note about it"

Another example follows, again with a collision that occurs when current speaker adds something to their turn, unaware that the other party is already in the process of responding:

Location #1			**Location #2**	
What Ann Says	**What Ann Hears**	**time in sec**	**What Bill Hears**	**What Bill Says**
Ann says: So, did you see		1.0		
that Beta Corp has		2.0	***Bill hears:*** *So, did you see*	
a new release coming		3.0	*that Beta Corp has*	
out? (pause just < 1 sec)		4.0	*a new release coming*	
I sent you a note about it.		5.0	*out? (pause just < 1 sec)*	
		6.0	*I sent you a note about it.*	**Bill says**: I saw your note about it.
	Ann hears*: I saw your note about it.*	7.0		
Ann says: Yeah. You would have thought		8.0		
when they were here they at least…		9.0	***Bill hears:*** *Yeah. You would have thought...*	

Figure 7: "I saw your note about it."

Ann asks Bill about a new Beta Corp release. Getting no response, she adds to her turn by specifying the form in which she believes he should have received the information ("I sent you a note about it."). This changes the terms of the prior turn and the appropriate response. After a further silence, she hears his response that he saw the note, at which point she continues with a comment about the people from Beta Corp.

Bill hears Ann's question, then hears a silence. He begins a response, "I saw your note about it." His response at 6.0, however, is overlapped by Ann's new turn in which she comments on the note he is currently talking about. After a couple of seconds, he hears Ann continue on about Beta Corp.

How does this look from the two perspectives represented here? Ann introduces a new topic of conversation, directly asking a question of Bill at 1.0. The framing of the first question anticipates a positive response. The lack of response she perceives suggests instead that she will receive a dispreferred response, and she appends a comment to her previous statement. From his perspective, however, Bill has responded promptly with the *preferred* response, directly referring to the note Ann sent.

What kinds of impressions does this exchange create? If Ann had sent email—perhaps mail which others had received successfully—she may wonder why Bill doesn't respond to her question. Additionally, his answer, framed for the original question, may no longer be the best response to the added unit, as it becomes a very formal restatement rather than a simple response. At the same time, Bill may perceive Ann as being rude for interrupting him. These inconsistencies are obvious to only one party at a time, and are a phenomenon of the delay.

4.3 Misapplied Feedback

In face-to-face conversation, participants who do not currently have the floor may say, "mm, hmm," or "yeah," may complete a sentence for the speaker (Lerner 1991, 1996), interject a word or phrase or assessment, etc. (Schegloff 1981; Goodwin 1986; Jefferson, 1984). These responses may play different roles. A "yeah" on the part of someone who does not have the floor, for example, can signal readiness to take over as speaker (Jefferson 1984). "Uh, huh" or "mm, hmm," on the other hand, can function as a way for a recipient to let the speaker know that the recipient is still with them and expects the utterance to continue. Schegloff (1982:81, 1997a) terms these utterances *continuers.* He writes that "uh, huh" in a possible turn position claims not only, 'I understand the state of the talk,' but embodies the understanding that extended talk by another is going on by declining to produce a fuller turn in that position.

This kind of feedback helps achieve alignment, giving both speaker and audience confidence that they are still paying attention and that they continue to understand (Jordan and Fuller 1975; Goodwin 1981). Schegloff points out that these utterances are particularly important at a point where an informational crux is being delivered. If the utterance is mistimed, it implies that the hearer has not grasped the point of what is being said. In video-supported remote conversation, the medium may fail to support these mechanisms (Dourish and Bellotti 1992; Sellen and Harper 1997). Because of the delay, participants may offer feedback at an appropriate point in time from their perspective, yet their utterance may be heard at a moment when it does not produce the intended meaning. We have identified several instances of this problem.

