SOCIAL AND COGNITIVE PROCESSES IN INTERPERSONAL COMMUNICATION: IMPLICATIONS FOR ADVANCED TELECOMMUNICATIONS TECHNOLOGIES*

SUSAN R. FUSSELL, Mississippi State University, Mississippi State, Mississippi, and

NICHOLAS I. BENIMOFF, AT&T Bell Laboratories, Holmdel, New Jersey

Interactive multimedia conferencing systems, in which two or more remotely located people can work on cooperative tasks through shared audio, video, and data, appear to be the wave of the future. However, because of great advances in the underlying technology of multimedia conferencing systems, many design decisions have been driven by what is technically feasible as opposed to what will best suit the needs of the users. In this paper we provide a framework for the design and evaluation of features in advanced telecommunications products and services which is derived from empirical research on interpersonal communication. We also discuss implications of this research for the development and use of advanced telecommunications technologies.

Running Title: INTERPERSONAL COMMUNICATION

Key words: telecommunications, interpersonal communication, computer-supported cooperative work, multimedia, electronic conferencing.

* Fussell, S. R., & Benimoff, N. I. (1995). Social and cognitive processes in interpersonal communication: Implications for advanced telecommunications technologies. *Human Factors*, **37**, 228-250. Please send reprint requests to: Susan R. Fussell, Human-Computer Interaction Institute, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213. Email: susan.fussell@cmu.edu

INTRODUCTION

Recent advances in a variety of telecommunications technologies have resulted in an extraordinary evolution in the ways in which people are able to communicate with one another. Interpersonal communication is in the midst of being whisked into the 21st century a few years early with the maturation of the set of technologies that enable multimedia conferencing. The ability to interact with people not in the same room in ways other than merely speaking by telephone has already been the focus of a great deal of attention, from the introduction of multimedia telecommunications products in the marketplace (e.g., videophones and PC-based video systems), to the burgeoning interest, in both the academic and business worlds, in the field of computer-supported cooperative work (*CSCW).* Before long, we are certain to witness particularly great advances in how remotely-located people interact to perform cooperative work activities, such as co-design a product, make marketing decisions, collaborate on a joint article, or attend classes without traveling to a classroom.

Human factors experts have had significant influence in the evolution of multi-media systems. Some have identified a general user interface issues for multimedia conferencing systems (e.g., display quality, interface design, privacy and security issues, etc.; e.g., Benimoff & Burns, 1993; Benimoff & Whitten, 1993). Others have evaluated specific systems in terms of usability and other criteria, while yet others have addressed issues concerning the structures and requirements of work groups that might benefit from advanced telecommunications technologies (e.g., Gabarro, 1990; Kraut, Egido, & Galegher, 1990; Kraut, Galegher, & Egido, 1987-1988; McGrath, 1990). Despite these contributions, however, and despite Kiesler, Siegel and McGuire's (1984) cautions to the contrary a decade ago, we believe that human factors expertise is often solicited late in the development cycle. As a result, many basic design decisions for multimedia conferencing products have been driven by what is technically feasible as opposed to what will best suit the needs of the users.

Our aim in this paper is to review the growing psychological literature on interpersonal communication with an eye toward identifying principles to guide the design and evaluation of

communication-enhancing features in advanced telecommunications products and services. We will use a broad definition of multimedia conferencing wherein it refers to two or more remotely located people electronically sharing audio, video, and data, either via their desktop PC or a group room system (see Figure 1 for a sample configuration). An extension of the more familiar audio teleconference, participants in a multimedia conference can not only speak with one another, but have the ability to see each other, as well as share documents, on-screen whiteboards, stored video clips, and the like.

Insert Figure 1 About Here

We conceive of multimedia-enabled conferences (and all teleconferences), as extensions of ordinary conversations; thus, the design and implementation of successful multimedia telecommunications systems rests on an understanding of what people do when they communicate face-to-face. Just as in face-to-face conversation (Austin, 1962/1975; Searle, 1969), teleconference attendees actively engage in producing and understanding messages (commands, questions, suggestions, etc.) as a fundamental element of their achievement of their joint goals (e.g., Winograd, 1987-1988).

Our basic assumption is that the users of multimedia conferencing technologies, like all speakers and hearers, actively attempt to construct a shared communicative context in which their messages can be produced and understood (Krauss & Fussell, in press, 1990). In this view, the participants in a conversation strive for a shared understanding of the situation, the task, and of one another's background knowledge, expectations, beliefs, attitudes, and the like. They also construct a body of shared knowledge and understanding (*common ground*) which they can draw upon in their subsequent communications (Clark & Brennan, 1991; Clark & Wilkes-Gibbs, 1986; Fussell & Krauss, 1992; Krauss & Fussell, 1991b). We believe that design decisions concerning all aspects of multimedia conferencing systems (e.g., shared workspaces, video transmission, etc.) should be considered in light of this notion of a shared communicative context.

For the sake of brevity, we will limit our discussion in several ways. First, we will focus upon computer-mediated *interpersonal* communication, rather than human-computer interaction,

although many of the same principles apply to both. (For discussions of conversational interfaces between users and computers, see Brennan, 1991; Hindus & Brennan , 1992, and Luff, Gilbert & Frohlich, 1990).

Second, we will focus specifically on multimedia conferencing systems. There is a substantial body of research readily available on audio-only and computer-based CSCW products and services (see, e.g., the papers in Galegher, Kraut, & Egido, 1990, Greif, 1988, and the CSCW '92 and CSCW '94 proceedings). It is well-documented that different means of communication (telephone, e-mail, computer bulletin boards, videoconferencing, etc.), because of their different methods of transmission (e.g., asynchronous vs. synchronous transmission, time-delayed vs. on-line interaction), restrict to varying degrees users' access to information that is available in face-to-face conversation. Research by Kiesler and her colleagues (e.g., Kiesler et al., 1984; Kiesler, Zubrow, Moses, & Geller, 1985; Kiesler & Sproull, 1992) has shown that such limitations affect both task-related and social aspects of cooperative work. Rather than reiterate previous discussions, we will focus here on the issue of information availability as it applies to on-line multimedia conferencing.

Finally, we will limit our discussion to findings from empirical research in the psychology of communication, rather than review the many human factors evaluations of specific systems that are widely available through the other sources listed above. A number of systems have already incorporated some of the principles we discuss, with varying degrees of success. With a small set of exceptions, however, most of multimedia conferencing products and services have been user-tested relatively late in the development cycle, and consequently, it is difficult to determine whether the constructed system was the best possible design.

Overview of Paper

In this paper we summarize basic research that bears on the construction of shared communicative contexts, and discuss implications of this research for the development and use of advanced telecommunications technologies. We begin by briefly describing the functions of several channels of communication (e.g., verbal messages, gesture, eye gaze), and discuss how

conversations are influenced by the presence or absence of these channels. Next, we review research on the importance of perspective-taking in the establishment of a shared communicative environment, and discuss how future products can facilitate the perspective-taking process. Then we consider conversational interaction, including turn-taking rules and the processes by which communicators ensure that messages have been correctly understood. Finally, we draw some general conclusions about videoconferencing features and design and discuss future directions that research and development might take to better accommodate end users' natural communicative processes.

SOURCES OF INFORMATION

It is customary to distinguish between *signals* and *signs* in communication (e.g., Clark, 1985; Krauss, 1979). Signals are intentional acts -- attempts to communicate a particular message to another person using conventional words or symbols. Listeners attempt to understand this message by identifying the speakers' intentions (Grice, 1957). In contrast, signs, such as blushing, yawning, frowning, and so on, are not intended to be communicative; however, they are often informative as to the state of the conversation (e.g., the listener's attention, comprehension, reactions, etc.). It is useful to keep the distinction between signals and signs in mind when considering the importance of various channels of information on the communicative process.

Channels of communication are usually categorized as *linguistic*, *paralinguistic* or *nonlinguistic*. Linguistic information (also called the *verbal channel*) consists of conventional signals, such as the words of a language, typically expressed via speech or written communication. Paralinguistic information includes intonation patterns and pauses, which are directly linked to the verbal channel. In contrast, nonlinguistic information includes eye gaze, posture, many gestures, and other signs that can convey meaning that is seemingly distinct from the verbal channel and thus often termed "nonverbal." There is not a one-to-one relationship between channels of communication and the production of signals vs. signs. While many signals

are verbal, others, such as certain hand gestures (e.g., the "victory" or "ok" hand gestures), are not.

