Communicative breakdown in copresent and technologically-mediated interaction

Eckehard Doerry, Sarah A. Douglas
Department of Computer and Information Science, University of Oregon

ABSTRACT: This paper presents a comparative analysis of communicative breakdowns experienced by participants collaboratively performing a task in three communication environments: face-to-face, mediated by an audio connection, and mediated by both an audio and a video connection. A quantitative analysis showed significantly less breakdown in the copresent condition than in either of the two technologically-mediated conditions; no differences were found between audio-only and audio-video conditions. Subsequent qualitative analysis revealed that breakdowns in the audio-video environment stemmed from pragmatic deficiencies in the access to nonverbal displays afforded by a remote video image in task-oriented scenarios.

INTRODUCTION

In an ideal world, all collaborative interactions would take place between participants who are face-to-face; the copresent condition allows participants to draw directly on a lifetime of communicative experience to organize their interaction. Unfortunately, the material and geographic constraints of the modern world make personalized interactions of this sort increasingly unlikely. Work groups may be distributed across widely separated subsidiaries of a large organization and may run into the tens or hundreds of participants.

Accordingly, an area within CSCW that has received much attention in recent years is exploring ways in which computer-based technologies can support the collaborative interactions of users who are geographically distributed. Increasingly powerful systems for desktop conferencing, group meeting, and distributed design (Root, 1988; Harrison and Minneman, 1990; Abel, 1993; Okada et al., 1994) promise to fundamentally change the way members of modern society interact with each other, both casually and in formal business contexts. Though the technologies applied in such systems vary widely, the underlying goal of all such systems is essentially the same: to provide a simulacrum of copresence that is functionally equivalent to face-to-face interaction, allowing users to accomplish their communicative and creative goals as easily and efficiently as if they were physically copresent. In other words, the goal of technologically-mediated environments is to support the same *communicative efficacy* as face-to-face interaction.

In recent years, there has been increasing interest in evaluating the extent to which current technologies achieve this goal by empirically comparing copresent to technologically-mediated interaction using metrics like user satisfaction, quality of work produced, or task-solution activities engaged in by participants (Whittaker et al., 1991; Apperley and Masoodian, 1995; Dykstra-Erickson et al., 1995; Isaacs et al., 1995; Olson et al., 1995; O'Malley et al., 1996) . Though all of these approaches provide a basis for asserting that interactions in one environment have a higher communicative efficacy than in another, they yield few insights as to *why* differences in efficacy exist. For instance, user satisfaction surveys can tell us that users prefer one communication environment over another, but do not reveal the communicative difficulties experienced by participants in a "less satisfying" environment that are presumably the root cause of their dissatisfaction. This limitation arises from the fact that metrics like user satisfaction, quality of work, and task-activity structure characterize the communicative efficacy of interactions indirectly, inferring the amount of communicative difficulty experienced by participants from the overall outcomes or structure of interactions.

Drawing on analytic techniques from several disciplines, the study presented in this paper compares the communicative efficacy of three different communication environments by painstakingly dissecting interactions to expose the low level communicative *breakdowns* experienced by participants. An important advantage of this approach is that it yields a concise characterization of the communicative troubles encountered by participants in a each environment, which supports focused investigation of *why* breakdowns occurred. By articulating causal relationships between the physical characteristics of an environment and the communicative troubles experienced by users, the analysis establishes a solid basis for future redesign.

Method

Using a between-subjects design, the study examined communicative breakdowns that occurred in three different communication environments: copresent, audio-only, and audio-video, with four pairs of participants in each condition. Interactions were task-oriented, with paired participants asked to collaboratively manipulate a cardiovascular simulator to answer a series of questions about the system's physiological behavior. Participants had no previous experience with advanced technologically-mediated environments.

Communication Environments

In the copresent condition, participants were seated side-by-side in front of a computer workstation and were able to communicate freely as they worked. In the audio-only condition, participants were seated in separate rooms and fitted with lapel microphones and a headset to provide a full-duplex, high-fidelity audio connection.

Shared access to the simulator workspace was provided by splitting the video output from the computer workstation on which the simulation was running to computer displays placed in front of each participant. Arrangements for the audio-video condition were identical to the audio-only condition, except that each participant was also provided with a large 27” color monitor to establish a video connection between participants. The monitor was placed adjacent to each participant’s workstation display, separated by approximately 45 degrees of angle. The camera capturing the remote image was placed within this angle and as close to the remote monitor as possible so that, when participants turned to direct their gaze at the remote image, it appeared to the remote observer that their partner was turning towards them. The remote image was framed to capture a participant’s upper body, including the tabletop and any writing or mouse manipulation that occurred there.