4.3.1 Example 1: Mistimed Feedback

Consider the following sequence, transcribed somewhat differently than those before in order to focus more clearly on the point at which feedback is given.[8]

[8] The transcript is to be read horizontally. In the first example, under "Ann's perspective, we see Bill's utterance, "Let me check...," with Ann's utterance, "OK," listed below. The point at which she believes she says "OK" is marked by a vertical line to indicate that Ann has timed her feedback to occur just after the name, "Carrie." In the second half, the line indicates that, from Bill's perspective, he hears the "OK" just after saying the word "if."

Ann's Perspective (Location #1)
Bill: Let me check with Carrie and see if she's got any issues. | **Ann:** OK
Bill's Perspective (Location #2)
Bill: Let me check with Carrie and see if she's got any issues. | **Ann:** OK

Figure 8: Mistimed Feedback: "OK"

Bill lets Ann know that he's going to check something with Carrie. Within that turn, he comes to the end of a grammatical clause ("let me check with Carrie"). From Ann's perspective, this is the point at which she provides feedback in the form of an "OK," functioning as a continuer by both letting Bill know she's still with him and that she is ready to hear more about Carrie's involvement. From Bill's perspective the feedback Ann provides comes at an inappropriate point. Instead of hearing her acknowledgment at the end of a grammatical clause, he hears it in the middle of a clause, overlapping with "she's."

While appropriately timed on Ann's part, her utterance arrives at the wrong moment, again leading participants on both sides of the link to experience the exchange differently. Ann has behaved competently in giving support to Bill, but her contribution does not display competence from Bill's perspective due to the effect of the delay. The "OK," intended to align the two participants in the exchange, has that effect on one side only.

In some cases, this form of mistimed feedback actually disrupts the discourse it was intended to support. The following excerpt illustrates how feedback received at an inappropriate moment can induce a noticeable pause or hesitation in the speaker's utterance. Prior to this excerpt, Ann asked whether or not the demonstration of the prototype would be carried out in a particular way. Bill says that option is not possible at the moment, and offers an elaborate explanation for how the demonstration will be organized this time around. In the excerpt below, he shifts from completing his explanation of the current demonstration to a discussion of how demonstrations could be organized differently in the future.

Ann's Perspective (Location #1)
Bill: ...so we could all sit around the table and look at it. But in the future, I think that's (pause) exactly why... | **Ann:** I see
Bill's Perspective (Location #2)
Bill: ...we could all sit around the table and look at it. But in the future, I think that's (pause) exactly why... | **Ann:** I see

Figure 9: Mistimed Feedback: "I see"

As Ann listens to Bill's plan for the demo, she says, "I see," directly after Bill has delivered a key point. She thus appropriately indicates comprehension and alignment on her part and, with "I see" functioning as a continuer, she also indicates that she is willing to hear more. Ann hears Bill continue, then pause for no reason apparent to her before going on. Bill, however, hears Ann's, "I see," at a conversationally inappropriate place. Instead of being placed at the end of a sentence or after a grammatical clause, and instead of reflecting back understanding and readiness, it comes in the middle of a new utterance. From Ann's perspective, she has done her part to help the conversation along and has no access to the reason why Bill might hesitate or pause in the middle of a sentence. From his perspective, however, she interrupted him. It broke his stride, leading him to pause a moment before continuing.

4.3.2 Example 3: Delayed Feedback and Backdowns

We also have examples of delayed verbal feedback which lead to some form of addition to a turn on the part of the speaker, changing the meaning of the original utterance. Frequently this addition will be a backdown, which weakens or negates the speaker's statement in some way. In these instances, the feedback is heard at a place that coincidentally happens to be grammatically appropriate, but which no longer achieves alignment as originally intended.