Current telecommunications network bandwidth limitations make it difficult to transmit all of the sources of information available in face-to-face conversation at the same rate and fidelity as they occur naturally. Even if the technology existed, cost considerations alone would suggest that we look carefully at the importance of each channel of communication for successful cooperative work. Current research supports the view that the verbal channel is by far the most important in communicating meaning (e.g., Krauss, Garlock, Bricker, & McMahon, 1977). Nonetheless, paralinguistic and nonverbal information play important roles in face-to-face conversation. Here we briefly describe the functions of two nonverbal channels -- eye gaze and gesture -- which can be transmitted in video-enabled multimedia conferencing systems but which are absent in audio-only and audiographics systems. (The interested reader should consult one of the recent collections of papers on nonverbal communication such as Feldman & Rime, 1991, Scherer & Ekman, 1982, and/or Siegman & Feldstein, 1987, for more detailed information.)

Eye Gaze

Several studies have addressed the importance and usability of specific features in multimedia conferencing systems, particularly the role of visual information (e.g., Fish, Kraut, Root, & Rice, 1992; Gale, 1991; Heath & Luff, 1991; Tang & Isaacs, 1993). While many studies have found that users desire video connectivity, the evidence thus far on the importance of video for interpersonal communication has been mixed. Although some investigators have concluded that video is relatively unimportant for successful communication, we agree with others, such as Angiolillo, Blanchard and Israelski (1993), that adding video to voice calls can enhance interpersonal interactions, so long as valid human factors techniques are employed to evaluate and optimize usability. In this paper we take the stance that the future of multimedia teleconferencing lies in video-based products and services, and that the task at hand is not to

evaluate video per se, but to examine the functions of video in face-to-face communication with an eye toward identifying design considerations for the new video age.

Communicators' eye gaze in face-to-face conversations serves several important functions (Argyle & Cook, 1976; Kendon, 1967; Rutter & Stephenson, 1977). As they are formulating their messages, speakers tend to look away except for short glances at the listener. As they move toward completion of the utterance, their gaze returns to the addressee. As we will discuss below, one function of this shifting of gaze is to help coordinate speaking turns. In addition, the shifts away from the listener are thought to help speakers formulate their messages by reducing cognitive load. Addressees, in contrast, tend to focus upon the speaker during the entire message. Speakers periodically glance at their addressees to ensure that they are receiving eye gaze, and lack of gaze is taken as a sign of boredom, disinterest, or, more generally, lack of engagement in the conversation. Gaze is likely to serve similar functions in multimedia conferences; thus, conferencing products should be designed such that they can convey gaze information differentially across multiple parties in a multimedia conference. As we discuss in detail below, this goal has a number of design implications revolving around the size and quality of the video window.

Gesture

In face-to-face conversation, people gesture frequently. Ekman and his colleagues (Ekman & Friesen, 1969) have identified several classes of gestures, some of which function as signals (i.e., as movements that carry meaning) and others of which act as signs (e.g., indications of uncertainty, nervousness, and the like). The repertoire of gestures with specific semantic content (such as a "V" for victory) is rather small, and these are readily replaced by verbal substitutes when there is no visual contact between communicators. Deictic gestures, in which one points to a particular location, object, or person, are frequent in face-to-face conversation and also may be replaced by verbal substitutes, though these substitutes may take more time and thus interrupt the smooth flow of conversation. For example, it is easier to say, while pointing to a specific sentence on a document, "change this," rather than "change sentence four in

paragraph three on page 24." In addition, some investigators (e.g., McNeill, 1992) have argued that gestures can enhance the meaning of the verbal utterance by providing additional clarifying information. The usefulness of transmitting gestural information will clearly depend on the type of gesture and the task being performed, but clearly there are many cooperative endeavors in which it would be beneficial to provide this information.

Gestures also have a role in the smooth transition of speaking turns (Duncan & Fiske, 1977), as we will discuss further below, and for that reason alone, their inclusion in video transmissions would be desirable. To incorporate gesture into the video feed, the video display content and fidelity must be such that it provides the correct information (by manipulating the camera lens visual angle or depth of field, or by manipulating video window size, increasing monitor resolution, or improving video transmission characteristics such as scan rate, frame per second, etc.). In addition, as illustrated in Figure 2, when the size of the video window must remain constant, the provision of gestural information will necessitate a trade-off between facial expressions (including eye gaze) and gesture. Furthermore, as will be discussed in greater detail later, discrepancies in transmission rates for audio and video information can lead to confusions when a speaker is simultaneously speaking and gesturing.

Insert Figure 2 About Here

In addition to the above considerations, recent research suggests that gestures also function as aids to the message formulation process (Krauss, et al., 1991; Morrel-Samuels & Krauss, 1992). When the movement of someone's hands is restricted, he or she produces more pauses and hesitations (Bilous, 1992), both of which are signs of difficulty in message formulation and articulation (Butterworth, 1980). Similarly, when people are having trouble thinking of a word or phrase, they often gesture (Morrel-Samuels & Krauss, 1992). This suggests that video systems should allow communicators' hands to be as free as possible (i.e., the audio connections should not be hand-held), to facilitate the use of gestures and to create a more natural communicative environment.

PERSPECTIVE-TAKING

Communication obviously does not occur in a social vacuum. Even when others are not physically present, our conceptions of them, as well as the social and linguistic context, shape how we formulate and interpret messages (e.g., Mead, 1934). More than 25 years of research has shown that speakers attempt to take their addressee's background knowledge, beliefs, attitudes, and so on, into account when they formulate messages, and that these acts of perspective-taking facilitate comprehension (e.g., Fussell & Krauss, 1989, 1992; Higgins, 1992; Isaacs & Clark, 1987; Schober , 1993; see Krauss & Fussell, in press, for an extensive review of this literature). Through perspective-taking, speakers and hearers strive to establish and maintain *common ground* -- a shared communicative context in which each participant is aware of the point of view of the others (Clark & Marshall, 1981; Krauss & Fussell, 1988; Krauss & Fussell, 1991b).

In this section we describe two types of perspective-taking that are fundamental to interpersonal communication -- assessing the physical context as seen by others, and determining others' background knowledge, social roles, and point of view. We also discuss design issues that arise in for perspective-taking in multimedia conferencing environments.

Creating a Shared Physical Context

Communication takes place in a physical location that is shared to greater or lesser degrees by all participants. In face-to-face discussions, this context includes the setting of the discussion and the objects and people therein (e.g., tables, overhead projectors, computers, and other conversational participants). In order to communicate successfully, each person must have an understanding not only of the location of these objects and people in their own visual fields, but in the visual fields of the other users as well (Clark & Marshall, 1981; Krauss & Fussell, 1988; Krauss & Fussell, 1991b; Schober, 1993). In face-to-face conversation, communicators assume that objects and entities that are visually available to all participants (that is, are *co-present*) are part of their common ground (Clark & Marshall, 1981). They can readily establish copresence by examining the physical setting of the conversation. For example, to use a deictic

expression such as "what do you think of this computer chip," the speaker must be able to ascertain that the chip is in the view of the addressee, and that he or she is attending to it.

In multimedia conferencing systems, a "virtual" physical setting is created by the shared workspace (e.g., an on-line whiteboard, shared documents), and users of these systems likewise must establish that objects and persons are co-present. For example, the speaker above might ask the same question about the computer chip in a videoconference in which the chip design is displayed in a shared graphics program. As in the face-to-face situation, he or she will need to establish that all conference attendees are looking at this graphics window.

Determining co-presence can be quite problematic in multimedia conferences. The difficulty is not simply that less of the visual field is shared -- research on single-line telephone communication has demonstrated that this limitation in itself is rarely problematic (e.g., Rutter, 1987), because telephone callers are aware that the only information they share is that provided by the voice channel. Rather, the problem stems from the fact that participants' visual fields in multimedia conferences may be shared to greater and lesser extents, and individual participants rarely have direct knowledge of what is seen by the others. In most present and planned desktop conferencing systems, individuals will have some control over the size and placement of on-screen windows that display video, shared data, and other information. As shown in Figure 3, this flexibility in window arrangements can lead to large discrepancies in the visual fields of various conference attendees. Thus, the addressee above may not have the graphics window with the chip design open, may have other windows overlayed upon it, or may have iconified it such that it cannot be read.

Insert Figure 3 About Here

A related problem arises from the need to assess the listener's position vis a vis the objects or persons being discussed. For example, in face-to-face conversation, speakers take into account that items on their right are on the addressee's left when formulating messages (e.g., Schober, 1993). This simple rule does not necessarily apply in electronically-mediated discussions. Even when all conference attendees have the same document open, individual

preferences in window placement may alter its relative location to other windows. Thus, expressions such as "the item in the left-hand window" can lead to potential confusions, especially when users must create and manipulate several window at the same time. This problem may be further compounded when the conferencing system provides individuals with either mirror-image or direct feedback from their own location.

An additional issue that arises in advanced videoconferencing systems is the potential for confusion between personal and public space. Most systems allow users to have open their own private files (e.g., datebook, notepad) in addition to the shared workspaces. Individual users may easily become confused as to which display windows are part of the conferencing environment and which are local to their own PC. As a result, they may erroneously assume that some information displayed is seen by others when it actually is not. Multimedia products need to ensure that this type of confusion between private and public workspaces will not occur.