Participants in all three conditions were each provided with their own mouse, empowering them to manipulate objects within the shared workspace. However, both mice controlled a single cursor within the shared workspace.

All of the connections between participants in the distributed scenarios, including the audio channel, the video channel, and the shared workspace, were implemented using analog technology, avoiding bandwidth limitations and latency problems often associated with packet-based networked implementations.

Subjects

Participants were recruited from university biology classes at all levels in the curriculum. The focus on biology students reflects an effort to locate participants who were naturally motivated to explore the task domain and would exhibit more vigorous and robust interactions. To further stimulate interaction, subjects were asked to sign up in pairs with a friend or lab partner.

Potential participants were asked to fill out a questionnaire and were screened to eliminate those with previous experience using interactive telecommunication technologies, including interactive text (e.g. MUDS, IRC), conference calling, and videoconferencing. None of the participants had ever used the cardiovascular simulator used in this experiment.

Seven female-female pairs, three male-male pairs, and two male-female pairs were selected and randomly assigned to the three communicative environments described earlier. Participants were paid $5 each for their participation.

Task

Participants were asked to use a cardiovascular simulator developed by the authors, called the Cardiovascular Construction Kit (CVCK) (Douglas and Doerry, 1994) to explore the physiological behavior of a simple cardiovascular system. To accomplish the task, participants had to first piece together a simple cardiovascular loop depicted in a printed “laboratory manual” given to each participant, attach gauges to measure blood pressure and flow at various places, and then run the simulation to answer a series of questions about the dynamic behavior of the system. Figure 1 shows the CVCK workspace with the completed construction and attached gauges.

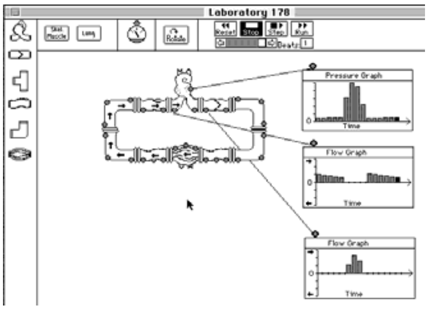

Figure 1: The simulator window showing the cardiovascular construct explored by participants, complete with attached graphs.

To ensure that there would be plenty of discussion, the laboratory manual was designed to provide only an abstract specification of the task, rather than giving detailed instructions. For instance, participants were simply asked to “attach pressure and flow gauges to the points shown in the diagram below”, without any indication of which components were gauges and how one might go about attaching them.

Procedure

Pairs of participants were scheduled for a single one hour session. After filling out consent forms, participants were taken to their workstations and briefly introduced to the technology. No training in the use of the simulation was provided. Participants were informed that the focus of the experiment was on their interaction and that their answers to the questions posed in the laboratory manual would not be graded. No time limit was set for completing the assigned task; participants were instructed to take their time and simply do their best.

In total, 12 pairs of students participated in the experiment, with four sessions recorded for each of the three communication environments. Images of the workspace and oblique upper body views of each participant were recorded onto a single videotape using a video processor to provide a synchronized and complete record of each session; audio was recorded on a separate audio track for each participant. In addition, the sessions were remotely monitored as they were recorded, and extensive field notes taken to provide an overview of events and structure of the interactions.

Analysis

To compare the amount of communicative breakdown that occurred in each of the three environments, the videotaped interactions were analyzed using a three phase analytic process we call Breakdown Analysis. Breakdown Analysis can be seen as a direct extension of the tools and techniques developed by conversation and interaction analysts for documenting the conversational regularities that interacting participants rely on to construct shared interpretations of each other's verbal and nonverbal communicative displays (Suchman, 1987; Jordan and Henderson, 1995) . More generally, Breakdown Analysis can be seen as a form of Exploratory Sequential Data Analysis (Sanderson and Fisher, 1994) .

Breakdown Analysis consists of three intertwined qualitative and quantitative studies that progressively refine our understanding of the communicative troubles experienced by participants in different environments, and of the way in which such breakdowns are related to the physical characteristics of the environments in which they occur. In the first study, interactions were qualitatively examined to identify consistent patterns of breakdown. The second study applied nonparametric statistics to expose significant differences in the incidence of breakdown between environments. The results of this statistical analysis were used to focus the third and final study, in which individual episodes of breakdown were examined in an effort to rationalize the differences found. The following sections describe each study in more detail.

