S. Bødker, M. Kyng, and K. Schmidt (eds.). *Proceedings of the Sixth European Conference on Computer-Supported Cooperative Work, 12-16 September 1999, Copenhagen, Denmark*

Meaning-Making Across Remote Sites: How Delays in Transmission Affect Interaction

Karen Ruhleder
University of Illinois, USA, *ruhleder@uiuc.edu*

Brigitte Jordan
Xerox Palo Alto Research Center, USA, *jordan@parc.xerox.com*

Abstract: Distributed organizations increasingly rely on new video-supported communication technologies that may be subject to transmission delays. These delays systematically misalign the feedback one side receives from the other. Through micro-analysis of video data from a video-supported meeting in a geographically distributed company, we examine the impact of delay in such communication systems We specify some ways in which they may subliminally affect communication between remote parties. We illustrate typical kinds of breakdown and conclude with observations about the impact of delays on distributed interaction and the ways in which these impacts can be mediated.

Remote Communication

Defining the Problem

Companies, government agencies and academic institutions increasingly rely on remote communication to carry out their daily work. Audio, video and data communication between remote teams are become increasingly common, and their use in certain settings is already almost 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 integral to an on-going dialogue within the CSCW community, that is, 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.

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 both designers, users and implementors who seek to integrate these new technologies into their worklife.

Delay

One feature of some remote communication technologies is that they generate transmission delays. Whenever this is the case, what is said and heard by users on each side of the communications link is 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. Consider a conversation between two remote collaborators. One person asks her collaborator a question, which he answers as he hears it. She, however, hears the response as coming after a gap determined by the length of the delay inherent in the technology. He thinks he has answered promptly, she thinks his response was delayed. This has, as we shall see, far-reaching consequences.

Some of the computer conferencing systems in current use generate a delay of one second or more. Earlier work has identified certain effects of video-based communication which may be related to this delay phenomenon (O'Conaill, et al., 1993; Sellen, 1995). This includes a lower frequency of backchannels, speaker feedback, and interruptions, as well as an increase in formal hand-overs. While Sellen finds no effect on turn frequency, duration, and distribution, O'Conaill, et al , noted that longer turns with deliberate hand-offs and trouble switching speakers characterized interactions via technologies with an inherent delay.

Our findings offer further evidence for and additional illustration of these observations. In our analysis, we see that the interaction of the delay with the conversation on either side leads to unusual phenomena, such as words getting swapped around or comments occurring out of place because of the delay. The causes and consequences of these phenomena are often invisible to participants. Only by detailed review of transcripts, based on a painstaking micro-analysis of the interaction, can one see the lack of alignment occasioned by the delay, and the ways in which the delay contributes if not to breakdown at least to potentially serious shifts in meaning. While our setting did not enable us to compare our results to face-to-face meeting situations between the same group of people, a subsequent meeting with an audio-plus-shared workspace set-up was not characterized by any of the problems we observed in the audio-plus-video meeting situation.

The work of O'Conaill, et al. (1993) and Sellen (1995) as well as our own may offer insight into why research across multiple environments has found greater discomfort, lower levels of trust and greater skepticism about others' competence in remote interaction (c.f., Isaacs, et al., 1995; Olson, et al., 1995; Storck and Sproull, 1995; c.f. also early work on delay and telephony, i.e., Riesz and Klemmer, 1963) The phenomenon we describe here, technology-generated delays, may play a role in these dynamics within the context of teleconferencing. We hypothesize that the mechanisms through which transmission delay affects trust and confidence between communicants is *turntaking,* a key element in demonstrating social and subject matter competence.

Understanding the impact of delay is particularly important in light of Sellen's (1995) and Whittaker's (1995) findings that audio forms the essential component in distributed interaction, and Heath and Luff's (1993) complementary finding that video mediated communication makes the performative impact of nonverbal behavior problematic. These findings are set against the significant positive contributions which video can make to collaborative interactions, especially in its ability to support the coordination of distributed collaboration by providing cues about availability, current state, and awareness (Sellen, 1995; Whittaker, 1995; Heath and Luff, 1993; Dourish & Bellotti, 1992). Furthermore, the availability of a synchronous, video-based medium as part of a broader suite of technologies facilitates interaction and improved communication, and long-term users of video-conferencing evolve both visual and verbal strategies to address certain kinds of interactional problems and strengthen certain benefits (Dourish, et al., 1996; Tang and Isaacs, 1993).