Communicators' Spatial Relationships. The other individuals present, either in person or by video links, are also part of the physical context. As physical entities, they are located in particular positions (for instance, around a table) and communicators make use of this location information when they are speaking and listening. Speakers frequently direct their eye gaze to particular individuals, often as an indication of whom they are addressing, and whom they expect to respond to the utterance (Argyle & Cook, 1976; Kendon, 1967; Rutter & Stephenson, 1977). When meeting attendees are co-present, communicators can also use deictic expressions and gestures to designate persons (e.g., point at someone rather than saying his or her name). Thus, shared knowledge of the arrangement of attendees plays an important role in the construction of shared communicative contexts.

These issues suggest that videoconference attendees would benefit greatly from the inclusion of features that create a shared "virtual" physical context in which knowledge of the location of individuals and objects is shared by all. An excellent example of the realization of this principle is the MAJIC system (Okada, Maeda, Ichikawaa & Matsushita, 1994), in which users face a large screen onto which is projected life-size images of up to four remotely-located

conference attendees, and in which spatial relationships between participants are maintained. In this impressive system, an extension of ClearBoard (Ishii, Kobayashi, & Grudin, 1992), all participants can track one another's gaze and gestures, and there is a strong sense of a shared physical context. However, the expense, size, and other limitations of ClearBoard, MAJIC, and similar systems suggest that they will not take the place of desktop-based ones at least in the foreseeable future. For these desk-top systems, designers will need to weight the benefits of providing users with flexible screen arrangements against the costs of this flexibility for the smooth flow of communication. We return to this issue in the final section of this paper.

Salience. Another aspect of physical context that is often overlooked in considerations of advanced features in telecommunications technologies is the importance of visual prominence. Research has shown that individuals who stand out from the others in a group are more memorable, and often judged to be more influential or important (e.g., Taylor & Fiske, 1975, 1978; see Fiske & Taylor, 1991, for a review), at least when the group consists of unacquainted individuals of equal status. A person's salience can be affected by individual and group composition, by intentional acts such as wearing a different style of clothes than the others, and, most importantly for our current purposes, by aspects of the shared physical environment. Individuals who are in one's direct line of view are judged to play a more central role in the discussion, to contribute more often, and to make better points than others who talk equally often and make equally valid contributions (Taylor & Fiske, 1975, 1978).

In face-to-face conversation, these effects of physical location are minimized because listeners can turn their head toward the speaker. Similarly, since none of the participants are visible in audio teleconferencing, no single individual or group of individuals should be perceived as more salient than the others. However, salience imbalances can easily arise in advanced telephone technologies, such as videoconferencing systems, due to attendance mode (e.g., audio-only versus A/V), equipment quality, and end users' individual preferences as to the arrangement of video feeds.

In addition, multipoint video conferencing is characterized by several video switching modes, including voice-activated switching, presentation mode, broadcast with autoscan, and chair control (W. Clark, 1992), each of which affects salience by providing various levels of control over who is seen and heard, and for how long. For example, with voice-activated switching, conference attendees see the person who is speaking the loudest (technically, the person with the greatest voice energy) and that person usually sees the last person to speak. This is likely to create the least salience imbalance for most attendees, since it most closely approximates everyday face-to-face conversation; however, it does create an imbalance for the current speaker by focusing his or her attention on the previous speaker.

Other modes of switching lead to similar problems. With presentation mode, which is often used in distance education efforts, everyone in the audience sees a predesignated person (e.g., the instructor), and whom that person sees is based on voice activated switching (e.g., a student asking a question). Presentation mode switching has a strong potential for creating salience imbalances because the last member of the audience to speak typically remains in the speaker's view until someone else speaks up. Broadcast mode is similar to presentation mode in the existence of a predesignated person who is seen by all, but it is an improvement in that this person views each of the members of the audience on a regular rotating basis. Aside from reducing salience imbalances, this method enables speakers to monitor the audience for signs of attention, confusion, and the like. Finally, with chair control video switching, a designated person has control over who everyone sees, and whom the viewed individual sees. Clearly, this method has the potential to make some individuals more visually prominent than others, but this will depend on the motives and goals of the person controlling the camera feeds.

When designing a system or service that requires video switching, the task demands obviously play a large part in selection of video switching mode; nonetheless, care must be taken to understand the effects of these choices with respect to the governing of participants' salience.

Creating a Shared Social Context

Successful interpersonal communication not only demands that speakers take into account their addressees' physical contexts and orientation toward objects and entities, but also that they consider these addressees' background knowledge, areas of expertise, attitudes, beliefs, motives, goals, and the like. The achievement of a variety of work as well as social goals rests in part on each participants' having an adequate understanding of the others' points of view. We first describe several areas in which knowledge of others' perspectives is essential to successful communication, both face-to-face and electronically-mediated. Then, we discuss how communicators assess their conversational partners' perspectives in face-to-face communication and how we might facilitate this process in multimedia teleconferencing products and services.

Arenas of Perspective-Taking. One level at which perspective-taking occurs is that of assessing individual differences in knowledge and expertise. Members of project teams, executive committees, and other work groups often have different educational backgrounds, and concepts and terminology that are used frequently by one subset of meeting participants may be unfamiliar to others. To communicate successfully, speakers need to take such differences into account (Fussell & Krauss, 1989, 1992; Krauss & Fussell, 1990; 1991; Isaacs & Clark, 1987). In addition, many of the complex tasks facing today's workers require contributions from individuals with different areas of expertise -- that is, they rely on *socially shared* cognition (e.g., Anderson, Heath, Luff, & Moran, 1993; Hutchins, 1990; Olson & Olson 1991; Resnick, Levine, & Teasley, 1991). In performing such tasks it is necessary that group members can quickly and easily identify the relevant individual(s) to consult for particular pieces of information.

Perspective-taking is not only required for message construction, but also for the creation of names and labels for shared entities in the multimedia workspace such as shared files, whiteboards, and notebooks. Conventions of discourse (Grice, 1975) and length limitations for file and window labels require that these names be brief yet meaningful to all users of a particular product or service. Research has shown that individual computer users readily create short labels for their own files, and that they can use these labels to recall what is contained in

those files even after several years delay (Carroll, 1985); however, there is little agreement *between* users as to what particular file names indicate (Furnas, Landauer, Gomez, & Dumais, 1987). Similar problems are likely to arise for visual icons: a schematic or abstract drawing may resemble one thing to one person and something entirely different to another (Fussell & Krauss, 1989; Clark & Wilkes-Gibbs, 1986). For developers and users of conferencing systems to create file and window names and icons that are readily understood by all conference attendees, they will need to take one another's perspectives into account.

Communicators also must take into account status and role differences between them. For instance, the knowledge that someone is in a managerial role enables others who are in lower positions in the business hierarchy to follow appropriate politeness conventions (e.g., using a title and a certain degree of restraint in expressing disagreement). Even cooperative group members at the same level of an organizational hierarchy may come from different parts of a company, or from different companies, with differing aims and goals. A developer of computer software, for instance, may perceive the goal to be the design of a product with elegant, error-free computer code whereas human factors experts may be specifically concerned with designing a compelling interface that is easy to use. An awareness of these differences in orientation is essential both for successful task completion and for the fulfillment of other goals such as establishing rapport and good working relationships, impressing one's superiors, and developing new friendships (Gabarro, 1990). In fact, workers who could potentially interact via teleconferencing sometimes decide to meet face-to-face instead for precisely these reasons.

Finally, individual differences in social and cultural background have the potential to alter the dynamics of cooperative work sessions by affecting one's attitudes, communication styles, interpersonal expectancies, and the like. For instance, men and women appear to prefer somewhat different styles of conversation (e.g., Tannen, 1990). Many companies and educational institutions today are emphasizing diversity training to increase employees' awareness of socio-cultural differences, in hopes of facilitating communication and reducing erroneous assumptions about others' characteristics. If one knows, for instance, that a fellow

meeting attendee comes from a background in which interruption is considered impolite, one can avoid mistakenly concluding that this person has nothing to say and instead ask him or her directly for input.

In short, there are many reasons why cooperative work group members benefit from being able to assess one another's points of view. In our view, designers of telecommunications products and services that aim to simulate face-to-face communication must take into account research on people's natural methods of ascertaining others' perspectives, and design systems with features that support these mechanisms. In some cases (e.g., very close personal or working relationships), people have relatively direct knowledge of others' perspectives. Festinger (1950) long ago demonstrated the importance of physical proximity in the establishment of close relationships , and, as Kraut and his colleagues note (e.g., Kraut, et al., 1990), an ideal collaborative tool would allow users to communicate informally as well as formally about their tasks. Since to date there are only a few systems that permit this type of informal interaction (e.g., CRUISER, Root, 1988), we focus here on situations in which people must infer others' perspectives from less direct information. Two primary sources of information are others' social and business category memberships and the conversational interaction itself (Fussell & Krauss, 1992; Krauss & Fussell, 1991b). We defer the topic of conversational interaction until the next section; here, we address social categorization, and in particular, the use of social schemas in social perception and inference processes.

