Towards a Theory of Mediated Communication

Steve Whittaker
Information Management Laboratory
HP Laboratories Bristol
HPL-91-78
June, 1991

interpersonal communication, media, shared workspaces

CSCW currently lacks a theory of mediated communication. A number of empirical studies have been conducted which detail the effects on communication of using a variety of different media, but we do not have clear explanation for these effects. This paper attempts to show how an account of mediated communication can be informed by an analysis of face-to -face interaction.

Internal Accession Date Only

1 Introduction

CSCW researchers are being presented with two sets of new problems. First there are changes in the nature of work: we have seen the emergence of distributed groups collaborating on joint projects while working at different geographical locations. Second is the development of new communication devices such as FAX, video, cellular phone, and portable computers as well as new media such as digital video, audio, image, and animation. The question this raises is how these new devices and media can be combined to support the interactions of the new distributed work groups.

The claim here is that we require a general theory of mediated interaction which makes specific predictions about communication using these different devices and media. It is clear that intuitions about communication devices are not enough: the failure of products such as the Picturephone developed by AT&T, and the more recent video work carried out by Xerox and AT&T, indicates this[GA86, HL91, Stu86, Roo88]. In both cases the intuition was that broadening the bandwidth of communication would improve its efficiency but in neither case was this satisfactorily demonstrated. While there has been much empirical work comparing the effects of different modes on effectiveness of communication [Cha75, SWC76], this work has not succeeded in producing a coherent theory. These studies showed quite clearly that simply broadening the bandwidth of communication by adding visual information to the speech channel does not always increase the efficiency of communication. The work also showed that for information access tasks, speech is by far the most efficient communication medium. However this work gives no theoretical explanation either for the success of speech, or the failure of visual information to improve communication. I offer answers to both these questions here.

An alternative to the empirical approach has been to try and understand the fundamental processes of human communication and then to see how these are achieved with different communication media[OC89a, CB90, WBC91]. One problem here is that much of the theoretical work on human communication has been based on a very specific type of interaction. The data on which these theories are based is face-to-face interaction in which the communication is supported by media such as speech and gesture in a visually shared environment. The focus on these media may mean that the current theories contain features that are specific to those media. The aim here will be to try and factor out this specificity, so that the general principles can be applied to situations involving media other than speech, gesture and direct visual contact. In fact one of the arguments we will advance here is that the nature of communication depends critically on the precise characteristics of the available media.

In what follows, I will (a) give a brief account of face-to-face interaction; (b) describe two studies in which the media differed from the face-to-face situation, explaining how the different media changed the nature of the communication; (c) talk about how features of the situation interact with media and devices, showing how one feature, physical situation, constrains what media can be used to communicate and how this influences the nature of communication.

2 An account of human communication

What then are the critical features of human communication? Communication is a joint activity which requires co-ordination of both process and content[CWG86]. Co-ordinating

have been understood, and that listeners give frequent evidence that they understand the utterances. Appropriate action can therefore be taken by either person to coordinate content if they detect a discrepancy in their respective beliefs. Kraut et el also stress the critical role of feedback in maintaining shared understanding[KLS82].

This is how co-ordination is achieved in human interaction, but to what extent does co-ordination depend on the specific media that are used in face-to-face communication, namely speech, gesture and shared visual environment? Speech has a number of key properties that support the co-ordination of both process and content. It can be produced and understood easily and quickly by native adult speakers and it also has a fast transmission time. These properties mean that speech can support phenomena like backchannels, clarifications and interruptions which depend critically on very precise timing. Another crucial feature of speech is that it enables two way (duplex) communication: thus one participant can backchannel to indicate affirmation of a given utterance and this act can occur while the speaker is still in the process of producing the utterance. This concurrency of feedback, and ease of achieving interruptions, mean that divergences from shared understanding can be detected and corrected quickly. So the incremental and interactive aspects of conversation are facilitated by speech. The visual medium also contributes to the co-ordination of process and content. The fact that participants share a physical environment enables them to achieve joint attention, and it also enables participants to monitor the attention of others (eg what people, objects and events someone else has observed). In addition eye-gaze and head nods can contribute to the achievement of smooth transitions between speakers and give feedback to speakers that their messages are being understood and accepted. In conclusion, speech and gesture in a shared visual environment seem to be a highly efficient solution to the problems of face-to-face communication. The solution to these problems seems, however, to depend critically on the properties of the specific media available, namely that the media jointly support duplex communication across several channels concurrently, with very short transmission times.