This paper is driven by these findings. It identifies particular classes of interactional problems, some of which may have no easy counter-strategy as they are not available to the user in real-time. We draw on these observations in our concluding section, in which we suggest design issues, implementation strategies, and work practices which may ameliorate the impact of delay and enable the realization of the powerful benefits video-conferencing offers.

Turntaking

Taking turns at talk is the basic mechanism for interaction, and is supported by both verbal and non-verbal cues (Kendon, 1967; Sacks, et al., 1974). It forms a hidden, underlying foundation for order in human interaction. Conversational turntaking is as critical in informal interaction between individuals as it is in corporate negotiations. Communicants have the expectation that other participants in an interaction will appropriately enact the rules which 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; Schegloff, et al., 1977).

To be social means to take appropriate 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. 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 competent member of a particular social group.

In verbal exchanges, speakers signal the end of a turn in a number of ways. 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!"). Who gets to speak next is governed by a basic set of rules for turntaking, here paraphrased from a seminal paper by Sacks, Schegloff and Jefferson (1974):

(1) When the current speaker reaches a point at which they may be done taking a turn, one of three things happen:
  (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 who 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) These rules hold sway throughout the conversation.

Applying rules of this sort is something speakers are competent to do as members of a particular social system. Their application is invisible, and requires split second timing. This timing is finely coordinated between speakers and listeners. A disruption of this system leads to anywhere from discomfort to breakdown or open rupture.

For current purposes we are interested in the turntaking that happens in conversations between groups of people whose talk is mediated by technologies subject to transmission delay, and in the ways in which that differs from what would happen if the parties were physically co-present. Current understanding of

conversational rules derives largely from face-to-face settings. Yet we can predict that the circumstances under which people "talk" to each other will become increasingly electronic in the society of the future.

In face-to-face conversation and other speech exchange systems such as lectures or court proceedings, when trouble occurs (if, for example, somebody answers in a way that the hearer finds inappropriate) a repair is promptly initiated and carried out (Jordan and Fuller, 1975; Schegloff, et al., 1977; Schegloff, 1979b). People ask, "huh?" or look confused, or don't take their turn when they should. The original speaker immediately repairs the problem by repeating or rephrasing the utterance. The repair is initiated by a "trouble flag"— the "huh?" or confused look that tells the speaker that something is not right. Key here is that the trouble flag serves as a *shared resource* for multiple participants in an exchange: both parties present understand that something has gone wrong and must be fixed.

In technology-mediated remote 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 will be heard differently on each side of the link. As a consequence, no specific repair can be initiated. This leaves participants with a vague but pervasive feeling that something is "not quite right." In "normal" conversation, hitches in the sequencing of turns are routinely seen and treated as evidence for trouble in the interaction and are remedied by repairs of various sorts. The disturbances in turntaking generated by transmission delays are particularly insidious because they raise no easily identifiable trouble flags.

Data and Methods

The Data

Our data are derived from a broader ethnographic study of a now-defunct holding company which used cutting-edge groupware and communication technologies to support a distributed work environment (Ruhleder and Jordan, 1997; Ruhleder, et al., 1996). Company headquarters managed several business units distributed across the United States. Lotus Notes® and remote meeting technologies such as LiveBoard® and PictureTel® were widely disseminated in order to facilitate interactions between headquarters and these geographically distributed holdings.

We carried out fieldwork over a period of four months, collecting data at headquarters and several business units through participant observation, unstructured interviews, and review of on-line and paper materials. The resulting data set includes a number of video tapes capturing both sides of various small- and large-group remote interactions using both audio-plus-shared workspace and audio-plus-video technologies, from which we derive the data for this paper.

