How to Use Gaze Information in Video Communication

Osamu MORIKAWA
Psycho-informatic laboratory, National Institute of Bioscience and Human-Technology
1-1 Higashi, Tsukuba, Ibaragi 305,Japan
+81-298-54-6731
morikawa@nibh.go.jp

Abstract

This report has examined the way that gaze is transmitted in videophons and video teleconferencing systems. Our results showed that face-to-face communication resulted in an accuracy rate of 90%, while the rate for positioning directly in front of a camera was 80%. However, as the visual angle increased from the camera, the accuracy rate fell dramatically to about 5%.

As some subjects gained experience of speaker, they could understand the characteristics of the video image, which affected their gaze recognition. This result may have occurred increasing the level of information processing by making an image of the alter ego.

KEYWORDS: gaze, video communication, human factor, cognitive feature

1. Introduction

Electronic technology has been changing the ways in which people communicate with one another. Until recently, face-to-face and telephone conversations were the only ways that people could communicate in real time; now, however, video phones and video teleconferencing systems which enable the use of image information on an individual basis have become possible. This paper examines how this added visual information can be used with the most effectiveness.

Virtual reality technology is one of the most important elements of media-based communication. In order for people to effectively utilize virtual reality technology, the environment supplying the system must be suitable for their characteristics. Therefore, people's characteristics must be understood. Suitability does not just refer to elementary sensory level such as sight, hearing, touch and etc.; it must also include the cognitive processing level of people. This paper examines gaze information in person-to-person communication, taking special note of people's cognitive processing level.

Experiments were used to measure how accurately people understood gaze information during communication. Then, a human information processing model was used to examine the results.

2. Non-monotonicity of subjective evaluation

Usually, the addition of a new service expands the options available for communication, which is increasingly being regarded in a favorable light. However, people's subjective evaluation does not necessarily improve with the addition of services. In other words, subjective evaluation is characterized by the existence of non-monotonicity. Let us consider the following studies.

According to a survey conducted on 549 college students in Kansai suburbs by Yoshida et al[1], the advantages of telephones are that they are 1) fast, 2) can be used at any time, 3) do not require a location to be chosen, 4) do not require the communication partner to be seen, and 4) enable things to be said which otherwise could not be said. Disadvantages are that 1) the communication partner cannot be seen, and 2) subtle feelings cannot be conveyed. Especially interesting is the fact that "not being able to see the communication partner" was seen as both an advantage and a disadvantage.

Therefore, we can see that different people feel differently in the same media environment. Even the same person can feel either comfort or discomfort in the same medium, depending on the circumstances. Furthermore, even if the media environment, users, and purposes are all equal, we can see that often people who originally had misgivings become comfortable with the new environment, and vice versa.

Although such evaluation characteristics of a person are seemingly contradictory, many of these cases can be explained rationally by taking the following into account. In other words, people's evaluation of their media environment is comprehensive, including not only the physical attributes of this environment, but also their purpose for and specific means of using it. Expressed as a formula, we can think of the evaluation function f as being expressed not as f(environment), but as f(environment, purpose, user method). Although, technically speaking "mapping" is a more appropriate term than "function," since subjective evaluation also contains quantitative aspects, this paper will proceed with the discussion of evaluation content as scalar values.

User methods change depending on the user's ability; for instance, if one is not an expert, he or she might not be able to use certain methods. Therefore, perhaps the evaluations of such users should be expressed as f(environment, purpose, user ability); in other words, this function would show evaluations based on the most appropriate user method considering the ability of the user.

Therefore, understanding evaluations of the same environment can help us find the causes of differences in purposes and user methods. Furthermore, case studies of the same individual's learning function can help us to find the causes of changes in user abilities. This concept is being considered in ISO 9241, Part 11: Guidance on usability [2].

3. Human Information Processing Model

One of the features of the human information processing model is the non-monotonicity of the subjective evaluation function. Human information processing is characterized by "hardware" and operational "software" that interact with one another in the head. Features of human "hardware" include the limitations of problem-solving ability due to the capacity of short-term memory (STM) and working memory (WM), and the search tendencies of long-term memory (LTM) and access speed. In addition, features of the information input-output section include the sensory/perceptive and motor systems, which have been researched in experimental psychology, cognitive psychology, and so on. Taking the results of these studies into consideration, this report will propose a human information processing model [3]. The basic concepts of the proposed human information processing model are parallel processing and the limitation of mental resources, the fact that most of the information which is merely processed is not used, and the existence of template strategy as knowledge. People gather as much information as they can from the outside, and their mental resources are distinguished by function and classified into regions. Mental resources demanded of the processing strategy are sequentially defined according to "region, volume, time."