Social Categorization and Inference Processes. A substantial body of research in social cognition research has addressed the ways in which we perceive, categorize, and recall information about others (see Fiske & Taylor, 1991, for a detailed review of this literature). Research has shown that upon first meeting someone, we categorize them along a number of dimensions, including gender, ethnicity, age, personality type (e.g., aggressive, quiet, manipulative), and social roles, both business-related (e.g., engineer, executive) and non business-related (e.g., single woman, father, fellow tennis player). Much of the time, these categorization processes occur rapidly and outside of our conscious awareness. They are often

based on people's physical characteristics (facial features, clothing), and other visible accomplements (e.g., briefcases, portable PCs), although other information such as accent, conversational style, and job title are also important.

This rapid categorization process is useful in that it enables people to draw inferences about others that will guide them in their interactions with them. Once someone has been categorized, social schemas -- cognitive structures that contain information of all sorts about members of that category -- are activated, and used as the basis for further inferences about that individual (see Fiske & Taylor, 1991, or the collection of papers in Wyer & Srull, 1994, for an in-depth review). For instance, after classifying a new acquaintance as a computer developer, we would assume that she had extensive knowledge of programming languages. We also might make inferences about stereotypical traits of developers, such as: she tends to work long hours, she is likely to complain about the incorporation of human interface features, etc. Several studies have documented substantial interpersonal agreement and fairly high accuracy in people's judgments of the background knowledge, attitudes, and behaviors of members of specific social categories (e.g., Fussell & Krauss, 1991; 1992; Kunda & Nisbett, 1986); however, this accuracy is based on a correct initial categorization. In fact, inferences based on erroneous social categorizations are likely to remain unchanged even when correct category information is later provided.

Accurate perspective-taking thus depends on communicators' abilities to assess others' work and social category memberships and to draw correct inferences from these memberships. In many current telecommunications environments (e.g., e-mail, computer conferencing, audio teleconferencing), such categorization is made more difficult by the unavailability of visual information. For example, although one can usually identify the manager in a group of workers rather quickly from his or her more formal dress, this information must be supplied directly, if it is available at all, when only text or audio channels are present (Krauss & Fussell, 1990). Newer technologies that incorporate video have the potential to reduce categorization problems by

providing the physical cues to an individual's social and work-related roles that are present in face-to-face communication.

Thus, a key issue for the future development of telecommunications products and services is the identification of the information most useful for correctly categorizing others, and developing products that transmit this information. Electronic meeting systems should provide more than icon-sized bitmap displays of each participant. Iconified images do not convey the full richness of social information that is present in face-to-face conversation. Images may be so small that users cannot detect the social information contained in them; they also may not be up-to-date: some systems allow users to supply a single picture of themselves, which they use in every situation. Systems in which users have the capability to transmit stored scanned images of themselves plus some identifying information such as name, title, etc. are an improvement; nevertheless, unless real-time video is present, important sources of information about others, particularly their engagement and interest in the meeting itself, will be lacking. As illustrated in Figure 4, even when real-time video is present, the images still may be too small to communicate information about eye gaze, gesture, clothing, and/or physical context. In addition, as shown in Figure 2 above, one may need to make tradeoffs between various sources of information such as facial expressions, gestures, and full-body representations, when the video window size must remain constant.

Insert Figure 4 About Here

Overhearers and Privacy Issues

In addition to the intended addressees in a conversation, there may also be overhearers (both intended and unintended). When communicators know of the presence of unwanted overhearers, they are fairly adept at creating messages that are communicative to the desired audience but incomprehensible to these eavesdroppers (Clark & Schaefer, 1987; Schober & Clark, 1989) . However, some communications technologies make the detection of overhearers problematic. For example, if a caller uses a speakerphone, conference attendees may not be aware of silent participants in the same room as speakerphone users. Also, in audio

teleconferencing, if someone calls up early, after the conferencing bridge is established but before most participants have joined in, his or her presence may go undetected by the others. Since future video conferencing systems are expected to provide at minimum a roster of names of all participants, eavesdropping should be less problematic; however it is important to keep in mind the need for a feature that allows for the identification of overhearers in all such systems.

COMMUNICATION IS COORDINATED ACTIVITY

Herbert Clark (1987) has likened conversation to a game of tennis: each participants' moves (messages) must be coordinated at every moment to what the others are doing. This coordination helps in the establishment and maintenance of a shared communicative context. Speakers and listeners coordinate their messages on a variety of levels, two of which we briefly review in this section: the orderly exchange of conversational turns and the establishment of shared understandings (see Clark, 1985, and Clark & Brennan, 1991, for overviews of this literature).

Turn-taking

On the whole, speakers' contributions to conversation are well coordinated: there is little overlap between speaking turns, and when such overlap does occur, it is usually resolved quickly, with a single individual taking over the floor (Sacks, Schegloff, & Jefferson, 1974). In addition, there is usually little or no perceptible delay between the end of one person's turn and the start of the next one. When longer delays do occur, they are often experienced as awkward silences. It is critical for such coordination that communicators be able to *project* the end of the prior speaker's turn. The cognitive processes involved in formulating and articulating a message take some time (Levelt, 1989), so if speakers waited until the end of the previous message before beginning their own, noticeable gaps in the conversation would occur.

There are a number of cues that communicators use to project the end of speaking turns (e.g., Duncan & Fiske, 1977; Sacks et al., 1974). Some of these cues are linguistic -- based on the meaning and syntax of the sentences uttered. Others are paralinguistic, such as the falling intonation characteristic of the end of declarative sentences, or the rise typical of questions.

These two sets of cues alone appear to be sufficient to coordinate understanding in two-party telephone calls (Rutter, 1987). However, substantial research indicates that the resulting conversations contain more pauses, interruptions, and the like than face-to-face conversations, at least for task-oriented dialogues (e.g., Argyle, Lalljee, and Cook, 1968; Boyle et al., 1994). Consequently, the amount of time taken to communicate a given piece of information is substantially longer. Given the current high cost of video teleconferencing, it stands to reason that users will benefit financially if, given a constant bandwidth, that bandwidth can be used to its full communicative potential.

Examination of turn-taking in face-to-face dialogues suggests that speakers also have the option of selecting the next person to talk (Sacks, Schegloff, & Jefferson, 1974). Although this can be done by verbal means, gestures and eye gaze can facilitate the process. Deictic gestures, such as pointing to a particular individual, can be used in lieu of that person's name to select him or her to be the next speaker. Direct eye gaze can serve similar functions (Duncan & Fiske, 1977; Kendon, 1967). In order for gaze and gestures to function as a turn-coordinating devices, however, participants must share a "virtual" physical environment. Thus, the ease of coordinating contributions to conversations in multimedia systems will be influenced by many of the same design issues we have previously considered, including user flexibility in video feed placements, the conference's mode of video-switching, and the like.

"Grounding" of meaning

Herbert Clark and his colleagues (e.g., Clark & Brennan, 1991; Clark & Wilkes-Gibbs; Wilkes-Gibbs & Clark, 1992) have argued that communication is a collaborative process, in which speakers and addressees work together to establish that the message has been understood as intended. We examine two processes by which communicators establish that a message has become part of their common ground: feedback from backchannel responses and conversational interaction.

Backchannel Responses. In everyday conversation, listeners are not inactive as speakers are talking. Rather, they provide feedback through what are known as backchannel responses --

brief utterances such as "I see", or "uhuh", head nods, looks of puzzlement, and the like. Speakers monitor these backchannel responses to ensure that the audience is attending to them and that the message is understood (e.g., Clark & Wilkes-Gibbs; Duncan & Fiske, 1977; Krauss et al., 1977; Kraut & Lewis, 1982). When an addressee withholds backchannel responses, it is typically viewed as a sign of lack of either comprehension or attention, and speakers adjust their behaviors accordingly.

In two-party telephone conversations, feedback is restricted to the verbal and vocal channels of communication but this limitation does not appear to have a major impact on people's ability to communicate effectively (e.g., Rutter, 1987). However, in larger groups, verbal feedback from individual participants is usually either inappropriate (e.g., there would be too many people speaking at once) or inaudible (as in a large lecture hall). As anyone who has given face-to-face lectures or talks knows, the most reliable signs of comprehension and attention in group situations are the eye gaze and facial expressions of the audience. In addition, when listeners are in view of one another, each can monitor the responses of the others. Thus, as is the case for supporting turn-taking and perspective-taking, successful multimedia products and services will need to provide as much visual contact between meeting participants as possible if they are to support the full range of feedback mechanisms present in face-to-face conversation. Some features that are common to such systems may restrict access to feedback. For example, broadcast mode switching does not allow the speaker to scan the audience for signs of understanding. And, since messages are adapted specifically for those providing feedback (Kraut, Swezey, & Lewis, 1984), desktop videoconferences in which some people are not connected by video will increase speakers' difficulty in assessing how well they are communicating their ideas to *all* meeting participants.