Study #1: Patterns of Breakdown

The initial qualitative study established the comparative framework for the Breakdown Analysis by applying the techniques of Interaction Analysis (Jordan and Henderson, 1995) to identify distinct categories of communicative breakdown that occurred during interactions. The analysis was structured by recognizing three fundamental conversational tasks[1] that must be continuously addressed by participants in any interaction in order to maintain mutual intelligibility:

1) **Turntaking.** Participants must organize their contributions to the interaction. For example, participants must regulate access to the verbal floor in order to avoid overlapping talk.
2) **Topic.** The notion of topic constitutes a fundamental organizational mechanism, establishing a basis for defining the notion of "progress" in conversation.
3) **Reference.** As participants in any interaction converse, they must continually match references that appear in a partner's utterance with entities that exist in the referential context. For example, in the utterance "Let's move the pump thing over to the side", there are at least two references which must be disambiguated in order to construct the meaning of the utterance: which object is the referent of "the pump", and what spatial position is meant by "the side".

---

[1] This framework draws on the extensive literature in Conversation Analysis and linguistics where each of these tasks has generated considerable interest (e.g. Sacks et al., 1974; Clark and Wilkes-Gibbs, 1986) .

These three fundamental organizational activities provide a framework for understanding what it means for "breakdown" to occur in interaction; a breakdown occurs when participants' organizational efforts fail, resulting in divergent conceptions of whose turn it is to contribute to the conversation, what the current topic is, or what object or entity is the referent of an immediately preceding utterance.

Using this framework, all of the sessions were examined to identify specific categories of communicative breakdown experienced by participants. A specialized notational schema was developed and all of the sessions were transcribed in their entirety to provide a stable, textual record totaling 343 pages for the 12 sessions; examples of this transcription are presented later in this paper. Finally, these transcripts were iteratively analyzed to refine the categories of breakdown identified earlier, developing objective, observable criteria for recognizing breakdowns in each category. In total, approximately 145 hours of analysis were required for each hour of videotaped interaction.

The analysis revealed four major categories of breakdown experienced by participants:

- **Verbal turntaking breakdowns** were defined by the failure to regulate access to the verbal channel, resulting in confusion over whose turn it was to speak. In general, episodes of Verbal turntaking breakdown were evidenced by the presence of overlapping talk. However, situations in which the overlapping speaker had ample and timely evidence that a partner was currently speaking were considered to be willful interruptions and were not counted as Verbal turntaking breakdowns.
- **Cursor turntaking breakdowns** were defined by participants' failure to regulate access to the shared cursor, resulting in confusion over whose turn it was to use the cursor to gesture or manipulate objects in the workspace. Episodes of Cursor turntaking breakdown were evidenced by simultaneous attempts to control the shared cursor, which resulted in readily apparent erratic, i.e. jerking, behavior of the shared cursor in the workspace.
- **Topic breakdowns** were defined by participants' failure to maintain shared topical orientations, resulting in a situation in which one participant believed the discussion had moved on to some next topic, while the other still believed discussion to be focused on a previous topic. Topic breakdowns were primarily evidenced by verbal repair sequences in which a participant explicitly raised the issue of "what are you working on?", creating an opportunity for participants to resynchronize their topical orientations.
- **Reference breakdowns** were defined by the failure to establish shared reference to objects or entities in the workspace, causing either the speaker or the listener to become uncertain that a linguistic reference produced in an immediately preceding utterance had been understood by the listener. Reference breakdowns were evidenced by explicit verbal repair sequences, initiated either by the speaker or the listener, in which the referential confusion was made apparent to both participants and collaboratively resolved.

Though space does not permit more detailed discussion (but see Doerry, 1995), the criteria developed to operationalize each of these categories were very conservative. Though some amount of inference on the part of the analyst is inevitably required in any qualitative analysis, emphasis was placed on minimizing such inference by firmly grounding the criteria for recognizing breakdowns in objective, readily-observable features of interaction.

Study #2: Exposing Differences

In the second phase of Breakdown Analysis, all transcripts were repeatedly examined in their entirety, applying the evidentiary criteria developed earlier to expose all episodes of breakdown in each category that occurred in each session; the analysis was considered complete when, after four passes, no further breakdowns had been detected. The results are summarized in Table 1.

	Breakdowns			
Session	**Verbal**	**Cursor**	**Reference**	**Topic**
FF2	48	2	3	3
FF3	18	0	1	2
FF4	39	0	4	2
FF5	22	0	4	3
FF-Total	**127**	**2**	**12**	**10**
FF-Mean	31.8	0.5	3.0	2.5
FF-StDev	12.3	0.9	1.2	0.5
AO2	16	4	4	5
AO3	75	27	7	11
AO4	63	4	11	5
AO5	38	9	4	4
AO-Total	**192**	**44**	**26**	**25**
AO-Mean	48.0	11.0	6.5	6.3
AO-StDev	22.8	9.5	2.9	2.8
AV2	40	3	7	4
AV3	67	24	12	12
AV4	13	0	3	1
AV5	13	5	6	7
AV-Total	**133**	**32**	**28**	**24**
AV-Mean	33.3	8.0	7.0	6.0
Av-Stdev	22.4	9.4	3.2	4.1

Table 1: Total number of breakdowns in each session. FF= copresent (face-to-face); AO= audio-only; AV= audio-video.