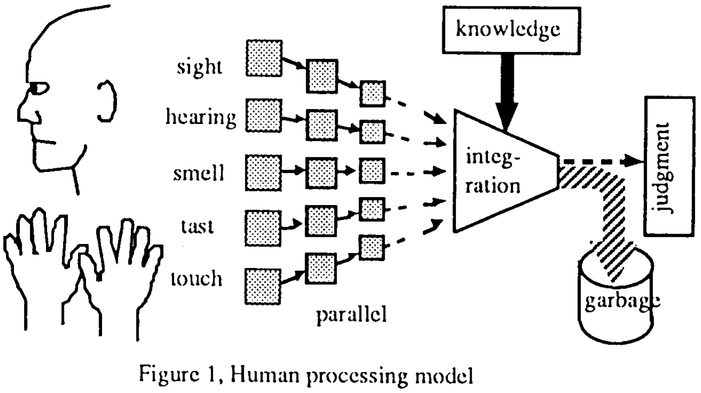

Figure 1, Human processing model

For example, in this model information that is ultimately observed to be useless can be hidden from observation by using a different processing method; thus, processing results which contradict other types of received information and existing knowledge are discarded. Or, the information can be discarded when it strays too far from the subject at hand.

Human information processing, not just in face-to-face communication but in people's everyday lives, involves a multiple paths information which leads to one conclusion. The greater the number of flows, the higher will be the evaluation of the reliability and safety of the conclusion. There is preliminary information for the conclusion which leads to bias in the overall judgment. Recognition processing of the communication partner's gaze is no exception.

4. Gaze information

As was mentioned in Section 2, the recognition characteristics of humans varies depending on the type of information that one is trying to obtain. Therefore, before discussing characteristics of recognizing gaze direction during communication, we must consider the role of gaze information.

Gaze activity acts as the starting point in interpersonal relationships. We judge the gaze direction of our partner and determine the focus of this gaze. This provides us with clues in knowing where our partners are directing their attention and the object(s) in which they are interested.

Gaze not only to gives clues about the start of communication, it also, as Kendon [4] has pointed out, has the following three functions during the course of the communication:

(1) Monitoring-- determining whether the conversation should continue or end based on whether or not the listener is gazing intently at the speaker.

(2) Regulation-- inferring from the movement of the partner's eyes where he or she is interested in the conversation, then adjusting the topic so the partner gazes at the speaker with interest

(3) Expressiveness-- non-verbal communication to the speaker to indicate whether the conversation is effective or is being rejected

Line of vision in such types of gaze information is important whether or not the speaker is being seen, and there is probably not much meaning in information about where the listener is looking.

Gibson et al.[5] and Anstis et al. [6] have measured recognition characteristics of the other person's gaze. According to them, if the speaker is not looking directly at the listener, there will exist a constancy error of judgment between the actual gaze direction and the gaze direction at which one feels the listener can be seen. In other words, when the face is turned toward the listener's right, the speaker, when looking at the listener, will tend to feel that he is looking to far to the listener's left. On the other hand, the speaker will actually be looking a little farther toward the listener's right than he thinks he is. Backward calculations made from Gibson et alís data show that when the face is turned away 30 degrees, recognition is off by about 3 degrees. This represents an error of about 10 cm for a distance of 2 meters. At such a small degree of error, the gaze direction toward oneself cannot be mistaken for the gaze direction toward a neighbor.

Furthermore, Anstis et al's research concluded that tilting the video monitor (convex surface of 60cm in diameter) led to the occurrence of constancy error in the direction of the tilting (about half the size for face situation).

All of these studies involved not conversational conditions, but experimental conditions in which all of one's effort is put into judging gaze. Furthermore, the images used for stimulation were either static of barely moving. However, the main purposes of the act of conversation are to understand the content of what is being said and to transmit this accurately, not to be aware of gaze. Therefore, it is necessary to measure the characteristics of gaze recognition at dividing of one's attention (cognitive resources) while at the same not impeding the progress of the main purpose which is communication. In addition, its is possible that gaze direction can be determined when the face and eyes are moving, especially through the integration of various types of non-image information obtained from facial images. Therefore, it is necessary to conduct quantitative experiments on gaze recognition characteristics of three types of sight information: that which mainly involves static images, that which consists of moving images, and sight information as one element in communication.

