

# What is Connected by Mutual Gaze? — User's Behavior in Video-mediated Communication —

Naoki Mukawa, Tsugumi Oka, Kumiko Arai and Masahide Yuasa

School of Information Environment, Tokyo Denki University

2-1200 Muzai-Gakuendai, Inzai, Chiba, 270-1382 Japan

{mukawa@,01ie020@ed.,01id002@ed.,yuasa@}sie.dendai.ac.jp

## ABSTRACT

Video-mediated communication systems such as teleconferencing and videophone have become popular. As with face-to-face communication, non-verbal cues such as gaze, facial expression, head orientation and gestures in visual systems play an important role. Existing systems, however, do not support mutual gaze because the lay-out of the camera and monitor is restricted. Thus, conversations using visual systems differ from those in face-to-face communication. This paper clarifies the problems of the video-mediated system, specifically for comparing the system with communication using eye-contact and with communication using no-eye-contact. This study focuses on the protocol of opening communication, e.g. establishment of a visual-audio link, person identification and confirmation of the acceptance of conversation. We conducted experiments using the two systems. Analysis of recorded video sequences revealed that the system using communication with eye-contact induced behavior similar to the system using face-to-face communication.

## Author Keywords

Video-mediated communication, gaze, eye contact, ethnographical approach

## ACM Classification Keywords

H5.3. Information interfaces and presentation (e.g., HCI): Group and Organization Interface.

## INTRODUCTION

Face-to-face communication is the most important type of human communication style. Human beings exchange words, facial expressions, gestures, postures and gazes that enrich mutual understanding. To expand human communication to bridge the spatio-temporal gap, electronic communication media have been developed, and are gaining widespread use in the community. Video-mediated communication systems such as videophone (wired or mobile) and video conferencing have become popular as well as existing telephone and e-mail[1].

Researchers have pointed out that video-mediated communication systems need to facilitate mutual gaze between users [9] as in face-to-face communication. Some

laboratory systems have been designed to establish mutual gaze [4,7,8,10]. However, few systems in the marketplace support eye contact.

Psychologists have revealed the importance of gaze in face-to-face communication, where the human gaze indicates turn-taking cues in conversation, projects the sense of intimacy between persons and also represents the focus of attention [2]. As for existing video-mediated communication systems that do not support eye contact, Short et al. claim that the system often misinterprets conversation situations. They point out that these systems may convey distorted gazes that differ from the case of face-to-face and of eye-contact systems. Sending a distorted gaze is sometimes much worse than sending no visual information at all (like a telephone conversation). Unfortunately, there are no precise evaluation reports that compare the two systems, with/without eye contact from both a psychological and an engineering point of view.

Three approaches can be applied to an evaluation of the systems. One is the cognitive psychological approach. Chen investigated the sensitivity of the discrepancy of gaze when the user's gaze direction shifts from ideal eye contact, and showed that spatial sensitivity is not symmetric [5]. His results present a design basis for a practical system of establishing eye contact, but he does not compare the two systems. The second approach is based on questionnaires about usability [9]. Many results show that a system that provides eye contact is preferable to a non-eye-contact system. However, Grayson et al. point out that even non-eye-contact systems convey some eye contact information if the lay-out of camera and display is carefully chosen. Their experiments show that, after some experience of the system, users can interpret their partner's gaze as looking at them even though the gaze is distorted [3]. Their results suggest that a questionnaire-based approach alone by asking users' preferences does not provide a solid basis for evaluating the systems.

Still another approach is the ethnographic methodology that analyzes users' behavior throughout their conversing interactions. By analyzing the users' conscious and unconscious behaviors that are considered to compensate the lack or the distortion of gaze, the usability of the system can be evaluated from a number of aspects. Even if users are accustomed to using the systems and can converse fluently

Copyright is held by the author/owner(s).

CHI 2005, April 2-7, 2005, Portland, Oregon, USA.