3 Two studies of mediated communication

We now apply this framework to two studies of communication where the media are not speech, gesture and the shared physical environment. What happens to co-ordination if we change the properties of the communication media?

3.1 Communication using shared workspace and audio

We conducted a study to look at the nature of communication when media other than speech are used and when participants are not co-present[WBC91]. Speech and gesture are ephemeral so we first examined the effect of using exclusively permanent media for communication. In our first experiment, we looked at interaction when the sole means of communication was by typing, writing or drawing on a shared electronic Whiteboard. Participants were at different physical locations, so they could not communicate using speech or gesture. The tasks that we gave them were brainstorming and calendar coordination. This meant that all participants had information to contribute to the successful completion of these tasks[WS88, WW90] We found that communication using permanent media differed from face-to-face communication in four ways. First, permanent media did not require the serial unfolding of topics that characterises speech[Lev83]: Contributions persisted, so participant did not need reply to another person's input immediately, because they knew that the input would not disappear. This led to more parallel activity with permanent

while interactions with these characteristics can occur for reasons other than those listed below[1], the properties of the media in video-conferencing preclude interactions of a given type.

How can we explain these effects? There are a number of ways in which the media used in video-conferencing differ from those in face-to-face communication.

- The audio and video channels have a transmission time approaching 250ms. This is because the video has to be compressed and decompressed and the audio is buffered so that it can be presented in synchrony with the video. This means that the feedback that the listener gives is delayed and hence often disruptive to the speaker. It also means that successful interruptions or clarifying questions are difficult because they require precise timing.
- The audio channel is half-duplex and so only one person can speak at any time. This means that if listeners produce audio backchannels ("mm"), then this information takes over the channel. This means that concurrent audio feedback is disruptive. In addition, the audio channel can be taken over by accidental noises such as sneezing and coughing, without the speaker being aware of what was deleted by these noises.
- The quality of the video leads participants to complain that they cannot identify who is speaking. It also means that it is very difficult for speakers to detect visual feedback from listeners.

The fact that there is much reduced verbal and visual feedback means that speakers are much less able to determine whether listeners understand what they are saying. They therefore do not make the types of adjustments that speakers normally make when they perceive that they are not being understood. Listeners also cannot easily interrupt to ask questions or clarify when they do not understand because the problems of timing and the half-duplex channel make interruptions highly disruptive. Both lack of feedback to speakers and the inability to interrupt, contribute to the problems of non-interactivity, unidirectionality and reduced mutual understanding. The difficulty of interrupting may also explain why meetings are thought to be superficial because it is difficult to achieve the incremental exchanges that are necessary to completely understand what the speaker is saying. It also contributes to the agenda-based format of these meetings: the cost of interruptions means that conversation is difficult to divert from the agenda with the opportunistic interruptions that normally occur in face-to-face meetings.

People also report that video-conferences are somewhat confrontational with feelings of "sides" and "us and them" for the different ends of the conference link and that conversations seemed to be dominated by two chairpeople, one at each end of the link. The problem here is that while local conversations are easy to manage, conversations with the remote location are difficult. The problems of multiple speakers competing to converse over the remote link are resolved by directing conversations locally to the chair, who is then responsible for relaying these across the remote link. This feeling of "sides" is then exacerbated by the fact that listeners do not always orient to the speaker when that speaker is local: they tend to stare at the remote monitor. This contributes to listener's difficulty in identifying the remote speaker and also to the feeling that the local listeners are monitoring the remote ones rather

[1] Indeed some of these problems are said to occur in most face-to-face formal meetings[FKC90]

more verbose (and hence less incremental) messages. Structured messaging is an attempt to provide at least some of this context[MMM91, MGL+87, WF86].