The total number of breakdowns that occurred in the three environments was compared for each of the four categories using nonparametric statistical techniques. Independent variables were the three communicative conditions — copresent, audio-only, and audio-video; the four categories of breakdown were considered dependent variables. Pairwise Mann-Whitney U tests were performed ($p \leq 0.1$[†]), comparing the three environments for each category of breakdown. Results were as follows:

Copresent versus Audio-Only. The amount of breakdown in copresent interactions was significantly lower than in audio-only interactions for Cursor turntaking ($U=0$; $p \leq 0.014$), Reference ($U=2$; $p \leq 0.057$), and Topic ($U=0$; $p \leq 0.014$) breakdown; no significant difference was found for Verbal turntaking breakdown.

Copresent versus Audio-Video. The amount of breakdown in copresent interactions was significantly lower than in audio-video interactions for Cursor turntaking ($U=2.5$; $p \leq 0.064$) and Reference ($U=2.5$; $p \leq 0.064$) breakdown; no significant difference was found for Verbal turntaking or Topic breakdown.

Audio-only versus Audio-Video. No significant difference in the amount of communicative breakdown was found between audio-only and audio-video interactions in *any* of the four categories of breakdown.

In sum, copresent participants experienced significantly less breakdown in several categories than their distributed counterparts; in no category was the number of breakdowns in a technologically-mediated environment significantly lower than in the copresent environment. Based on these results, we conclude that the overall communicative efficacy of copresent interactions was higher than that of either audio-only or audio-video interactions. The absence of any significant differences between the audio-only and audio-video condition supports the conclusion that there was no difference in overall communicative efficacy between these two environments.

Study #3: Rationalizing Differences

For designers of technologically-mediated environments, knowing that the communicative efficacy supported by some technologically-mediated environment is inferior to copresent interaction is less important than understanding *why* this deficiency exists. Specifically, the results yielded by the quantitative analysis of breakdown raise two important questions:

1) How can the higher incidence of breakdown observed in the audio-only and audio-video environments be related to specific physical characteristics of these environments?
2) Why did the availability of a video channel not significantly reduce the amount of breakdown experienced in the audio-video environment?

As a basis for exploring these issues, we hypothesized that breakdowns in technologically-mediated interactions might be consistently associated with the use of certain kinds of verbal (e.g. utterance, prosodic effects, pace) and nonverbal (e.g. gesture, manipulation of the workspace) communicative displays, thus implying that

[†] While slightly higher than that used in traditional parametric analyses, this p-value is not unusual in nonparametric analyses of complex conversational phenomena that are not easily quantified (Siegel and Castellan, 1988) .

access to these displays was somehow constrained in the technologically-mediated environments.

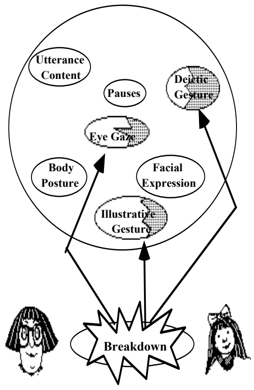

Figure 2: Communicative breakdown is more likely in environments in which access to verbal and nonverbal communicative displays is restricted.

As illustrated in Figure 2, maintaining shared understandings of an evolving conversation is based on the contextual interpretation of a partner's communicative displays (Garfinkel, 1967) . When access to these displays is restricted by the environment, this evidentiary process is crippled, leading to a higher overall likelihood of breakdown.

Each episode of Cursor turntaking, Reference, and Topic breakdown that occurred in a technologically-mediated interaction was re-examined to expose consistent patterns in the verbal and nonverbal communicative displays that participants were relying on to organize their interaction at the point at which the breakdown occurred. Copresent interactions were examined as well to establish that breakdowns did *not* occur when copresent participants relied on similar displays to inform their interactions.

The analysis revealed that Cursor turntaking, Reference and Topic breakdowns that occurred in audio-only and audio-video interactions were overwhelmingly associated with situations in which participants *were relying on nonverbal displays to organize their interaction*.

The segments of transcript shown on the following page present examples of Cursor turntaking and Reference breakdown that illustrate this insensitivity to nonverbal displays. An example of Topic breakdown is not given due to lack of space, but see (Doerry, 1995) .