In this paper, we draw on a 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 was organized as a preliminary discussion for a software prototype demonstration set up for the following day. Ann is one of the three accountants, and Bill is one of the three developers. Neither the genders nor the names necessarily correspond to those of the original participants, and the excerpts we are presenting were all drawn from interactions between the most senior members of each group. Neither the developers nor the accountants have had extensive experience with video-conferencing. Though all three accountants and the most senior member of the developers' group have vicarious knowledge of the technology through attendance at large, formal video-conference meetings, this is only the second meeting which they themselves initiated.

Our analysis in this paper focuses on the audio component of the transmission. 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 the 19 minute interaction which exhibited these characteristics and in which these characteristics were identified by multiple analysts without the aid of any special technological manipulation or assistance (slowing the sounds down, using a metronome, etc.).

Methods: Interaction Analysis

If turntaking is foundational to ordered human interaction and forms the basis for social life, then people or groups having two fundamentally different conversations with each other raises serious questions about what it means to "share" a conversation in a distributed setting. It challenges our notion of a mutually-constructed reality.

In order to examine this phenomenon, we analyzed videotapes of participants' activities on both sides of a remote communications link using video-based Interaction Analysis, 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, Interaction Analysis 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. However, 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, and can serve as one source of insight for designers and practitioners (Button and Dourish, 1996).

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 and furnishes a background against which the video taping is carried out. Video tape content logs provide an overview of the data corpus for locating sequences for further analysis.

Collaborative tape analysis is carried out within a multi-disciplinary research group. 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 according to emergent categories This, in turn, identifies new sequences for analysis, or suggests new venues for video taping. The application of Interaction Analysis to this particular project is outlined in more detail in (Ruhleder and Jordan, 1997).

Turntaking and Technology

Turning now to the analysis of videotaped remote interaction, we note that according to the rules of conversational interaction mentioned above, a current speaker may pass along a turn, or simply indicate that they are done with their turn, at which point another person present can take a turn. We are interested in what happens when turntaking is disrupted because of a delay in transmission. Consider the *hypothetical* conversation in Figure 1, in which no transmission delay occurs.[1]

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?*	
	Ann hears: *Yes, thanks*	2 00		**Bill says:** Yes, thanks

Figure 1: 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). Bill has access to multiple cues that suggest that a response is called for, including not only the tone of the speaker's voice, but also the speaker's gaze, body orientation, and gesture. Ann asks a question and Bill offers a response that is prompt and appropriate to the

[1] This fragment is derived from a video-mediated interaction We use a hypothetical example to illustrate how the same utterance may be experienced differently under different circumstances

situation. It may be worth noting here that the speed with which an answer is delivered is implicative. Face-to-face and on the phone a negative response to a question is routinely foreshadowed by a brief pause, often followed by "uh," "well..." or "yes, but...," or something of that sort. In other words, a pause in this spot potentially indicates an upcoming disagreement.

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 differences 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		
	(Ann hears silence)	2 00	***Bill hears:*** *Did you get the report I sent you?*	
	(Ann hears silence)	3 00		**Bill says:** Yes, thanks
	Ann hears: *Yes, thanks*	4 00		

Figure 2: 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. Additionally, due to the generally low quality of video as implemented in today's teleconferencing systems, both parties have only limited access to interactional cues such as a speaker's gaze, body orientation, and gesture. 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.

If from Bill's perspective, this particular exchange is essentially the same as above— Bill heard Ann's question and responded— 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 signal hesitancy or doubt, subconsciously triggering a set of questions. Why isn't Bill responding? Does he know what report she is talking about? Is he trying to remember whether the report was in yesterday's mail? Is he angry about the contents of the report and doesn't know how to say this? Within this turntaking system, waiting for the response to a simple question raises a trouble flag in the interaction, but in this setting, it would be a flag operational only for one party.

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. Depending on her choice, Bill might have come to believe that she is not all that competent or that she is uninterested in the issue on the table.