What is the result of all this? The fact that incremental exchanges cannot easily be achieved in asynchronous communication means a reduction in the frequency with which the listener gives and the speaker receives feedback. This combined with the reduced linguistic context, detracts from mutual understanding. Another problem with asynchronous situations is that there is a possibility that messages can lose their sequence due to variable transmission times: with synchronous communication this cannot occur because each message occurs in real time[CB90]. A final problem with asynchronous communication lies in knowing what to infer from a lack of response. The failure to obtain a response may result from reasons as diverse as the message never arriving, to the fact that the recipient does not like the content of the message and is spending some time constructing a tactful reply. The existence of feedback in face-to-face interaction can resolve some of these uncertainties.

In aynchronous communication, the lag between inputs and the lack of linguistic context, means that co-ordination of content and process is more difficult to achieve. There are however, some advantages to asynchronous communication: participants are freed from the pressure to produce messages in real time, they can thus edit their message before they produce it, and deliberate over and review other people's messages before responding. In addition it is possible with aynchronous communication to hold a number of different conversations concurrently, and the participants have a permanent record of all messages. As we have already seen, synchronous communications tend to be less concurrent and they also generate no permanent record.

4.2 The spatial dimension: Co-presence

The second critical characteristic of the physical situation is whether or not the participants are co-present. If they are co-present then this means that they have access to a shared physical environment, with the possibility of joint attention, non-verbal information, and affective communication. Furthermore they should be able to communicate using media such as speech and gesture. Despite the ease of communication with these media, they do have limitations, as we have seen: they do not leave a permanent record, nor is there the possibility of parallel communication, and there is also the pressure to produce quickly and respond immediately. Systems have been designed to overcome these limitations in face-to-face interaction[SFB+88]. However, care should be taken with the design of the interface to such systems in order that using the system does not disrupt people's ability to participate in face-to-face interaction. Some systems that attempt to supplement face-to-face meetings in this way, have suffered from the problem of disruption[TFB90].

What do participants lose when they are co-temporal but not co-present? Clearly they lose the ability to monitor attention, to achieve joint attention, and to refer to things in the physical environment by pointing. Once they are beyond earshot and out of sight, they also have to find some device to transmit their messages. Media like speech have been successfully supported by a pervasive device like the telephone, to support synchronous distributed communication. High quality telephone clearly allows incremental interaction, provided it is not used over huge distances, because large distances produce time lags and consequent disruptions of communication. Attempts to provide a visual analogy to the telephone for people who are not co-present have not met with great success however. One example of a failed attempt is the Picturephone, but there have been more recent attempts

mediated communication. I have also argued that the physical situation is a major influence on communication. What are the broader implications of this? I have specified a number of underlying media factors which determine that asynchronous interaction can never be incremental. This has implications for what communication situations we should try and support asynchronously. It seems that asynchronous communication is inappropriate for various types of activity that rely on negotiation or shared meaning because these require incremental communication. For synchronous situations in which partipants are not co-present, incremental interaction should be possible, but specific types of visual information such as eyegaze and gesture may be difficult to transmit. We have also seen from the analysis of video-conferencing that certain aspects of people's visual behaviour in this setting may be misleading for remote participants. However, there may be a substantial number of tasks that do not require this type of dynamic visual information and these may be efficiently carried out in the absence of video.