The transcription notation is an amalgam of approaches used in existing interaction analytic work (Heath, 1986; Suchman, 1987) . Briefly, the notation consists of two columns which document, respectively, verbal behaviors and nonverbal behaviors; nonverbal events are indexed either by superscript integers, when a nonverbal event occurs during an utterance, or by integers in double-parentheses when a nonverbal event begins during a silence between utterances. A speaker's direction of gaze is denoted by the typeface used to transcribe the utterance: plain typeface denotes gaze at the simulator workspace, italics denotes gaze at the laboratory manual, and boldface indicates gaze at the conversational partner. For non-speakers, changes in direction of gaze are marked in the same way as other nonverbal events, by noting them in the Nonverbal column of the transcript. Abbreviations (e.g. H1, U2, C1) that appear in descriptions of nonverbal behavior refer to components in the simulator workspace. R and M refer, respectively, to the participant seated in the "main" and "remote" rooms in our laboratory. Finally, a black dot in the small leftmost column is use to indicate the point at which a breakdown occurs.

	VERBAL	NONVERBAL
•	((1)) R: you wanna do tho:se ones? (.5) M: *I think we haveta have* it hooked [2](.5) [3]yaknow ((4)) M: awwwooh+ -wu+ -- let go of it R: [5]ooh -- there	1- M releases mouse to scratch her head, R rolls cursor over to L in pallete (.7) 2- R gazes to lab manual 3- M grabs mouse and jerks cursor over near H1 4- R raises gaze to workspace and cursor wobbles and jerks across construction as both control mouse (2.5) 5- R jerks his hand off mouse

Segment: AV3p6

	((5)) M: What [6]about the[7] second thing down	5- R gazes lab manual while M still gazes workspace (2.3) 6- M points to workspace 7- R gazes workspace and grabs mouse
•	R: (um that) [1]this one? M: yeah:: ((2))	1- R rolls cursor down to V in pallete 2- R rolls to HELP menu and does a "describe" on V; description dialog pops up; they read it (5.0) then R nods and gazes lab manual (1.0)

Segment: AV5p3

•	R: how do we get rid of[2] this? ((3)) M: try the[4]::: ((5)) R: tr-hi sump'mm [M: sump'mm (.7) R: hummm[6]mmm ((7))	2- R points and clicks on G2 graph with cursor; G2 highlights 3- R rolls cursor back and forth, M glances towards the table (1.7) 4- M grabs his mouse 5- R rolls mouse some more, then pulls back hand and shrugs as he speaks (1.2) 6- M clicks on G2 7- M hesitates, then drags G2 over to the biowaste; biowaste highlights as it is contacted. (4.2)

Segment: FF4p14

The exchange shown in segment AV3p6* exemplifies a common pattern of Cursor turntaking breakdown, in which control of the cursor is initially negotiated verbally, followed by a Cursor turntaking breakdown when participants rely on nonverbal displays to tacitly negotiate a subsequent transition in cursor control. As segment AV3p6 begins, R verbally offers control of the cursor to her partner, opening an

* Segment names denote the session from which they were drawn. FF= copresent (face-to-face) environment; AO= audio-only environment; AV= audio-video environment. Remaining characters denote the session number and the page on which the exchange appears.

explicit negotiation over who should control the cursor. However, when M fails to verbally respond to this overture, R assumes that he retains control of the cursor. Cursor turntaking breakdown results as both participants try to move the cursor. It is important to point out that the Cursor turntaking breakdown occurred *in spite of* compelling nonverbal displays of cursor control produced by M. Specifically, R apparently fails to notice that M has tacitly accepted his preceding verbal offer of control over the mouse by moving her hand to the mouse.

This insensitivity to a partner's nonverbal displays was implicated in Reference breakdowns as well, as illustrated in Segment AV5p3. Here, M uses finger pointing to identify her referent as she suggests the next component to install in the construction the pair are piecing together. The Reference breakdown becomes apparent as R initiates a repair to clarify M's reference to "the second thing down". Note that R never directs her gaze at the remote monitor, gazing instead at the laboratory manual and the workspace as M gestures. That is, R appears to be totally *unaware* of the deictic gesture that M is making available.

The breakdowns presented in these two exchanges each occur in situations in which participants were relying primarily on nonverbal displays like deictic gesture (pointing), direction of gaze, and hand position to organize their interaction; participants in technologically-mediated interactions appeared to be profoundly insensitive to such displays.

By contrast, an examination of copresent interactions showed that participants were intimately aware of their partner's nonverbal behaviors and were able to effectively access these displays to inform their interaction. For example, consider the exchange shown in segment FF4p14.