Some of Ann's choices would have led to crossovers and collisions in terms of her conversation with Bill. In all cases, however, what happened on her side of the link would have been *different* from what happens on the Bill's side in terms of the words they heard and the conversation in which they think they are participating.

Below, we look at these crossovers and collisions and the meanings they may generate on each side of the link from the perspective of the rules that govern turntaking in conversation. We do that by examining the data described above, drawing from 32 episodes in which we see the rules of normal turntaking undermined, and in which cues used to follow the rules are not accessible as they are in face-to-face settings. Our examples constitute discrete, identifiable instances in which the transmission delay inherent in the technology did not allow for the successful application of the conversational rules outlined in (Sacks, et al., 1974).

Collisions and Swaps

Unless the previous speaker selected a specific next speaker at the end of a turn (e.g., "How about you, Gitti?"), another person can start or the first speaker can resume (Sachs, et al., 1974:704). In settings where the audio signal is significantly delayed and visual cues are not helpful, the effectiveness of this turntaking system breaks down in subtle ways for the simple reason that what one side hears and responds to is not what the other side hears and responds to. In Figure 3, we see what happens when speakers are unable to perceive the cues for applying the rules.

what Ann says	what Ann hears	time in sec	what Bill hears	what Bill says
Ann says: Will Omega Group join us?		1 00		
		2 00	***Bill hears:*** *Will Omega Group join us?*	
Ann says: They're not being involved?		3 00		**Bill says:** No
	Ann hears: *No*	4 00	***Bill hears:*** *They're not being involved?*	
Ann says: OK		5 00		
		6 00	***Bill hears:*** *OK*	

Figure 3: "Will Omega Group join us?"

In this example, Ann and Bill are talking about an upcoming meeting. They are in the process of working out which groups will take part in the discussion. Note

that in the excerpt, Bill's answer, "no," to Ann's question, "Will Omega group join us?," is heard by Ann as the answer to her *second* question, "they're not being involved?"

Let's consider this exchange from Ann's perspective Ann, in asking, "Will Omega Group join us?" has completed a turn, designating Bill as the next speaker. She waits for his answer, but receives no indication that he has heard the question or that he needs to think about the answer ("Gee, I might have to check with so-and-so on that."). After a second has passed, Ann rephrases the question, "They're not involved?" She now hears what seems to be a prompt answer to her question, "No," which she acknowledges with equal promptness, "OK."

Let's consider this now from Bill's perspective. Bill hears the question and answers it promptly, "No." With this answer, Bill has completed a turn and, according to the rules for turntaking, it is quite appropriate for Ann to take the next turn with a response or another question. But Ann's response is not meaningful in context. Bill has already told her that Omega group is not involved. Her "OK" is too late and out of sequence if it is to serve as an acknowledgment of his answer to the original question.

On each side, the rules for conversation and the orderly interchange of words have been met, yet they have not facilitated the orderly creation of shared meaning. On both sides an awkward moment arises: Ann repeats a question she thinks Bill hadn't addressed, leading Bill to hear a question he thought he had already answered. Without an understanding of the delay phenomenon, and working in real-time, each instance could be interpreted negatively. Delayed or inappropriate responses could indicate the other party is not paying attention ("is Bill listening?"), doesn't have necessary information ("but he works with Omega group"), or challenges someone's statement ("I already told Ann 'no'").

It would be difficult if not impossible for individuals on either side of the link to identify the source of trouble— the technology delay— within their real-time experience, as the other's conversation appears to be occurring naturally and spontaneously. Multiple instances of this kind open up the possibility of rising confusion or discomfort across the link.

Unnecessary Rephrasings

The rules for turntaking imply that when one person is done, another person may pick up the conversational ball. If a speaker comes to the end of a turn without indicating who should go next, and finds that there are no self-selecting next speakers, he or she may resume with several options. One of these options is to rephrase the previous utterance. This rephrasing might elaborate on part of the previous utterance, and may also serve to specifically select or narrow down the choice of next speaker.