Ann's Perspective (Location #1)	
Bill:	...we don't have to worry too much about them at this point in time, with this project.
Ann:	| Mm, hmm
Bill's Perspective (Location #2)	
Bill:	...we don't have to worry too much about them at this point in time, with this project.
Ann:	| Mm, hmm

Figure 10: Delayed Feedback: "Mm, hmm"

At this point in the meeting, Bill tells Ann that two reporting-related issues she was concerned about at their previous meeting really aren't all that critical—"we don't have to worry too much about them at this point." Bill doesn't hear Ann's aligning "Mm, hmm" at the point she intended, and adds a backdown, "with this project," thus limiting the circumstances under which these issues don't count. Ann's "mm, hmm," heard after that backdown, no longer applies to the broader statement of not having to worry about these issues, but applies only to the narrower assurance on Bill's part that they aren't important *with this project.*

The following example illustrates this same phenomenon. In this case, the backdown follows on a quasi-joke about working hard on a joint project. It follows a somewhat tense discussion about the readiness of particular modules and the adherence to previously established schedules.

Ann's Perspective (Location #1)		
Bill:	...We work a lot of late hours here, too, <1 sec> when we have to.	<3 sec> I think it's really important....
Ann:	Yeah (laugh).	OK*
		*not heard by the other side
Bill's Perspective (Location #2)		
Bill:	...We work a lot of late hours here, too, <1 sec> when we have to.	<3 sec> I think it's really important....
Ann:		Yeah (laugh).

Figure 11: Delayed Feedback: Laughing "Yeah"

Ann has just asked Bill whether or not a certain target date will be met. At this point, they have tentatively settled the issue, and Ann hears Bill say that they work a lot of late hours, too (referring to the long hours traditionally put in by Ann's team at headquarters). She immediately acknowledges this with her "Yeah" and an appreciative laugh, then hears Bill add, "when we have to," to which she responds, "OK." Several seconds later, she hears him continue talking about the development phase under discussion.

Bill, however, perceives their exchange differently. After his utterance about the long hours, he hears no uptake from Ann. As in the other example, he adds a backdown, "when we have to," what sounds like Ann's response to that backdown (the "yeah" and the appreciative laugh). Bill doesn't hear Ann's subsequent "OK" because it didn't carry across the link. Several seconds later, he returns to their earlier discussion. Bill might not have added, "when we have to," if he had received her response at the appropriate place.

What does the addition of the backdown, "when we have to," mean to a person on each side of the link? Ann hears Bill highly circumscribing the claim that he has made on behalf of his group. Bill, meanwhile, receives no acknowledgment of his team's hard work as he makes his statement about long hours, but hears Ann's "Yeah" and laugh in response to his backdown. The change in meaning by virtue of the timing alters the perception each person has of the other.

5. Conclusions and Implications

Talk is not just about the exchange of information, but about shared meaning-making on multiple levels. The examples above illustrate how delay impacts the ability of conversational participants to create shared meaning through talk via remote communication technologies. In each case:

- the delay engenders some kind of trouble,
- this trouble disrupts the turn-taking system, and
- participants cannot identify the true source of the trouble.

Participants may sense that something might be "wrong"— in several examples above, for instance, one person appears to interrupt or say something contextually inappropriate. The nature of the distributed technology, however, precludes people from identifying the trouble and making repairs. Even when people know about the delay as a technical specification of the system, they may have difficulty recognizing and adjusting their construction of meaning *in real-time* in the course of producing talk over the link as part of a broader set of work activities. In conversations repeatedly punctuated by episodes such as the ones above, explanations for the unanticipated behavior of conversational partners may be overwhelmingly negative—trouble piles itself upon trouble over the course of the conversation.[9]

The potential consequence is a pervasive sense of uneasiness similar to that described in Jordan and Fuller (1975) where Maya and English speakers using Spanish as their *lingua franca* were unable to repair trouble in conversation. They knew something was wrong, but they were unable to locate the source of trouble. We find significant potential for troublesome miscommunication and interpretation in our own data. Here, too, people are unable to identify and repair trouble as it occurs because its origin is obscured.