There are further considerations in the development of conferencing products that allow for audio and video feedback. For communicators to be provided with the feedback they need, it is essential that information transmittal rates be high, and that all channels of communication be properly synchronized. Several early studies found that transmission delays adversely affected

speakers' message production processes (Krauss & Bricker, 1966; Krauss & Weinheimer, 1966). Most current and prospective multimedia conferencing products are characterized by an asynchrony between the audio and video channels, the latter often following the former by a third or half second or more, depending on a variety of factors. This asynchrony, due in large measure to the video compression/decompression (CODEC) algorithms the products employ, can easily lead to confusion when listeners are providing visual feedback regarding their states of comprehension (e.g., frowns, puzzled looks). Specifically, speakers may receive this feedback after they moved on to their next point, and thereby be mistaken about what aspects of their messages are confusing to the addressee. Delaying the audio an amount equal to that necessary to have audio and video presented simultaneously may result in a second or so delay from the speaker to the audience, which itself could cause problems in turn-taking behaviors.

Interactive Grounding. In addition to providing feedback during message production, communicators work to ensure that each message has been understood correctly before they move on to a new message -- that is, they *ground* each contribution via interactive dialogue (Clark & Brennan, 1991; Clark & Wilkes-Gibbs, 1986; Wilkes-Gibbs & Clark, 1992). Listeners may make a variety of responses to a speaker's message, and these responses are informative as to those listeners' understanding (and thus the current status of common ground). For instance, addressees may indicate a lack of comprehension (e.g., "what did you say"?) or ask for clarification of specific points. For example, if a speaker says, "I've been finding a lot of bugs recently," the addressee may ask "what kind of bugs?" In other cases, the listener may think he or she has understood the message, and go on to say something else relevant to the topic (e.g., in terms of the example above, "well why don't you call an exterminator?"). Often the listener will be correct in making this assumption, and the conversation will progress. At times, however, the response the addressee makes may be inappropriate to the message, and the original speaker also uses this form of feedback to assess whether or not common ground has been established. For instance, if the speaker above was initially trying to convey that he or she had found a lot of

computer bugs in his programs recently, then a response suggesting that an exterminator be called would clearly indicate a lack of comprehension.

Although a large part of the grounding process occurs through the audio channel, communicators also make use of physical behaviors in assessing comprehension. If, in the example above, the listener had supplied a can of insect repellent, this act would be as informative about his or her comprehension as a verbal response would be. Thus the ability to observe others' behavioral responses to a message is very helpful in the maintenance of common ground. This suggests that the "virtual" shared physical contexts created by some multimedia conferencing systems should be able to indicate each time a response is made, either in the shared workspace or in others' personal computing environments.

The Problem of Private Conversations. In large meetings, individual participants often speak privately to others who are present. Although sometimes necessary, these private conversations are a threat to the creation and maintenance of common ground. In face-to-face meetings, other participants may not know what these private conversations are about; however, they do know when, and between whom, they are occurring. This information can tell them who may not be up-to-date on the status of the general conversation, or it can alert them to possible dissension in the audience. Although future videoconferencing products are expected to indicate to conferencees when other participants have disengaged themselves from the larger group, it is further necessary, for the maintenance of a shared communicative environment, that they be designed to indicate when the disengaged parties are conferring amongst themselves.

IMPLICATIONS FOR PRODUCT DESIGN

The research discussed in this article describes a number of characteristics of the communicative process which should be considered in the design of multimedia conferencing products and services. Our conceptualization of multimedia conferencing is based heavily on the assumption that users actively attempt to create shared communicative environments for their discussions and activities. We believe that the design of advanced telecommunications products, particularly their user interfaces, must take into account the social and cognitive processes that

underlie such communication. Nonetheless, the realization of these processes requires a number of complex design decisions. In this section we consider several of these issues as they apply in particular to desktop multimedia conferencing systems (see Angiolillo et al., 1993, for discussion of technical issues concerning the transmission of visual information.)

(1) What Information Should be Included in the Visual Field?

We have discussed many reasons why information from facial expressions, eye gaze, gesture, and appearance are important to communication -- to monitor others' comprehension, to assist in the identification of others' category memberships, to coordinate turn-taking, etc. In addition we noted that more than a static representation of one's face alone should be transmitted to others. What is the optimal amount of visual information must be conveyed in real-time live video transmission of attendees? Reasonable limits for the visual field in desktop conferencing systems range from face alone to full body. However there are a number of reasons why full body video is not feasible for most purposes (e.g., the camera optics, video fidelity, and video window size that would be required). Does a head and shoulders view provide the necessary information? Probably not, as illustrated in Figure 2. We suggest that the best visual field for a desktop video system is one that communicates information conveyed by facial expressions and most gestures; i.e., the head, shoulders and arms.

The issue of visual field becomes even more complex when group video systems are included. Too often, the decisions made by product teams regarding these types of questions are driven by considerations unrelated to interpersonal communication (e.g., camera cost and size, video encoding and decoding algorithms). To understand more fully how to create a system that optimizes the transmission of visual field information for a fixed video window size requires further research on the communicative functions of eye gaze, facial expressions, gestures, and appearance. However, based on the concepts discussed in this paper, it seems clear that when the visual field is limited to the head and shoulders only, there is great potential for incomplete or misleading communication (as illustrated in Figure 2). Of course the video parameters of the system (resolution, video board quality, etc.) will dictate to some extent the range of visual field

that is reasonable, but we feel that most currently-available systems support acceptable perception when the visual field is extended beyond the head and shoulders to include information communicated via hand and arm gestures.

In addition to research on the communicative functions of eye gaze, facial expressions, gestures, and appearance, researchers must use measures of communicative success that go beyond those traditionally used in this area. For instance, with the exception of work on computer-based communication, we know of few studies that have examined turn-taking and speaker selection as a function of mode of communication. Work by Brennan and her colleagues (Clark & Brennan, 1990) has addressed this issue as it pertains to the use of current products and services. We believe that rather than focusing on comparisons of communication using audio and audio-visual channels of information, future research should systematically vary the amount, type, and fidelity of information contained in the video feed.

As noted earlier, visual field implications for group systems are much more complex, since group situations are characterized by a greater number of parameters that are free to vary (e.g., number of people, distance from system to people, etc.). Unless the system is being custom-built for a specific group setting, the visual field should be optimized for reasonable viewing of the average number of participants in a group-system room -- about four to six individuals. Note that for group systems reasonable viewing, similar to desktop systems, includes not only heads and shoulders but sufficient body for communication of gestures. It is likely, however, that the most effective multimedia conferencing products, or at least site-specific versions of them, will be those that are custom-built or adapted to the size of the group using the system as well as its purpose. Perhaps for a specific set of users or needs, eye contact with ten colleagues in a group meeting will be more beneficial than the ability to see each speaker's gestures. In contrast, for presentations, lectures, and the like, we might best have the screen filled with a large enough video field to show gesture, dress, etc. Broad visual fields may also be useful for first meeting other people, the times at which social categorization processes

are most important. To examine these hypotheses, studies must also systematically manipulate the number of participants in the discussion and the type of task being performed.

(2) What Size Should A Video Window Be?

Ideally, a video window should be large enough to provide easy viewing of facial expressions and gestures without taking up more screen space than necessary. As illustrated in Figure 4, the nonverbal cues communicated by a given visual display (in this case, a person's facial expression) can be rendered ineffectual by reducing its size below some threshold value. Recall also that eye contact plays important roles in turn-taking, and potential users have reported a desire for eye contact. At what video window size does such eye contact become moot? To a great extent the answer to this question is again dictated by the video parameters (resolution, quality of video board, etc.) of the particular system. A very high-quality video display such as those found in high-end workstations can support eye contact and communicate subtle facial expressions within presentations as small as a postage stamp. More moderate systems require larger video displays. This variance regarding users' personal computing environments highlights a major challenge for designers of desktop video systems -- without being able to control for aspects such as video board, monitor resolution, etc., they are placed in the quandary of having to decide between designing to ensure usability for the lowest common denominator (i.e., for moderate-quality systems) or to provide a high-quality user experience for those with high-end systems.