We have identified 12 instances in our 32 episodes where this orderly progression is disturbed in some way. The cases fall into one of two categories. In

some instances, the speaker on one side of the link had selected the next speaker, but restated or rephrased their question when they perceived a lack of response due to the delay (see Figure 3, above). On the other side of the link, however, the speaker had responded appropriately. In other cases, the speaker on one side of the link acted in accordance with a perception that no self-selection of next speaker had taken place. On the other side of the link, however, the next speaker had self-selected. The result in both cases was somewhere between confusion and collision.

Figure 4 illustrates one form of collision resulting from unnecessary rephrasing. In this excerpt, Bill is just finishing up an answer to a previous question about the status of a particular component of the software being developed.

What Ann Says	What Ann Hears	time in .5 sec	What Bill Hears	What Bill says
		1 00		**Bill** says through the report screens
		1 50	(Bill hears silence)	
	Ann hears *through the report screens*	2 00		**Bill says** Maybe I
Ann says So you're		2 50		**Bill says** didn't communicate
	Ann hears. *Maybe I didn't communicate*	3 00	***Bill hears*** *So you're*	**Bill says** that clearly
	Ann hears *that clearly*	3 50		
Ann says Yeah		4 00		
Ann says You're farther along		4 50	***Bill hears*** *Yeah*	
Ann says than we expected		5 00	***Bill hears*** *You're farther along*	

Figure 4: Communicating Clearly?

Ann hears Bill finish an explanation about the report screens. He has completed a turn, and she attempts to take the next turn. Just as she begins to speak, she hears Bill break in with a statement about not having communicated the status clearly. She responds to him, "Yeah," 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. He adds a statement about half a second later, suggesting that his explanation wasn't clear enough, thereby downgrading his own competence ever so slightly. Towards the end, Bill hears the start of Ann's next turn overlap with the completion of his own. Then there is a second's worth of silence, an affirmative, and Ann's continuation.

Note what happens here and how this interaction may be interpreted as a lack of competence on multiple levels. From a turntaking perspective, both speakers have

behaved inappropriately, Bill cutting Ann off at 3.00 (from her perspective), and Ann not waiting until Bill is finished at 3.00 (from his perspective).

The ordering of utterances and duration of silences here also causes problems in terms of each side's ability to judge the competence or understanding of the other. Bill's comment at 2.00 about not having communicated clearly, which comes in reaction to a lack of response from Ann, implies that he thinks she didn't understand his explanation. But she did, and her attempt at introducing a follow-on to his explanation at 2.50 shows that she has understood it and knows what it means within the context of the project. That acknowledgment, which would have signaled successful communication, gets cut off by Bill's comment about a lack of communication. After her "yeah" she continues her original statement about how far along the project is.

In these small ways, competence is repeatedly called into question on multiple levels. People appear to behave incompetently as turn takers in a conversation, even as they themselves work hard to play by the rules. And people make statements that appear inappropriate within their local context because that context is not shared by both parties. What is the impact of these kinds of situations where trouble is apparent to only one side of the link? And what happens when multiple exchanges pile trouble upon trouble on both sides of the link?

Misapplied Feedback

In face-to-face conversation, participants who do not currently have the floor communicate with the speaker through various types of feedback (c.f., Jefferson, 1984; Yngve, 1970; but esp. Schegloff, 1981). Feedback to the speaker can provide acknowledgment or confirm understanding on the listener's part through expressions such as "mm, hm" and "yeah"; sentence completion; requests for clarification ("such as?"); brief restatements ("in Omega" following "we'll ask everyone in Omega group"); and head nods and shakes. They offer an opportunity for shared meaning-making, giving both speaker and audience confidence that they are on the same wavelength (Jordan and Fuller, 1975).

The sequencing of these kinds of messages is systematically related to signals from the speaker They occur when the speaker has been looking away and turns towards the listener or completes a grammatical clause. They may also be a response to the speaker's elicitation of confirmation or acknowledgment (Schegloff, 1981). At these points, it is appropriate for the listener to signal that they are still paying attention and that they continue to understand (c.f., Goodwin, 1981; Kendon, 1967).