5.1 Implications for Learning and Technology Development

Technological systems that support distributed communications are already challenging us to rethink the notion of interaction and to revisit our current understanding of exchange systems. Our paper contributes to this dialogue by examining in close detail one particular kind of twist on "ordinary" conversation: the impact of a technology-generated delay on the workings of a conversational system that evolved within a co-located, face-to-face context. We would like to consider this impact in the following terms:

What people can compensate for. As noted earlier, there is already evidence that people pick up on the delay phenomenon and learn to moderate their manner of speaking (O'Conaill et al. 1993). As people become more accustomed to these technologies through on-going use over time, they will become more facile at adapting their manner of communication to the medium.[10] For years to come, however, corporations and other organizations will have to deal with a continuing stream of novices or people whose very sporadic use of these technologies mitigates against any long-term development of effective delay-compensating work practices.

What people must learn to reinterpret. What people cannot pick up are the more serious classes of problems, those where the trouble that is generated by the delay is apparent only to the person on one side of the interaction. This kind of trouble occurs in real-time for participants. The underlying cause may only be recognizable through analysis of both sides of the interaction, something reserved for analysts who have the time to do this kind of detailed work. However,

---

[9] See also Garfinkel's (1967:pp 36–53) experiments with his students, in which they deliberately violated interactional expectations. Those with whom they engaged in interaction interpreted these violations extremely negatively.

[10] Technology-generated modification of interaction is well documented by conversation analytic studies of telephone conversations, in which greetings, leave-takings and turntaking differ from face-to-face interactions and require participants to draw on different resources in managing turntaking and repair (Schegloff, 1979a).

with continued use it may be that people learn to systematically discount apparent hesitancy, rudeness, or inattention. They may learn to recognize and challenge negative reactions to their distributed conversational partners.[11]

How these findings can guide meeting strategies. If certain forms of conversational patterns are difficult to sustain in interactions characterized by delay, then people might want to restrict the use of delay-generating technologies to those interactions in which trouble is less likely to arise. While quick give-and-takes, brainstorming, and trouble-shooting sessions are probably highly vulnerable to this kind of trouble, interactions with explicitly laid-out rules for turntaking, such as structured events and formal meetings, are less likely to suffer from the kinds of trouble illustrated above. News reporting, for example, consists of a highly-structured set of interactions that support smooth transitions between speakers (Clayman and Whalen, 1988/89; Greatbatch, 1988; Heritage and Greatbatch, 1991). For brainstorming or design sessions, especially for teams that know each other, it may be more productive to forgo the video connection and use instead audio-only channels without appreciable delay and some form of shared electronic workspace. By understanding the impact of particular technologies on real-time interactions, users and implementers can more effectively chose an appropriate technology set.

How these findings contribute to technology development efforts. Current development efforts are concerned with improving ways of indicating presence and attention in remote meetings, such as through gaze and peripheral awareness of movement and sound. An extension of these efforts may yield a variety of ways of signaling the current state of a conversation, the readiness of a workgroup to move on to the next topic, the nature of feedback to the speaker given by different workgroup members, etc. For example, just as web browsers tell us what percentage of a page has loaded, systems for collaborative communication may incorporate representations of a distributed conversation—perhaps showing participants how their words are traveling to the other side, indicating the earliest point at which they might hear back.[12] Nevertheless, our work points out the importance of concentrating development efforts on reducing delay as much as possible, as even a delay as small as .2 seconds can significantly affect an interaction.

Our ideas are obviously speculative at this point, and it will take some amount of experimentation and prototyping to transform what we know about remote communication into new interfaces and applications. Technology development efforts will need to continue to improve the ability of a workgroup on one side of the link to judge the state or readiness of members on the other side of the link, thereby contributing to their ability to manage the impact of the delay. When systems development is driven by inaccurate or incomplete assumptions about conversation, the system itself will fail to effectively support the communication process (Tatar et al. 1991).

---

[11] Users of email will recognize this phenomenon, as they must occasionally remind themselves that their perception of the tone of an email message as abrupt or angry may be a product of the medium, and not the intention of the other party.

[12] We thank Praesun Dewan for suggesting this idea.