The problem becomes much more complex when we consider group videoconferences with many attendees. Clearly, current screen size limitations demand that tradeoffs be made between ease of viewing and the number of people viewed. Again, systematic research will be necessary to determine whether minimal eye contact with all fellow conference attendees is superior to greater contact with a selected few, and in what contexts. If the latter turns out to be preferred, we will further need to establish how this greater contact will come about -- through individual changes on one's personal computer, or by switching characteristics of the conferencing system.

(3) How Can Spatial Relationships Between Participants Be Established And Maintained?

We have reviewed work that indicates the importance of gaze and gesture direction on turn-taking and communication. This work strongly suggests that a "virtual" physical arrangement of meeting participants be created that is similar to seating arrangements in a face-to-face conversation. By creating a shared environment of this sort, any individual participant's line of gaze could be seen by all as directed towards a particular member of the group. The HYDRA™ system (Sellen, 1992) is one example of a system that provides remotely located individuals with a spatial arrangement typical of a face-to-face conversation. In this system, which was applied to a three-party multimedia conversation, each participant has a separate monitor for each of the others, and their spatial relationships are maintained such that eye contact between any two of the participants can be seen as such by the third. Unfortunately, the equipment demands of this system render it unreasonable for almost any real-life work scenario, particularly those with more than three participants. The challenge then becomes one of identifying the critical components of the HYDRA™ system that can be carried over into an environment in which each conferee has a single monitor.

One of the most important considerations in this regard is whether multimedia conferencing system should introduce a spatial relationship between attendees and enforce this relationship graphically. The evidence suggests that this will facilitate communication. Yet, how do we deal with the fact that many PC users like to customize their screens by moving, sizing, and minimizing windows, thereby perhaps negating the positive effects provided by the introduction of inter-person spatial relationships? We suggest that the benefits of fixed spatial orientation will outweigh its costs, but research is necessary to evaluate this claim.

One promising step in the direction of providing a framework that supports a common spatial relationship between participants is an on-screen meeting participant manager based on the "meeting room" metaphor (Benimoff et al., 1995). The meeting room metaphor, which takes advantage of people's existing knowledge about how face-to-face meetings operate and about the properties of real-world objects, provides an on-screen presence that looks like a meeting room.

Key to this user interface is the graphical representation of all meeting participants placed at positions around a table. Across participants' screens the placement of individuals is identical, thus providing a shared sense of spatial relationships. One can imagine extending the meeting room concept by enforcing specific roles to different seat locations; e.g., the seat at the head of the table would have capabilities associated with the meeting leader. Again, more study is required to ascertain the extent to which the meeting room metaphor is useful, especially when multi-participant communications other than prototypical meetings, such as broadcast presentations or distance education efforts, are considered.

(4) How Can We Reduce Mismatches Between The Assumptions Individuals Make About What Others Are Viewing and What They Are Actually Viewing?

We have discussed a number of ways in which a videoconference attendee might make erroneous assumptions about what is in the field of view of the other attendees. Consider an application common to PC-based CSCW systems -- a shared, on-screen whiteboard, viewed simultaneously by all participants, that allows them to make revisions, draw figures, etc. Like many applications presented within a window, it is resizable, moveable, and minimizable. A user of this system may make notations on his or her own whiteboard, and reasonably assume that the others see these notations, just as in face-to-face meetings markings made on a whiteboard are seen in all in attendance. However, as was illustrated in Figure 3, the others on the electronic conference, each of whom has the whiteboard application running, may take advantage of the application's window management capabilities (resizing, moving, etc.) and optimize their own screen layout. Thus, Person A will incorrectly assume that Person B is viewing what is being written, and Person B will incorrectly assume that Person A is viewing specific data, and the basis of successful communication will be compromised.

The potential for such mismatches between what is understood to be viewed and what is actually viewed is expanded when we consider more advanced electronic whiteboards that have multiple layers, or "pages," and individual users can be working on separate layers at the same time. We suggest that there are at least two acceptable alternative design strategies for such an

application to minimize the occurrences of these sorts of mismatches -- either enforce the principle that all individuals must view exactly the same whiteboard display (e.g., temporarily disable window resize) or provide supporting information to all participants informing them of what is on the screen at the other locations. The latter solution is similar to that employed by Intel's ProShare™ Personal Video System, a person-to-person conferencing product that provides a shared electronic whiteboard in the form of a multi-page notebook. Two people may be on different pages of the notebook, but there is always an indication of what page the other is on. This implementation, which works well in the person-to-person product, is not so easily extended to the multi-participant world due to the large amount of information that would need to be displayed and monitored by each user. Creative screen designs will be necessary in order to ensure shared "virtual" environments in multimedia systems.

(5) How Should Systems Deal With Workspaces That Are Only Partially Shared?

A related design decision revolves around situations in which some participants are not able to attend a meeting using the full system capabilities; i.e., via all available media. As we discussed earlier, attendees who must participate by voice only are likely to have less salience, and to be able to provide less feedback, than those who attend in a full audio-video-graphics mode. We also discussed the problems that arise from this discrepancy, such as differences in perceived contributions to the discussion, and limitations in the extent to which speakers can adapt their messages to the attendee's comprehension. On the one hand, these considerations suggest a "least common denominator" approach: Group conferences should be established such that all participate equally. For instance, if some participants do not have video access but are able to share documents and audio, then the conference should be held in audiographics mode. On the other hand, it is unlikely that either vendors or users of advanced videoconferencing products will find this solution satisfactory. We suggest that the resolution of this issue will involve careful consideration of the tasks being performed, the degree of acquaintance between conference attendees, and the distribution of equipment across users (i.e., how many individuals lack the video link). If individuals are allowed to participate in a meeting with different media

capabilities, then it must be clear to all participants the media with which all other participants are attending. This should minimize, though probably not entirely eliminate, miscommunications based, for example, on assumptions that everyone has access to points made on a shared whiteboard when there may be voice-only or voice-video attendees present.

(6) How Can We Facilitate the Construction of a Shared Social Context?

We have argued extensively that successful interpersonal communication requires that participants take one another's background knowledge, attitudes, etc. into account. We have also noted that many inferences about others' perspectives are derived from assumptions about their social and work-related category memberships. Given that much of the information that aids conversationalists in identifying these category memberships is visual in nature, basic considerations in the design of multimedia systems concern how much of this visual information can be presented, and how. As noted above, this decision may involve a trade-off between the presentation of category cues such as style of dress and transmission of eye-gaze and facial expressions. To determine what configuration will work best for a given conferencing system, careful attention must be paid to the costs of errors in each of these domains.

(7) Can We Improve on Face-To-Face Communication?

A final issue, and one that is beyond the scope of the current article, is that of the possibility of enhancing communication beyond face-to-face conversation. A number of discussion-enhancing features have been suggested in the literature. For example, there is the possibility of reviewing on-line video recordings of the preceding parts of the conversation. While this feature may be useful for those who join a meeting in progress, because it allows them to catch up on what they have missed, it may also detract attention from the ongoing discussion. On the other hand, review of previous statements might be beneficial in tasks such as group decision making. A second example is providing a conferencing feature that allows side-conferences to occur. Participants may wish to temporarily "leave the meeting" and hold a private conversation with another individual, and they may desire not to have their temporary absence known. On the one hand, this feature may meet user needs. On the other hand,

however, unless designed to inform meeting attendees when particular participants have temporarily left, there is great potential for miscommunications based on the assumption that all parties are in attendance and have access to presented information. Are these new features enhancements or potentially negative alterations of normal communicative processes? The research reviewed throughout this paper clearly suggests that such features not be incorporated into a videoconferencing system without extensive examination on their effects on the dynamics of interpersonal communication.

Summary of Design Implications

In this section we outlined several design implications, trade-offs, and guidelines that ought to be considered in the creation of a multimedia collaboration system. Specific design requirements, while beyond the scope of this article, are in addition problematic to discuss completely since most are best considered with respect to specific systems and in specific computing and equipment environments. Plus, as we noted there is a great deal of research yet to be conducted on a variety of issues that will provide us some of the answers we seek. In the absence of such knowledge, the clear method for proceeding with the creation of any multimedia collaboration system would have to include thorough systematic usability testing with people who are expected to be typical users of the system. In addition, the advantages promised by the usability testing program will be enhanced if it can be conducted with users in real-world settings, so that their real-world needs, capabilities, limitations, etc. can be identified and measured.

CONCLUSIONS

To date, decisions about features of advanced telecommunications products and services have been driven heavily by the underlying technological capabilities of these systems. When the current and future users of these products and services are considered at all, it is usually by means of customer surveys rather than experimentation. To what extent should we rely on user surveys to answer questions about interface design? Casual observation suggests that people are unaware of the impact of various communication channels on their responses (Gale, 1991).