In video-supported remote conversation, the medium may fail to support these mechanisms (Dourish and Bellotti, 1992; Sellen and Harper, 1997). In fact, feedback may actually disrupt the discourse it was intended to support because participants provide feedback or indicate understanding at what is an appropriate point from their point of view, but is heard at a moment when it does not produce

the intended meaning. Figure 5 illustrates how intended supportive feedback can induce a noticeable, unintended pause or hesitation in the speaker's utterance.

Ann's perspective	
Bill:	we could all sit around the table and look at it But in the future, I think that's (pause) why
Ann:	I see
Bill's perspective	
Bill:	we could all sit around the table and look at it But in the future, I think that's (pause) why
Ann:	I see

Figure 5: "But in the future..."

Ann is listening to Bill's plan for the demo. Her comment, "I see," is spoken directly after Bill has delivered a key point, appropriately indicating comprehension on her part. Ann hears Bill continue, then pause for no reason apparent to her before going on. Bill hears the dialogue differently. On his side of the link, Ann's, "I see," is heard 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, it comes in the middle of a new utterance.

From Ann's perspective, she has behaved helpfully by confirming an understanding of Bill's arrangements for an upcoming meeting. She has no access to the reason why Bill might hesitate or pause in the middle of a sentence. From his perspective, however, her utterance formed an interruption. It broke his stride, leading him to pause a moment before continuing.

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:

- some kind of trouble arises,
- this trouble disrupts the turntaking system, and
- the trouble source *cannot be identified* by participants.

Participants may sense that something might be "wrong"— in several examples above, for instance, one person appears to violate the rules of the turntaking system by interrupting or saying something contextually inappropriate. The nature of the distributed technology, however, may preclude 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

meaning construction *in real-time* in the course of producing talk over the link as part of a broader set of work activities.

The potential consequence is a pervasive sense of uneasiness similar to that described in Jordan and Fuller (1975) where non-native 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 the trouble. We find significant potential for troublesome miscommunication and interpretation in our own data. Here, too, people were unable to identify and repair trouble as it occurred because its origin was obscured.

The point here is not that people break conversational rules Rather, it is that by seeming to behave in ways that violate the expectations of their conversational partner, multiple explanations for those violations may be generated. Explanations may be generated for seemingly aberrant behavior. In conversations repeatedly punctuated by episodes such as the ones above, these explanations may be overwhelmingly negative when judged within the cultural expectations for turntaking in face-to-face settings.

Implications for Design and Training

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 delay on the workings of a turntaking system developed within a co-located, face-to-face context. We would like to consider this impact in the following terms:

What people can learn to 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, they will become more facile at adapting their manner of communication to the medium.[2] 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 class of problems, those where the trouble that is generated by the delay is apparent to only one side *or to neither side* of the interaction. This kind of trouble may be recognizable only through retrospective analysis, while generating

---

[2] Early research in conversation analysis includes the study 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)

real-time reactions among participants. As such, people may need to learn to develop a kind of "meta narrative" which analyzes and challenges negative assumptions about one's conversational partners, in the way that people must occasionally remind themselves that the abrupt or angry tone of an email message may be a product of the medium, and not the intention of the other party.

How these findings can guide meeting strategies. If certain forms of conversational patterns are difficult to sustain in interactions characterized by delay, then the use of delay-generating technologies can be focused on those interactions in which trouble is less likely to arise. For instance, 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. Broader participation in these events through the use of video-conferencing technologies can additionally help draw junior members of the organization more deeply into a distributed community of practice (Ruhleder, et al., 1996; Ruhleder, forthcoming). Brainstorming or design sessions, however, may be more feasible using non-delayed audio-only channels and some form of shared electronic workspace. By understanding the impact of particular technologies on real-time interactions, users and implementors can more effectively chose an appropriate technology set.

How these findings contribute to development efforts. Current development efforts are concerned with ways of indicating presence and attention in remote meetings, such as through gaze. 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. These development efforts will 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. This may contributed 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).