5.2 In Conclusion

Finally, we suggest that the turntaking system itself be considered with respect to the development and implementation of new communication technologies. In face-to-face conversation, the sequencing of turns forms a basis for meaningful interaction. Hitches, problems and false starts are treated as trouble in interaction that can be repaired. Yet in the interactions we have examined, the affordances of the technology prevented identification of the problem; hence, no repair was possible in real-time. Furthermore, while this paper focuses on the impact of delay, we believe other characteristics merit similar attention. In particular, we would like our analysis and argument not to be restricted to the delay, but to apply to the visual, aural and spatial cues which enable participants to create shared meaning and mutual understanding. New technologies shape, constrain, and expand the ways in which participants in a mediated dialogue perceive cues and engage each other. Future research and development in technology-mediated communication must account for the interaction between the affordances of a technology and the turntaking system.

6. Acknowledgments

Our work has benefited immensely from collaborative working sessions with Emanuel A. Schegloff. We also would like to thank participants in the Interaction Analysis Laboratories we conducted at the Xerox Palo Alto Research Center, the Institute for Research on Learning, and the University of Illinois for the ways in which they challenged and aligned with our take on these issues.

Our research was supported by NSF Grant #9712421, the Xerox Palo Alto Research Center, the Institute for Research on Learning, the Graduate School of Library and Information Science, the University of Illinois Research Board, and the now-defunct Holding Company.

7. References

Clayman, Steven E. and Whalen, Jack (1988/89): When the Medium Becomes the Message: The Case of the Rather-Bush Encounter. *Research on Language and Social Interaction,* vol. 22, pp. 241-272.

Dourish, Paul, Annette Adler, Victoria Bellotti, and Austin Henderson (1996): Your Place or Mine? Learning from Long-Term Use of Video Communication," *Computer Supported Cooperative Work: The Journal of Collaborative Computing,* vol. 5, no. 1, pp. 33–62.

Dourish, Paul, and Victoria Bellotti (1992): Awareness and Coordination in Shared Workspaces. *Proceedings of the Conference on Computer Supported Cooperative Work ,* pp. 107-114.

Garfinkel, Harold (1967): *Studies in Ethnomethodology.* Cambridge, MA: Blackwell Publishers.

Goffman, Erving (1982): *Forms of Talk.* Philadelphia, PA: University of Pennsylvania Press.

Goodwin, Charles. (1979): The Interactive Construction of a Sentence in Natural Conversation. In Psathas, George, ed., *Everyday Language: Studies in Ethnomethodology.* New York: Irvington Publishers, Inc., pp. 97–121.

Goodwin, Charles (1980): Restarts, Pauses, and the Achievement of a State of Mutual Gaze at Turn-Beginnings, *Sociological Inquiry,* vol. 50, nos. 3–4, pp. 272–302.

Goodwin, Charles (1981): *Conversational Organization: Interaction between Speakers and Hearers.* New York: Academic Press.

Goodwin, Charles (1986): Between and Within: Alternative Treatments of Continuers and Assessments. *Human Studies,* vol. 9, pp. 205-217.

Greatbatch, David (1988): A Turn-Taking System for British News Interviews. *Language in Society,* vol. 17, no. 3, pp. 401-430.

Heath, Christian and Paul Luff (1993): Disembodied conduct: Interactional asymmetries in video-mediated communication. In Graham Button (ed), *Technology in Working Order: Studies of Work, Interaction, and Technology,* Routledge, London, pp. 35–54.

Heritage, John and David Greatbatch (1991): On the Institutional Character of Institutional Talk: The Case of News Interviews. In D. Boden and D. H. Zimmerman (eds), *Talk and Social Structure.* Berkeley: University of California Press, pp. 93-137.

Isaacs, Ellen A., Trevor Morris, Thomas K. Rodriguez, and John C. Tang (1995): A Comparison of Face-to-Face and Distributed Presentations. *Proceedings of CHI'95.* Denver CO: ACM Press.

Jefferson, Gail (1973): A Case of Precision Timing in Ordinary Conversation: Overlapped Tag-Positioned Address Terms in Closing Sequences. *Semiotica,* vol. 9, no. 1, pp. 47–96.