ACM 1-59593-002-7/05/0004.

after experience, this approach can clarify that the systems are affected by distorted gaze. However, there are few research works that compare the eye contact system with the non-eye-contact system based on the approach..

This paper clarifies the feasibility of video-mediated systems, and compares the eye-contact system with the non-eye-contact system by observing users' verbal and non-verbal behaviors.

Whatever communication style is chosen, the communication session can be divided into three stages: opening, maintaining and closing. Almost all human-to-human communication requires a specific protocol for starting a conversation such as the initial finding-out about each other, and salutation. So far, there are few research results on the opening stage of the visual communication link. By comparing existing face-to-face and telephone communication, our study will clarify the characteristics of eye-contact and non-eye-contact video-mediated systems.

### PROTOCOL OF ESTABLISHING VIDEO-MEDIATED COMMUNICATION

Schegloff found that the following confirmation protocol is needed in telephone conversation [6].

(1) Confirmation of physical connection of telephone channel  
 "Hello!" "Yes." With these words, the telephone channel connection is established.

(2) Person identification

"This is Taro speaking. Is that Hanako?" "Speaking." With these words, the caller and receiver are identified. Note that this step can be skipped in the case of cellular phones that display the phone number or name of the caller.

(3) Confirmation of receiver acceptance

"Can I talk to you now?" "Is this a good time to talk?" This confirmation is needed because telephone is a medium where the caller interrupts the receiver.

This kind of protocol observed in telephone conversation must also exist in other electronic communication media such as videophone and teleconferencing systems. Examples of protocol patterns are summarized in Table 1. By verifying the three-step protocol, we investigate users' cognitive burden of gaze understanding and recognitive burden of distorted gaze compensation. For this purpose, observations of users' verbal and non-verbal behavior should be carried out.

### HYPOTHESES

This section formulates hypotheses on human behavior at the opening of communication based on findings in social psychology and findings from the evaluation results of visual communication systems. Short et al. [9] compared two media: telephone and face-to-face. They point out that telephone users mainly use verbal and paralinguistic information, while face-to-face communicators use a variety of rich non-verbal information. However, there has been little

Step	Communication media		
	Face-to-face	Video-mediated	Telephone
Audio-visual link	Approach, gaze, raise hand, "Hi!" smile	Gaze, raise hand, "Hi!" smile, "Can you see me?"	"Hello!"
	(Same as above)	(Same as above)	"Yes."
Person identification	Shake hands, bow, "My name's Taro." (do not know each other)	Bow, "My name's Taro." (do not know each other)	"This is Taro speaking. Is that Hanako?"
	(Same as above)	(Same as above)	"Speaking ."
Conversation acceptance	"Can I ask you something?" "Are you free now?" "Well."	"Can I ask you something?" "Are you free now?" "Well."	"Is this a good time to talk?"

Table 1. Examples of conversation opening protocol

research on how participants use non-verbal information when they use visual communication media e.g. video conferencing and videophone systems. The role of non-verbal information in the opening of a communication channel based on observation of users' behavior is unknown. Referring to Schegloff's work in telephone case and our own empirical findings, the following hypotheses have been formulated: "Those who use a video-mediated, non-eye-contact system shift their gaze less to confirm the connection than those who use an eye-contact system," and "To compensate distorted gaze, such communicators use more utterances, gestures and facial expressions than those using an eye-contact system." Considering the above three-step protocol, the hypothesis is reorganized as follows:

Hypothesis 1: Confirmation of audio-visual connection

Users who utilize an eye-contact system repeatedly gaze and avert their gaze. This is because the audio-video link connection is confirmed by mutual gaze. Immediate awareness of connection follows social greetings such as smiles and raises of the eyebrows.

On the other hand, users who utilize a non-eye-contact system gaze more than users of an eye-contact system.