While I have focussed on the interaction between physical situation and medium here, it is clear that there are other factors that influence the communication, including the goals of the participants and their number, status and knowledge. A more complete account would have to show how these combine with physical situation to generate communication demands. The object will then be to select appropriate combinations of media and devices to meet those demands.

6 References

[BFP87] Susan E. Brennan, Marilyn Walker Friedman, and Carl J. Pollard. A centering approach to pronouns. In *Proc. 25th Annual Meeting of the ACL*, pages 155–162, 1987.

[Bly88] Sara S. Bly. A use of drawing surfaces in different collaborative settings. In *Proceedings of the Conference on CSCW*, 1988.

[CB90] Herbert H. Clark and Susan E. Brennan. Grounding in communication. In L. B. Resnick, J. Levine, and S. D. Bahrend, editors, *Perspectives on socially shared cognition.* APA, 1990.

[Cha75] A. Chapanis. Interactive human communication. *Scientific American*, 232:34–42, 1975.

[CM81] Herbert H. Clark and Catherine R. Marshall. Definite reference and mutual knowledge. In Aravind K. Joshi, Bonnie Lynn Webber, and Ivan Sag, editors, *Elements of Discourse Understanding*, pages 10–63. Cambridge University Press, Cambridge, 1981.

[CWG86] Herbert H. Clark and Deanna Wilkes-Gibbs. Referring as a collaborative process. *Cognition*, 22:1–39, 1986.

[Egi88] Carmen Egido. Video-conferencing as a technology to support group work: A review of its failures. In *Second conference on Computer Supported Co-operative Work*, pages 13–24, 1988.

[FKC90] Robert Fish, Robert Kraut, and Barbara Chalfonte. The videowindow system in informal communication. In *Proceedings of the Conference on Computer Supported Co-operative Work*, pages 1–12, 1990.

[Roo88] R. Root. Design of a multi-media vehicle for social browsing. In *Proceedings of the Conference on Computer Supported Co-operative Work*, 1988.

[SFB+88] Mark Stefik, Gregg Foster, Daniel Bobrow, Kenneth Kahn, Stan Lanning, and Lucy Suchman. Beyond the chalkboard: Computer support for collaboration and problem solving in meetings. In Irene Greif, editor, *Computer-Supported Co-operative Work*. Morgan Kaufmann, San Mateo, Ca., 1988.

[Sid81] Candace L. Sidner. Focusing for interpretation of pronouns. *American Journal of Computational Linguistics*, 7(4):pp. 217–231, 1981.

[SSJ74] Harvey Sacks, Emmanuel Schegloff, and Gail Jefferson. A simplest systematics for the organization of turn-taking in conversation. *Language*, 50:pp. 325–345, 1974.

[Stu86] R. Stults. Media space. Technical report, Xerox PARC, 1986.

[SWC76] J. Short, E. Williams, and B. Christie. *The Social Psychology of Telecommunications*. Wiley, London, 1976.

[TFB90] Deborah Tatar, Gregg Foster, and Daniel Bobrow. Design for conversation: Lessons from cognoter. *International Journal of Man-Machine Studies*, 1990.

[TL88] John Tang and Larry Leifer. A framework for understanding the workspace activity of design teams. In *Proceedings of the Conference on CSCW*, pages 244–249, 1988.

[WBC91] Steve Whittaker, Susan Brennan, and Herbert H. Clark. Co-ordinating activity: An analysis of computer supported co-operative work. In *Proceedings of CHI91*, 1991.

[WF86] Terry Winograd and Fernando Flores. *Understanding computers and cognition.* Ablex Press, 1986.

[WS88] Steve Whittaker and Phil Stenton. Cues and control in expert client dialogues. In *Proc. 26th Annual Meeting of the ACL, Association of Computational Linguistics*, pages 123–130, 1988.

[WW90] Marilyn A. Walker and Steve Whittaker. Mixed initiative in dialogue: An investigation into discourse segmentation. In *Proc. 28th Annual Meeting of the ACL*, 1990.