Jefferson, Gail (1984): Notes on a Systematic Deployment of the Acknowledgment Tokens 'Yeah' and 'Mm hm.' *Papers in Linguistics,* vol. 17, no. 1–4, pp. 197–216.

Jefferson, Gail (1989): Preliminary Notes on a Possible Metric which Provides for a 'Standard Maximum' Silence of Approximately One Second in Conversation. In D. Roger and P. Bull (Eds.), *Conversation: An Interdisciplinary Perspective,* Clevedon: Multilingual Matters, pp. 166-196.

Jordan, Brigitte (1992): Technology and Social Interaction: Notes on the Achievement of Authoritative Knowledge in Complex Settings. *IRL Technical Report No. IRL92-0027, Palo Alto, CA: Institute for Research on Learning.*

Jordan, Brigitte, and Nancy Fuller (1975): On the Non-Fatal Nature of Trouble: Sense-Making and Trouble-Managing in Lingua Franca Talk. *Semiotica,* vol. 13, no. 1, pp. 1–16.

Jordan, Brigitte, and Austin Henderson (1995): Interaction Analysis: Foundations and Practice. *The Journal of the Learning Sciences,* vol. 4, no. 1, pp. 39–103.

Kendon, Adam (1967): Some Functions of Gaze-Direction in Social Interaction. *Acta Psychologica,* vol. 26, pp. 22–63.

Lerner, G. H. (1991): On the syntax of sentences-in-progress. *Language in Society*, vol. 20, pp. 441-458.

Lerner, G. H. (1996): On the 'Semi-Permeable' Character of Grammatical Units in Conversation: Conditional Entry into the Turn Space of Another Speaker. In E. Ochs, E. A. Schegloff and S. A. Thompson (Eds.), *Interaction and Grammar,* Cambridge: Cambridge University Press, pp. 238-276.

Mauss, Marcell (1990): *The Gift: The Form and Reason for Exchange in Archaic Societies.* Translated by W. D. Hall. London: Routledge.

O'Conaill, Brid, Steve Whittaker, and Sylvia Wilbur (1993): Conversations over Video Conferences: An evaluation of the spoken aspects of video-mediated communication. *Human-Computer Interaction,* vol. 8, pp. 389–428.

Olson, Judy S., Gary M. Olson, and D.K. Meader (1995): What Mix of Video and Audio is Useful for Small Groups Doing Remote Real-time Design Work? *Proceedings of CHI'95.* Denver, CO: ACM Press.

Pomerantz, Anita (1984): Agreeing and Disagreeing with Assessments: Some Features of Preferred/Dispreferred Turn Shapes. In J. M. Atkinson and J. Heritage (eds), *Structures of Social Action: Studies in Conversation Analysis.*

Riesz, R. R., and E.T. Klemmer (1963): Subjective Evaluation of Delay and Echo Suppressors in Telephone Communications, *The Bell System Technical Journal,* November, pp. 2919–2933.

Ruhleder, Karen, and Brigitte Jordan (1997): Capturing Complex, Distributed Activities: Video-Based Interaction Analysis as a Component of Workplace Ethnography. In Lee, Allen S., Jonathan Liebenau, and Janice I. De Gross (eds), *Information Systems and Qualitative Research.* London, UK: Chapman and Hall, pp. 246-275.

Ruhleder, Karen, Brigitte Jordan, and Michael Elmes (1996): Wiring the 'New Organization': Integrating Collaborative Technologies and Team-Based Work. *Annual Meeting of the Academy of Management.* http://alexia.lis.uiuc.edu/~ruhleder/publications/96.academy.html.

Sacks, Harvey (1987 [1973]): On the Preferences for Agreement and Contiguity in Sequences in Conversation. In Graham Button and J.E.R. Lee (eds), *Talk and Social Organization,* Clevedon, UK: Multilingual Matters, pp. 54-69.