Indeed, people have limited access to *most* of their own cognitive processes (Nisbett & Wilson, 1977). Nonetheless, product development teams frequently solicit data from potential users regarding the desirability of particular features. We would argue that this is the wrong strategy to take in constructing successful multimedia conferencing products, and it can lead developers of these products astray. For example, some users of conferencing systems have expressed the desire to have a clear view of the other participants' eyes, on the assumption that they will be able to detect whether or not someone is lying from this information. Unfortunately, research suggests that this is not the case (e.g., Kraut, 1973).

While acknowledging that marketing considerations are important, we suggest that developers of new telecommunications technology focus on the actual features that facilitate the construction of shared communicative environments, and then work to persuade potential users that it is these features that they need to accomplish their work-related and interpersonal goals. In short, we believe that products and services should be designed foremostly for successful interpersonal communication. By doing so, marketplace success is virtually guaranteed.

We tend to disagree with those who argue that video teleconferencing is not in line to replace face-to-face meetings (e.g., Egido, 1990). Rather, we believe that by incorporating features that facilitate people's ordinary communicative processes we can, in effect, create a medium that is the "next best thing to being there." To develop such products and services, however, careful attention will need to be paid to the actual dynamics of interpersonal communication.

Acknowledgments

We would like to thank Robert M. Krauss, Amir M. Mane, Judy S. Olson, William B. Whitten II, and two anonymous reviewers for their thoughtful comments on a previous version of this manuscript, and Edmond W. Israelski and Max S. Schoeffler, guest editors, for their extensive comments, advice, and patience.

REFERENCES

Anderson, R. J., Heath, C. C., Luff, P., & Moran, T. P. (1993). The social and the cognitive in human-computer interaction. International Journal of Man-Machine Studies, 38, 999-1016.

Angiolillo, J. S., Blanchard, H. E., & Israelski, E. W. (1993). Video telephony. AT&T Technical Journal, 72(3), 7-20.

Argyle, M., & Cook, M. (1976). Gaze and mutual gaze. London: Cambridge University Press.

Argyle, M., Lalljee, M., & Cook, M. (1968). The effects of visibility on interaction in a dyad. Human Relations, 21, 3-17.

Austin, J. L. (1962/1975). How to do things with words (2 ed.). Cambridge, MA: Harvard University Press.

Benimoff, N. I., Altom, M. W., Farber, J. M., Kirby, D. J., Mane, A. M., Montero, R. C., Pastore, R. L., Roberts, L. A., Sauer, R. F., Todd, S., andWhitten, W. B. II (1995). A user interface design for desktop multimedia collaboration. Proceedings of the Human Factors in Telecommunications International Symposium, Melbourne, Australia, March 6-10, 1995, 21-25.

Benimoff, N. I., & Burns, M. J. (1993). Multimedia user interfaces for telecommunications products and services. AT&T Technical Journal, 72(3), 42-49.

Benimoff, N. I., & Whitten, W. B. I. (1993). Human factors issues in multimedia conferencing. Proceedings of the Human Factors in Telecommunications International Symposium, Darmstadt, Germany, May 11-14, 1993, 227-236.

Bilous, F. R. (1992) The role of gestures in speech production: Gestures enhance lexical access. Unpublished doctoral dissertation, Department of Psychology, Columbia University.

Boland, R. J. J., Maheshwari, A. K., Te'enl, D., Schwartz, D. G., & Tenkasi, R. V. (1992). Sharing perspectives in distributed decision making. Proceedings of the 1992 Conference on Computer-Supported Cooperative Work, Oct. 31-Nov. 4, Toronto, Canada.

Boyle, E. A., Anderson, A. H., & Newlands, A. (1994). The effects of visibility on dialogue and performance in a cooperative problem solving task. Language and Speech, 37, 1-20.

Brennan, S. E. (1991). Conversation with and through computers. User modeling and user-adapted interaction, 1, 67-86.

Butterworth, B. (1980). Evidence from pauses in speech. In B. Butterworth (Eds.), Speech and Talk. London: Academic Press.

Carroll, J. M. (1985). What's in a name? New York: Freeman.

Clark, H. H. (1987). Four dimensions of language use. In J. V. &. M. Bertuccelli-Papi (Eds.), The Pragmatic Perspective: Selected Papers from the 1985 International Pragmatics Conference Amsterdam/Philadelphia: John Benjamins.

Clark, H. H. (1985). Language use and language users. In G. Lindzey & E. Aronson (Eds.), Handbook of social psychology (pp. 179-231). New York: Random House.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), Perspectives on socially shared cognition Washington, DC: APA.

Clark, H. H., & Marshall, C. E. (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. L. Webber, & I. A. Sag (Eds.), Elements of discourse understanding (pp. 10-63). Cambridge: Cambridge University Press.

Clark, H. H., & Schaefer, E. F. (1987). Concealing one's meaning from overhearers. Journal of Memory and Language, 26, 209-225.

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

Clark, W. J. (1992). Multipoint Multimedia Conferencing. IEEE Communications Magazine, 30, (5), 44-50.

CSCW 92: Sharing Perspectives. (1992) (J. Turner & R. Kraut, Eds.). New York: ACM Press.

CSCW '94: Transcending Boundaries. (1994) (R. Furuta & C. Neuwirth, Eds.) New York: ACM Press.

Duncan, S., & Fiske, D. (1977). Face-to-face interaction: Research, methods, and theory. Hillsdale, NJ: Erlbaum.

Egido, C. (1990). Teleconferencing as a technology to support cooperative work: Its possibilities and limitations. In J. Galegher, R. E. Kraut, & C. Egido (Eds.), Intellectual teamwork: Social and technological foundations of cooperative work (pp. 351-371). Hillsdale, NJ: Erlbaum.

Ekman, P., & Friesen, W. V. (1969). The repertoire of nonverbal communication: Categories, origins, usage, and coding. Semiotica, 1, 49-98.

Feldman, R. S., & Rime, B. (Ed.). (1991). Fundamentals of nonverbal behavior. New York: Cambridge University Press.

Festinger, L. (1950). Informal social communication. Psychological Review, 57, 271-282.

Fish, R. S., Kraut, R. E., Root, R. W., & Rice, R. E. (1992). Evaluating video as a technology for informal communication. In Proceedings of the Conference on Computer Human Interaction (CHI) '92 (pp. 37-48). Monterey, CA:

Fiske, S. T., & Taylor, S. E. (1991). Social cognition (2nd ed.). New York: McGraw-Hill.

Furnas, G. W., Landauer, T. K., Gomez, L. M., & Dumais, S. T. (1987). The vocabulary problem in human-system communication. Communications of the ACM, 30, 964-971.

Fussell, S. R., & Krauss, R. M. (1989). The effects of intended audience on message production and comprehension: Reference in a common ground framework. Journal of Experimental Social Psychology, 25, 203-219.

Fussell, S. R., & Krauss, R. M. (1991). Accuracy and bias in estimates of others' knowledge. European Journal of Social Psychology, 21, 445-454.

Fussell, S. R., & Krauss, R. M. (1992). Coordination of knowledge in communication: Effects of speakers' assumptions about what others know. Journal of Personality and Social Psychology, 62, 378-391.

Gabarro, J. J. (1990). The development of working relationships. In J. Galegher, R. E. Kraut, & C. Egido (Eds.), Intellectual teamwork: Social and technical bases of collaborative work (pp. 79-110). Hillsdale, NJ: Erlbaum.

Gale, S. (1991). Adding audio and video to an office environment. In J. Bowers & S. Benford (Eds.), Studies in Computer-Supported Cooperative Work: Theory, Practice and Design (pp. 49-62). Amsterdam: North-Holland.

Galegher, J., Kraut, R. E., & Egido, C. (Eds.). (1990). Intellectual teamwork: Social and technological foundations of cooperative work. Hillsdale, NJ: Erlbaum.

Greif, I. (1988). Computer-supported cooperative work: A book of readings. San Mateo, CA: Morgan Kaufmann.

Grice, H. P. (1957). Meaning. Philosophical Review, 64, 377-388.

Grice, H. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), Syntax and semantics: Speech acts. New York: Academic Press.

Heath, C., & Luff, P. (1991). Disembodied conduct: Communication through video in a multi-media office environment. In Proceedings of the Conference on Computer Human Interaction (CHI) '91 (pp. 99-103). New Orleans, LA:

Higgins, E. T. (1992). Achieving 'shared reality' in the communication game: A social action that creates meaning. Journal of Language and Social Psychology, 11, 107-125.

Hindus, D., & Brennan, S. (1992). Conversational paradigms in user interfaces. Tutorial presented at the ACM Conference on Human Factors in Computing Systems (CHI), Monterey, CA, May 3-7, 1992.

Hutchins, E. (1990). The technology of team navigation. In J. Galegher, R. E. Kraut, & C. Egido (Eds.), Intellectual teamwork: Social and technical bases of collaborative work (pp. 191-220). Hillsdale, NJ: Erlbaum.