From a strictly verbal perspective, M's aborted utterance "Try the:::" appears to be a request for R to perform some action. The fact that M's utterance actually represents an implicit request for control of the cursor is only apparent in light of M's movement of her hand to her mouse. The subsequent transfer of control over the shared cursor is then progressively negotiated entirely by nonverbal means, with both participants clearly orienting to each other's nonverbal displays to inform the transaction. Specifically, M does *not* immediately move the cursor after issuing her tacit request for control of the cursor, demonstrating her awareness that R is still using the cursor and is not yet ready to give up control. Only after R tacitly acknowledges M's request for control by removing his hand from his mouse does M actively assume control of the shared cursor. This behavior demonstrates how copresent participants were able to rely on nonverbal displays, not only to infer a partner's beliefs about who currently controls the cursor, but also as "feedback" during tacit negotiation of cursor control.

In sum, the qualitative analysis of breakdowns that occurred in audio-only and audio-video interactions revealed that breakdowns were related to the apparent insensitivity of participants to each other's nonverbal displays. This strongly implies that access to these displays was somehow restricted in these environments.

Explaining how this constraint arises from the design of the audio-only environment is trivial: since no visual connection between participants was provided, the nonverbal displays of a partner were fundamentally inaccessible.

It is less clear why participants in audio-video interactions were insensitive to their partner's nonverbal displays — nonverbal behaviors like finger pointing, movement of the hand towards that mouse, and direction of gaze were all readily discernible in the remote video image available to each participant. Why, then, were participants in audio-video interactions unable to access these communicative displays to more effectively organize their interaction?

To explore this question, a further analysis investigated the way in which participants in audio-video interactions used the remote video connection.

As a way of characterizing the extent to which participants used the remote video image in audio-video interactions, we counted the total number of times that a participant directed gaze towards the monitor displaying the remote video image. Copresent interactions were also examined, noting the total number of times that participants directed gaze directly at a partner, i.e. turned to look directly at the person seated next to them. The results of this analysis are presented in Table 2.

Audio-Video Environment	AV2	AV3	AV4	AV5
# gazes at remote video	16	86	9	23
Copresent Environment	**FF2**	**FF3**	**FF4**	**FF5**
# of gazes at partner	16	20	9	2

Table 2: Number of gazes at partner in audio-video and copresent interactions.

With one exception (AV3), participants in both audio-video and copresent interactions turned to gaze directly at their partner relatively infrequently, devoting almost all of their attention to the laboratory manual and the workspace. A Mann-Whitney U test showed that there was no significant difference between copresent and audio-video conditions.

Juxtaposed with the results of the earlier analysis, this observation suggests that the access to nonverbal displays provided by a video image is fundamentally unlike copresent access to such displays. Despite the fact that both copresent and distributed participants rarely gazed directly at their partner, copresent participants exhibited a keen sensitivity to nonverbal displays like direction of gaze, hand position, and manipulation of the laboratory manual while distributed participants appeared to be generally insensitive to these same displays. Clearly, copresent participants were able to rely on *peripheral* perceptual mechanisms to maintain an awareness of their partner's nonverbal behaviors. The remote video image did not appear to afford the same kind of peripheral access to distributed participants; accessing nonverbal displays available in a video image apparently requires participants to focus attention explicitly on that image.

Though this insight may explain the insensitivity to nonverbal displays in audio-viode interactions, it raises an obvious question: Why did participants not compensate for the inability to peripherally access the nonverbal displays available in the remote video image by simply directing their gaze at the remote image more often? An analysis of the occasions on which participants in audio-video interactions *did* direct their gaze at the remote video image revealed two prominent difficulties:

Video Schizophrenia. The need to attend explicitly to the remote video display to access a partner's nonverbal displays creates a classic competition for attention situation in which participants had to continually decide which nonverbal displays were "most relevant"; attending to the remote video image to access nonverbal displays like direction of gaze and finger pointing implied not attending to a partner's manipulations of the CVCK in the shared workspace. As a result, participants who tried to utilize the remote video image exhibited a sort of *video schizophrenia*, snapping their gaze back and forth from one video space to the other. This behavior was frequently related to one or more breakdowns, as participants missed crucial events taking place in one space or the other.

Perceiving Details. The constraints imposed by (fixed) framing and resolution limited the utility of the remote video image. Though coarse-grained phenomena like direction of gaze and hand position were readily apparent, it was impossible to read the laboratory manual, or to discern exactly what a participant was pointing at. Consequently, participants gazing at the remote video often had to interrupt the interaction to ask questions like "what are you pointing at?" or "what are you doing?".

In short, it was not clear that the communicative benefits of explicitly attending to the remote video image outweigh the costs; accessing the nonverbal displays available in the remote image may cause at least as much communicative trouble as it avoids. This conclusion is supported by the fact that the incidence of breakdown in session AV3, in which participants *did* gaze frequently at the remote video image, was substantially higher in all four categories (see Table 1) than in the other audio-video sessions. These observations explain why participants in audio-video interactions did not heavily utilize the remote video image and concentrated primarily on the workspace and laboratory manual.