Sacks, Harvey, Emanuel A. Schegloff, and Gail Jefferson (1974): A Simplest Semantics for the Organization of Turn-Taking for Conversation. *Language,* vol. 50, pp. 696–735.

Schegloff, Emanuel A. (1979a): "Identification and Recognition in Telephone Openings." In Psathas, G., *Everyday Language: Studies in Ethnomethodology,* pp. 23–78.

Schegloff, Emanuel A. (1979b): The relevance of repair to syntax-for-conversation. In T. Givon, (ed): *Syntax and Semantics 12: Discourse and Syntax,* New York: Academic Press, pp. 261-288.

Schegloff, Emanuel A. (1982): Discourse as an Interactional Achievement: Some Uses of 'uh huh' and other things that come between sentences. In Deborah Tannen (ed), *Analyzing Discourse: Text and Talk.* (Georgetown University Round Table on Languages and Linguistics, 1981), Washington, DC: Georgetown University Press, pp. 71-93.

Schegloff, Emanuel A. (1987): Recycled turn beginnings: A precise repair mechanism in conversation's turn-taking organization. In G. Button and J.R.E. Lee (eds), *Social Organisation,* Clevedon: Multilingual Matters, pp. 70-85.

Schegloff, Emanuel A. (1988): On an Actual Virtual Servo-Mechanism for Guessing Bad News: A Single Case Conjecture. *Social Problems,* vol. 35, no. 4, pp. 442-57.

Schegloff, Emanuel A. (1990): On the Organization of Sequences as a Source of "Coherence" in Talk-in-Interaction. In Bruce Dorval (ed), *Conversational Organization and its Development.* Norwood, NJ: Ablex Publishing Corporations, pp. 51-77.

Schegloff, Emanuel A. (1992): Repair after Next Turn: The Last Structurally Provided Place for the Defense of Intersubjectivity in Conversation. *American Journal of Sociology,* vol. 95, no. 5, pp. 1295-1345.

Schegloff, Emanuel A. (1996): Turn Organization: One Intersection of Grammar and Interaction. In Elinor Ochs, Emanuel A. Schegloff and Sandra A. Thompson (eds), *Interaction and Grammar,* Cambridge, UK: Cambridge University Press, pp. 52-133.

Schegloff, Emanuel A. (1997a): Practices and Actions: Boundary cases of other-initiated repair. *Discourse Processes,* vol. 23, pp. 499-545.

Schegloff, Emanuel A. (1997b): Sequence Organization, unpublished manuscript.

Schegloff, Emanuel A. (1997c): Third Turn Repair. In Gregory R. Guy, Crawford Feagin, Deborah Schiffrin and John Baugh (eds), *Towards a Social Science of Language. Volume 2: Social Interaction and Discourse Structures.* Philadelphia: John Benjamins Publishing Company, pp. 31-40.

Schegloff, Emanuel A., Gail Jefferson and Harvey Sacks (1977): The preference for self-correction in the organization of repair in conversation. *Language* 53:361-382.

Sellen, Abigail J., (1995): Remote conversations: The effects of mediating talk with technology, *Human-Computer Interaction,* vol. 10, pp. 401–444.

Sellen, Abigail J., and Richard Harper (1997): Video in Support of Organisational Talk. In Finn Kensing, Abigail Sellen, A. and Sylvia Wilbur (eds)*, Video-Mediated Communication.* Lawrence Erlbaum Associates, Mahwah, New Jersey, pp. 225–244.

Storck, John and Lee Sproull (1995): Through a Glass Darkly: What Do People Learn in Videoconferences? *Human-Computer Research,* vol. 22, no. 2, pp. 197-219.

Tang, John C., and Ellen Isaacs (1993): Why do users like video? *Computer Supported Cooperative Work: Journal of Collaborative Computing,* vol. 1, no. 3, pp. 163–196.

Tatar, Deborah G., G. Foster, and Daniel G. Bobrow (1991): Design for conversation: lessons from Cognoter, *International Journal of Man-Machine Studies,* vol. 34, no. 2, pp. 185–209.