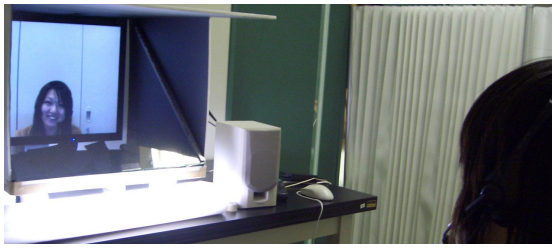


Figure 1. Experiment with eye-contact system.



Figure 2. Experiment with non-eye-contact system.

This is because a non-eye-contact system gives erroneous gaze direction and causes the compensated behavior of the user such as, “Can you see me?” and waving his/her hand at the monitor to confirm the connection. As a result, confirmation behavior requires a continuous gaze.

#### Hypothesis 2: Personal identification

If the caller and receiver are acquaintances, the confirmation behavior is not observed. This is because confirmation is performed by the displayed face only. If they are not acquaintances, salutations and self-introductions are observed.

#### Hypothesis 3: Confirmation of the receiver’s acceptance

An utterance corresponding to “Can I talk now?” is observed. A speaker who utilizes the eye contact system glances briefly. A user who utilizes the without-eye-contact system continuously gazes at his/her partner.

### EXPERIMENTS

Experiments were performed to investigate the above hypotheses through the conversations of the participants.

#### Experimental System

An experimental video-mediated communication system with eye contact was produced using a half-silvered mirror with 30-percent transmittance. The system was arranged so that the camera center axis meets the conversational partner’s gaze. Figure 1 shows the experimental system with eye contact. “A non-eye-contact condition” was realized by placing the camera on the right-hand side of the monitor as shown in Figure 2. A two-way audio-video link was established by the user pressing the button. The participants wore headsets with a microphone. Video and audio were recorded for analysis.

#### Procedure

Seven pair of students between the ages of 20 and 23 participated in the experiments. The students were

remunerated after the experiments. Five pair of participants were intimate friends, and the others were not acquainted with each other. They debated the “Pros and cons of capital punishment,” “Which do you prefer, cats or dogs?” and other topics.

They were requested to insist on their own opinion as far as possible. In addition, they were informed that the purpose of the experiment was to evaluate the video-mediated systems. No information was given that the actual purpose of the experiments was the establishment of communication channels. Before the debate, the participants were instructed to answer a questionnaire which did not concern the discussion topic. One of the participants was requested to push a switch to connect the audio-video channel after finishing the work. After the debate, the participants were instructed to fill in a questionnaire sheet about the usability of both systems.

### Experimental Results

The recorded video sequences were analyzed based on the ethnographical approach. An example of an analyzed sequence is shown in Figure 3. From observations of the all video scenes with eye contact system, the subjects seem to consider that visual-audio link was immediately connected only by repetition of mutual gaze and averted gaze accompanied by smiling. This resulted in some participants forgetting to confirm the audio connection at first and having to confirm the audio link connection later again. On the other hand, the non-eye-contact system needed quite a long time for confirmation of the connection by attempting behavior such as saying “Hi!” gestures, and smiling. This (probably unconscious) behavior can be interpreted as confirmation that the partner is looking at the subject, which is one kind of compensation behavior that recognizes distorted gaze. As a result, in the non-eye-contact system, the audio-link turned out to be confirmed by unintended behavior and there was no needed to confirm it.

Hypothesis 1 agrees with the experimental results. As for the use of the non-eye-contact system, participants tried to draw the partner’s attention by greeting and waving their hands, which was not observed in the eye-contact system. The users’ behavior in the eye-contact system, that repeatedly gazed and averted their gaze, meets the intimate equilibrium theory. The behavior can be interpreted as prolonged gaze in close proximity sometimes being considered rude.

Hypothesis 2 agrees with the experiments. When the participants are known to each other, the partner’s face on a monitor did not draw the confirmation behavior. On the other hand when the participants were not known each other, they gave salutations and self-introductions.

Hypothesis 3 partially agrees with observations where the participants uttered “Well...,” “Let’s begin,” for both systems. However, no apparent gaze differences were found.