The increased integration of audio-only and video-plus-audio channels into applications for collaborative work also requires the development of technological components and workplace strategies for determining awareness and calling attention to specific activities and elements in the shared workspace. The content of a conversation can be used to maintain awareness (i.e., Dourish and Bellotti, 1992; Hindmarsh, et al., 1998; Hindus, et al., 1996); more importantly from our perspective, the *structure* of an utterance can be used to generate attention (Goodwin 1980). Delays disrupt this fundamental function. While we do not explicitly address this issue in our current paper, our data suggest this as an important consideration in further analyses.

Finally, we suggest that the turntaking system itself be considered with respect to the development and implementation of new communication technologies. In

normal, 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 which 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. Further research and development in technology-mediated communication must account for the ways in which affordances interact with and affect the turntaking system.

Acknowledgments

Our work has benefited immensely from collaborative working sessions with Manny Schegloff. Valuable input was provided by participants in Xerox Palo Alto Research Center and University of Illinois Interaction Analysis Laboratories.

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

References

Button, G., and Dourish, P. (1996): "Technomethodology: Paradoxes and Possibilities," *Proceedings of the Conference on Human Factors in Computing Systems,* pp. 19–26.

Dourish, P., Adler, A., Bellotti, V. and Henderson, A. (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, P., and Bellotti, V. (1992). "Awareness and Coordination in Shared Workspaces," *Proceedings of the Conference on Computer Supported Cooperative Work*, pp. 107-114

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

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

Heath, C. and Luff, P. (1993): "Disembodied conduct: Interactional asymmetries in video-mediated communication," in Button, G., *Technology in Working Order: Studies of Work, Interaction, and Technology,* Routledge, London, pp. 35–54.

Isaacs, E. A., Morris, T., Rodriguez, T. K., and Tang, J. C. (1995): "Comparison of Face-to-face and Distributed Presentations," *Proceedings of the Conference on Human Factors in Computing Systems,* pp. 354–361.

Jefferson, G. (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.

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

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

O'Conaill, B., Whittaker, S., and Wilbur, S. (1993): "Conversations over video conferences: An evaluation of the spoken aspects of video-mediated communication," *Human-Computer Interaction,* vol. 8, pp. 389–428.

Olson, J S., Olson, G. M. and Meader D. K. (1995): "What Mix of Video and Audio is Useful for Small Groups Doing Remote Real-time Design Work?" *Proceedings of the Conference on Human Factors in Computing Systems,* pp. 362–367.

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

Ruhleder, K., and Jordan, B. (1997): "Capturing Complex, Distributed Activities." in Lee, A. S., Liebenau, J. and De Gross, J. I. *Information Systems and Qualitative Research,* Chapman and Hall, London, UK, pp. 246–257.

Ruhleder, K., Jordan, B., and Elmes, M. (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, H., Schegloff, E. and Jefferson, G. (1974): "A Simplest Systematics for the Organization of Turntaking for Conversation," *Language,* vol. 50, pp. 696–735.

Schegloff, E , Jefferson, G., and Sacks, H. (1977): The preference for self-correction in the organization of repair in conversation.," *Language,* vol. 53, pp. 361-382.

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

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

Schegloff, E. (1981). "Discourse as an Interactional Achievement: Some Uses of 'Uh huh' and Other Things that Come Between Sentences," in Tannen, D., *Analyzing Discourse: Text and Talk,* Georgetown University Roundtable on Languages and Linguistics, pp. 71–93.

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

Sellen, A. J., and Harper, R. (1997): "Video in Support of Organisational Talk," in Finn, K., Sellen, A. and Wilbur, S., *Video-Mediated Communication.* Lawrence Erlbaum Associates, Mahwah, New Jersey, pp. 225–244.

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

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

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

Whittaker, S. (1995): "Rethinking video as a technology for interpersonal communications: theory and design implications." *International Journal of Human-Computer Studies,* vol. 42, no. 5, pp. 501–29.

Yngve, V. (1970) On Getting a Word in Edgewise. *Papers of the 6th Regional Meeting of the Linguistic Society of Chicago;* pp. 564–577.