Discussion

The results of Breakdown Analysis show that communicative efficacy in the audio-only and audio-video environments was lower than copresent interaction; significantly more breakdowns were documented in three out of four categories for audio-only interactions, and in two out of four categories for audio-video interactions. This supports the conclusion that copresent interaction represents the "best case" communicative scenario and, importantly, that the two technologically-mediated environments were not functionally equivalent to copresent interactions. By explicitly revealing differences in the amount of communicative breakdown between the copresent and distributed environments, the results of this study provide a causal rationale for existing comparisons of user satisfaction, which consistently show that users overwhelmingly prefer copresent to distributed interaction (Apperley and Masoodian, 1995; Isaacs et al., 1995; Olson et al., 1995) .

The finding that participants in audio-only and audio-video interactions encountered significantly more breakdowns also supports and rationalizes Olson et al.'s (1995) analysis of task-solution activities, which found that participants in audio-only and audio-video interactions invested significantly more time organizing their work and clarifying confusions than their copresent counterparts; presumably this difference can be attributed to participants' efforts to repair a higher number of breakdowns. At the same time, evidence that the quality of work does not vary significantly between copresent and technologically-mediated interactions (Olson et al., 1995; O'Malley et al., 1996) suggests that, even though the communicative efficacy of technologically-mediated interactions may be lower, the quality of the end-product is not necessarily degraded. That is, participants may be able to compensate for lower communicative efficacy by investing more time and effort in the interaction.

No difference in communicative efficacy was found between audio-video and audio-only interactions; participants in both settings experienced similar amounts of breakdown. This result demonstrates that the availability of a video image does not necessarily improve communicative efficacy in a technologically-mediated environment, complementing the findings of existing comparisons . The conclusion that the video channel provided little or no advantage over an audio-only connection to participants in this study supports a growing body of evidence that the value of a video connection has been overestimated by designers. While a video connection has been found to be useful for supporting passive awareness between collaborators, e.g. "Is Mary is her office?", (Mantei et al., 1991; Dourish and Bly, 1993) , its utility in support of active problem-solving interactions is questionable (Hollan and Stornetta, 1993; Kraut et al., 1994; Apperley and Masoodian, 1995; O'Malley et al., 1996) .

Through detailed analysis of participants' low-level communicative troubles, this study presents a rationale for these observations, revealing that participants in audio-video interactions were unable to access nonverbal displays available in the video image to inform their interactions. As a result, the audio-video environment was rendered functionally equivalent to the audio-only environment.

Our analysis of why nonverbal displays were inaccessible via the remote video image sheds new light on recent studies documenting the effects of fixed camera framing, lack of direct eye contact, disparate frames of reference, and difficulty gaining a partner's attention on the perception or interpretation of a partner's nonverbal displays during video-mediated interactions (Heath and Luff, 1993; Sellen, 1995) . Where these studies have focused on mundane personal conversations, in which participants are able to focus attention solely on the remote video image, this study has focused on task-oriented interactions, in which participants must also attend to a shared workspace. Our results suggest that the problems associated with trying to monitor both spaces at once may overshadow any other problems; if one can't consistently attend to the remote video at the appropriate moment, then problems associated with camera framing and supporting direct eye contact are less relevant.

Conclusions

A number of comparative studies have documented differences in user satisfaction, quality of work, and task-activity structure between copresent interaction and interaction in various technologically-mediated environments. The goal of this study was to delve deeper, examining interactions in detail to reveal and compare the communicative breakdowns that occur in copresent, audio-only, and audio-video environments. This has allowed us to move beyond revealing differences in communicative efficacy to explain how these differences arise, causally relating communicative breakdowns to specific characteristics of the environment.

Our findings suggest that the role of video in support of distributed, task-oriented interactions must be carefully reconsidered. In particular, the intuition that environments that provide a video connection between participants are inherently more robust than environments that don't is overly simplistic; there is a great deal of difference between providing access to nonverbal displays and the *practical* utility of such upgrades to participants.

Given that a remote video image may not provide reliable access to a partner's nonverbal displays (at least in task-oriented interactions), it makes sense to think about how one might provide alternative compensatory resources to overcome this limitation. For example, there may be ways to use the audio channel or the shared workspace to somehow indicate a partner's current point of attention or control of the cursor. Another approach might be to move away from traditional video monitors to provide participants with more natural, wide-angle visual access to collaborating partners (Okada et al., 1994) .

There is also some evidence that participants may, over time, develop novel communicative practices to compensate for constraints imposed by a technologically-mediated environment (Dykstra-Erickson et al., 1995) . Though our analysis showed no decreasing trend in the number of breakdowns over the course of interactions, this observation motivates a future longitudinal study to document the effects of long-term experience on the incidence of communicative breakdown.