**CONCLUSION AND FUTURE STUDY**

The experiments indicate that eye-contact systems provide immediate awareness of visual connection through users' mutual gaze. These results correspond to the behavior observed in face-to-face communication. On the other hand, participants using non-eye-contact systems are likely to need confirmation of opening the conversation by waving their hands and uttering a greeting. This can be regarded as compensating for distorted gaze direction.

Interestingly, it was observed that a pair of participants using the non-eye-contact system confirmed the three-dimensional consistency of the space between the two, such as, "Am I in the middle (of the monitor)?" This suggests that "confirmation of spatial consistency" can be included in audio-visual connections. As future work, we will continue our observation of users' behavior and clarify the relationship between the behavior and the usability of video-mediated communication systems.

**ACKNOWLEDGMENTS**

This research was partially supported by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Scientific Research (B)(2)16300032, 2004 and by the Research Center for Advanced Technologies, Tokyo Denki University.

**REFERENCES**

1. Finn, K.E., Sellen, A.J., and Wilbur, S.B. (eds.): *Video-mediated Communication*, Lawrence Erlbaum Associates, Inc., 1997.
2. Argyle, M. and Cook, M. *Gaze and Mutual Gaze*. London: Cambridge University Press. 1976.
3. Grayson, D.M. and Monk, A.F., Are You Looking at Me? Eye Contact and Desktop Video Conferencing, *Trans. on CHI*, SVol.10, No.3, 2003, pp.221-243
4. Ishii, I. and Kobayashi, M., ClearBoard: a Seamless Medium for Shared Drawing and Conversation with Eye Contact. In *Proc. CHI92*, ACM Press, 1992,pp. 525-532
5. Chen, M. Leveraging the Asymmetric Sensitivity of Eye Contact for Videoconferencing In *Proc. CHI2002*, ACM Press, pp.49-56, 2002
6. Schegloff, E. A., Beginning in the Telephone, in J.E Katz and M.A. Aakhus eds., *Perpetual Contact: Mobile Communication, Private Talk, Public Performance*, Cambridge Univ. Press, 2002
7. Okada, K., Maeda, F., Ichikawa, Y. and Matsushita, Y. Multiparty Videoconferencing at Virtual Social Distance: MAJIC Design. In *Proc. CSCW*, pages 385-393, 1994.
8. Rosenthal, A.H. Two-way Television Communication Unit. US Pat. 2,420,198, 1947.
9. John Short, Ederyn Williams and Bruce Christie: *The Social Psychology of Telecommunications*, John Wiley & Sons, 1976
10. Vertegaahl, T., Van der Veer, G. C., and Vons, H. Effects of Gaze on Multiparty Mediated Communication. In *Proc. Graphics Interface 2000* pp.95-102, 2000

Item	Subject	Time line									
		0 sec	2 sec		4 sec				10sec		
Hand gesture	A			Raise hand							mic
	B			Raise hand		Raise hand					
Gaze	A	Front	Left		Front			Below left	Front	Below	Front
	B		Front	Below left	Front	Below left	Front	Below left		Below left	Front
Facial expression	A	Smiling									
	B										
utterance	A					Hi,		Hi			well
	B		Oh!		Hi,		Hi				

(a) System with eye contact

Item	Subject	Time line										
		0 sec	2 sec		10 sec		18 sec		24sec			
Hand gesture	A			touch mic								
	B			touch microphone					mic		mic	
Gaze	A			Front				Below		Front	Left	
	B	Front	CAM	Front	CAM	Front	Below left	Front	CAM	Front	B-Left	CAM
Facial expression	A		Surprised									
	B											
utterance	A		What's wrong?	Aside?			On my left.		You push me away			
	B		Hi!	You look aside.		Both of us. Here's a camera.		Well....	No.			

(b) System with non-eye-contact

Figure 3 Behaviors of Participants just after the system on.