Isaacs, E. A., & Clark, H. H. (1987). References in conversation between experts and novices. Journal of Experimental Psychology: General, 116(26-37).

Ishii, H., Kobayashi, M., & Grudin, J. (1992). Integration of inter-personal space and shared workspace: Clearboard design and experiments., Proceedings of CSCW'92. New York: ACM.

Kendon, A. (1967). Some functions of gaze direction in social interaction. Acta Psychologica, 32, 1-25.

Kiesler, S., & Sproull, L. (1992). Group decision making and communication technology. Organizational Behavior and Human Decision Processes, 52, 96-123.

Kiesler, S., Zubrow, D., Moses, A. M., & Geller, V. (1985). Affect in computer-mediated communication: An experiment in synchronous terminal-to-terminal discussion. Human-Computer Interaction, 1, 77-104.

Kiesler, S., Siegel, J., & McGuire, T. W. (1984). Social psychological aspects of computer-mediated communication. American Psychologist, 1123-1134.

Krauss, R. M. (1979). Communicative models and communicative behavior. In R. Schiefelbusch & J. Hollis (Eds.), Language Intervention from Ape to Child. University Park Press.

Krauss, R. M., & Bricker, P. D. (1966). Effects of transmission delay and access delay on the efficiency of verbal communication. Journal of the Acoustical Society of America, 41, 286-292.

Krauss, R. M., & Fussell, S. R. (1988). Other-relatedness in language processing: Discussion and comments. Journal of Language and Social Psychology, 7, 263-279.

Krauss, R. M., & Fussell, S. R. (1990). Mutual knowledge and communicative effectiveness. In J. Galegher, R. E. Kraut, & C. Egido (Eds.), Intellectual teamwork: Social and technical bases of collaborative work Hillsdale, NJ: Erlbaum.

Krauss, R. M., & Fussell, S. R. (1991a). Constructing shared communicative environments. In L. Resnick, J. Levine, & S. Teasley (Eds.), Perspectives on socially shared cognition. Washington, DC: American Psychological Association.

Krauss, R. M., & Fussell, S. R. (1991b). Perspective-taking in communication: Representations of others' knowledge in reference. Social Cognition, 9, 2-24.

Krauss, R. M., & Fussell, S. R. (in press). Social psychological models of interpersonal communication. To be published in E. T. Higgins & A. Kruglanski (Eds.), Social Psychology: Handbook of Basic Principles. New York: Guilford Press.

Krauss, R. M., Garlock, C. M., Bricker, P. D., & McMahon, L. E. (1977). The role of audible and visible back channel responses in interpersonal communication. Journal of Personality and Social Psychology, 35, 523-529.

Krauss, R. M., Morrel-Samuals, P., & Colasante, C. (1991). Do conversational hand gestures communicate? Journal of Personality and Social Psychology, 61, 743-754.

Krauss, R. M., & Weinheimer, S. (1966). Concurrent feedback, confirmation and the encoding of referents in verbal communication. Journal of Personality and Social Psychology, 4, 343-346.

Kraut, R. E., Egido, C., & Galegher, J. (1990). Patterns of contact and communication in scientific research collaborations. In J. Galegher, R. E. Kraut, & C. Egido (Eds.), Intellectual teamwork: Social and technological foundations of cooperative work (pp. 149-171). Hillsdale, NJ: Erlbaum.

Kraut, R. E., Galegher, J., & Egido, C. (1987-1988). Relationships and tasks in scientific research collaboration. Human-Computer Interaction, 3, 31-58.

Kraut, R. E., & Lewis, S. H. (1982). Feedback and the coordination of conversation. In H. Sypher & J. Applegate (Eds.), Cognition and communication. Hillsdale, NJ.: Erlbaum.

Kunda, Z., & Nisbett, R. E. (1986). The psychometrics of everyday life. Cognitive Psychology, 18, 195-224.

Levelt, W. J. M. (1989). Speaking: From Intention to Articulation. Cambridge, MA: The MIT Press.

Luff, P., Gilbert, N., & Frohlich, D. (Eds.). (1990). Conversation and Computing. London: Academic Press.

McGrath, J. E. (1990). Time matters in groups. In J. Galegher, R. E. Kraut, & C. Egido (Eds.), Intellectual teamwork: Social and technical bases of collaborative work (pp. 23-61). Hillsdale, NJ: Erlbaum.

McNeill, D. (1992). Hand and mind: What gestures reveal about thought. Chicago: University of Chicago Press.

Mead, G. H. (1934). Mind, self and society. Chicago: University of Chicago Press.

Morrel-Samuels, P., & Krauss, R. M. (1992). Word familiarity predicts temporal asynchrony of hand gestures and speech. Journal of Experimental Psychology: Learning, Memory and Cognition, 18, 615-623.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. Psychological Review, 84, 231-259.

Okada, K., Maeda, F., Ichikawaa, Y., & Matsushita, Y. (1994). Multiparty videoconferencing at virtual social distance: MAJIC design. In R. Furuta & C. Neuwirth (Eds.), Proceedings of the ACM 1994 Conference on Computer Supported Cooperative Work. New York: Association for Computing Machinery.

Olson, G., & Olson, J. (1991). User-centered design of collaboration technology. Journal of Organizational Computing, 1, 61-83.

Resnick, L. B., Levine, J. M., & Teasley, S. D. (1991). Perspectives on socially shared cognition. Washington, DC: American Psychological Association.

Root, R. W. (1988). Design of a multi-media vehicle for social browsing. Proceedings of the ACM 1988 Conference on Computer-Supported Cooperative Work, New York.

Rutter, D. (1987). Communicating by telephone. Oxford, Eng.: Pergamon Press.

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

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. Language, 50, 696-735.

Scherer, K. R., & Ekman, P. (Eds.). (1982). Handbook of methods in nonverbal behavior research. Cambridge: Cambridge University Press.

Schober, M. F. (1993). Spatial perspective-taking in conversation. Cognition, 47, 1-24.

Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. Cognitive Psychology, 21, 211-232.

Searle, J. R. (1969). Speech acts: An essay in the philosophy of language. Cambridge: Cambridge University Press.

Sellen, A. J. (1992). Speech patterns in video-mediated conversations. Proceedings of ACM Conference on Human Factors in Computing Systems, Monterey, CA, May 3-7, 1992, 49-59.

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

Siegman, A. W., & Feldstein, S. (Ed.). (1987). Nonverbal behavior and communication, 2nd ed. Hillsdale, NJ: Erlbaum.

Tang, J. C., & Isaacs, E. (1993). Why do users like video? Studies of multimedia-supported collaboration. Computer-Supported Cooperative Work, 1, 163-196.

Tannen, D. (1990). You just don't understand: Women and men in conversation. New York: Ballantine Books.

Taylor, S. E., & Fiske, S. T. (1975). Point-of-view and perceptions of causality. Journal of Personality and Social Psychology, 32, 439-445.

Taylor, S. E., & Fiske, S. T. (1978). Salience, attention, and attribution: Top of the head phenomena. In L. Berkowitz (Ed.), Advances in Experimental Social Psychology (Vol. 11) (pp. 249-288). New York: Academic Press.

Wilkes-Gibbs, D. & Clark, H. H. (1992). Coordinating beliefs in conversation. Journal of Memory and Language, 31, 183-194.

Winograd, T. (1987-1988). A language/action perspective on the design of cooperative work. Human-Computer Interaction, 3, 3-30.

Wyer, R. S. J., & Srull, T. K. (Eds.) (1994). Handbook of social cognition, (2nd ed.). Hillsdale, NJ: Erlbaum.

LIST OF FIGURES

SUSAN R. FUSSELL and NICHOLAS I. BENIMOFF (Social and Cognitive Processes in Interpersonal Communication: Implications for Advanced Telecommunications Technologies)

SUSAN R. FUSSELL received her Ph.D. in Social and Cognitive Psychology from Columbia University in 1990. From 1987-1988 she worked as a research assistant at AT&T Bell Laboratories in Lincroft, New Jersey. After receiving her degree, she spent three years as a post-doctoral fellow in the Department of Psychology at Princeton University. Since 1993 she has been an Assistant Professor of Psychology at Mississippi State University. She has published a number of articles and chapters in the area of interpersonal communication. Her current research interests include multi-party conversations, the communication of emotional states, and computer-supported cooperative work.

NICHOLAS I. BENIMOFF received his Ph.D. in Experimental Psychology from Columbia University in 1985. Since 1986 he has been employed as a member of the technical staff at AT&T Bell Laboratories in Holmdel, New Jersey. He has worked on a variety of projects at Bell Labs including automatic speech recognition systems, multimedia and graphical user interface design, and voice messaging services. He is currently responsible for user interface design and evaluation for AT&T WORLDWORX SM Personal Conferencing Service, an AT&T service that supports desktop multimedia collaboration.