References

Abel, M. J. (1993): "Experiences in an Exploratory Distributed Organization", in R. M. Baecker (Ed.): *Groupware and Computer-Supported Cooperative Work* , Morgan Kaufman Publishers, Inc., San Mateo, CA, 1993, 792-802.

Apperley, M. and M. Masoodian (1995). "An experimental evaluation of video support for shared workspace interaction", *Proceedings of CHI'95*, Denver, CO, ACM Press.

Clark, H. H. and D. Wilkes-Gibbs (1986). "Referring as a Collaborative Process.", *Cognition* 22, 1-39.

Doerry, E. (1995). "An Empirical Comparison of Copresent and Technologically-Mediated Interaction based on Communicative Breakdown", University of Oregon.

Douglas, S. A. and E. Doerry (1994). "The Cardiovascular Construction Kit", College Park, MD, Academic Software Development Group, University of Maryland.

Dourish, P. and S. Bly (1993): “Portholes: Supporting Awareness in a Distributed Work Group”, in R. M. Baecker (Ed.): *Groupware and Computer-Supported Cooperative Work* , Morgan Kaufman Publishers, Inc., San Mateo, CA, 1993, 809-815.

Dykstra-Erickson, E., et al. (1995). “Supporting Adaptation to Multimedia Desktop Conferencing”, *Proceedings of 15th International Symposium on Human Factors in Telecommunications*, Melbourne, Australia,

Garfinkel, H. (1967). *Studies in ethnomethodology*. Prentice-Hall, Inc., Englewood Cliffs, NJ.

Harrison, S. and S. Minneman (1990). “The Media Space: a research project into the use of video as a design medium”, *Conference on Participatory Design*, Seattle, WA, ACM Press.

Heath, C. (1986). *Body movement and speech in medical interactions*. Cambridge University Press, Cambridge.

Heath, C. and P. Luff (1993): “Disembodied Contact: Communication through video in a multimedia office environment”, in R. M. Baecker (Ed.): *Groupware and Computer-Supported Cooperative Work* , Morgan Kaufman Publishers, Inc., San Mateo, CA, 1993, 837-841.

Hollan, J. and S. Stornetta (1993): “Beyond Being There”, in R. M. Baecker (Ed.): *Groupware and Computer-Supported Cooperative Work* , Morgan Kaufman Publishers, Inc., San Mateo, CA, 1993, 842-848.

Isaacs, E. A., et al. (1995). “A Comparison of Face-to-Face and Distributed Presentations”, *Proceedings of CHI’95*, Denver, Colorado, ACM Press.

Jordan, B. and A. Henderson (1995). “Interaction Analysis: Foundations and Practice”, *Journal of the Learning Sciences* 4, 39-103.

Kraut, R. E., et al. (1994). “Life and Death of New Technology: Task, Utility and Social Influences on the Use of a Communication Medium”, *Proceedings of CSCW’94*, Chapel Hill, NC, ACM Press.

Mantei, M. M., et al. (1991). “Experiences in the Use of a Media Space”, *Proceedings of CHI’91*, New Orleans, LA, ACM Press.

O’Malley, C., et al. (1996). “Comparison of face-to-face and video-mediated interaction”, *Interacting with Computers* 8(2), 177-192.

Okada, K.-i., et al. (1994). “Multiparty Videoconferencing at Virtual Social Distance: MAJIC Design”, *Proceedings of CSCW’94*, Chapel Hill, NC, ACM Press.

Olson, J. S., et al. (1995). “What Mix of Video and Audio is Useful for Small Groups Doing Remote Real-time Design Work?”, *Proceedings of CHI’95*, Denver, Colorado, ACM Press.

Root, R. W. (1988). “Design of a multi-media vehicle for social browsing”, *Proceedings of CSCW’88*, Portland, OR, ACM Press.

Sacks, H., et al. (1974). “A simplest systematics for organization of turn-taking in conversation”, *Language* 50(4), 696-735.

Sanderson, P. M. and C. Fisher (1994). “Exploratory Sequential Data Analysis: Foundations”, *Human-Computer Interaction* 9(3,4), 251-318.

Sellen, A. J. (1995). “Remote Conversations: The effects of mediating talk with technology.”, *Human-Computer Interaction* 10(4), 401-444.

Siegel, S. and N. J. Castellan (1988). *Nonparametric statistics for the behavioral sciences*. McGraw-Hill, New York.

Suchman, L. (1987). *Plans and Situated Actions: The Problem of Human-Machine Communication*. Cambridge University Press, Cambridge, England.

Whittaker, S., et al. (1991). “Co-ordinating activity: An analysis of interaction in computer-supported co-operative work”, *Proceedings of CHI’91*, New Orleans, LA, ACM Press.