5. Experiments

To investigate how gaze information is altered by video images, two types of experiments were conducted: one on video conditions, the other on face-to-face conditions. There were 68 subjects of the experiments (51 of whom were female), ranging in age from teens to fifties. Experiments were conducted in groups of 4, with 6 groups having no previous face-to-face communications with one another, and 11 groups comprised of close friends.

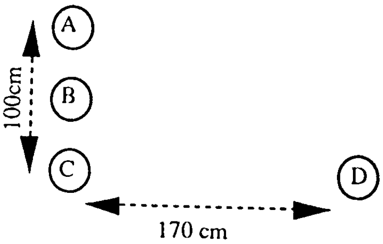

Figure 2 Layout of face to face communication

5.1 Face-to-face experiments

These experiments consisted of one experiments in which subjects stared at one of them without speaking (static images) and 2 experiments in which a designated person would talk to one of them (moving images). The positional relationship of the subjects is shown in Figure 2. Three listeners sat in a lengthwise row 50cm apart from one another. The speaker sat 170cm directly in front at the far right of the listeners (C-seat) . Psychologically speaking, this is the near phase of social distance [7]. It should be noted that the subject on the far left of the listeners (A-seat) was 200cm away from the speaker and the visual angle of listeners (from seat-A to seat-C) was 30 degrees from the speaker.

Communication conditions were arranged such that subjects were spaced together as in a group picture, that is, they were touching the shoulders of the people around them, and vertical spacing would ensure that no one's face would be hidden. The degree to which listeners felt they could be seen just from visual information was evaluated at five levels: "speaker looks at me", "probably looks at me", "can't say whether the speaker looks at me or not", "probably does not look at me', and "the speaker does not look at me". The gaze direction was recorded by listeners as the relative position, centered on themselves, which they judged to speaker to be looking at. In addition, in cases of general judgments which included such conditions as the content of the talk, subjects used the same five levels to evaluate whether or not they were seen.

The **only gaze condition** (O-condition) assumed that talking might begin. The speaker would look at target points among the listeners without uttering a word. Target points consisted of a total of 11 points in the vicinity around three listeners: one point each to the left and right of the listeners, and three vertical positions for each of the listeners. There were no special limitations on the direction in which people's bodies and faces faced; rather, instructions were given to look at target points in as natural fashion as possible.

On the other hand, the **interactive condition** (I-condition) assumed normal conversation conditions. The speaker would proceed talking to his or her partner at the target point while checking to see whether or not the partner was listening. At such time, since there were several subjects, vocal agreeable responses were prohibited, but listeners were instructed to nod at and make as much gaze as possible with the speaker. The conversation consisted of using maps to give road information. Speakers were instructed to interact at least three times during every talk.

In order to investigate whether information on gaze recognition during the I-condition came from the movements of the speaker or from the interaction of listeners and speakers, the **speak to condition** (S-condition) was added to the experiment. Here, the speaker was instructed to start talking to a listener as the former was moving, and while moving, keep looking at the listener without answering. Specifically, this consisted of introducing oneself and bowing while continuously looking at the listener

5.2 Video experiment

In this experiment, two rooms were prepared. In the listeners' room, a 90cm high X 120cm wide screen was set perpendicular to and 70cm above the floor. The projector was 170cm in front of the screen and 90cm above the floor, and the camera was placed on the right side of the screen (Fig. 3). The 3 listeners adjusted their chairs so that their line of vision was at exactly the same height as the camera. Also, the size of the display was adjusted so that the listeners appeared life size [8].

In the speaker's room, a screen was placed 180cm away from the speaker, and cameras were set up so that the gaze direction was in alignment with the listener on the far right. In more precise terms, two cameras were set up, one each in the following positions: 170cm and 80cm from the speaker, in direct line with the listener at the far right of the speaker and screen (seat C). This meant that, depending on the person and camera, the dimensions could differ from those of face-to-face communication due to the forward and backward movement of the speaker.

The topic in the video experiment was the same as in the face-to-face experiment. It should be noted that locational differences of the cameras were assumed to be negligible under conditions

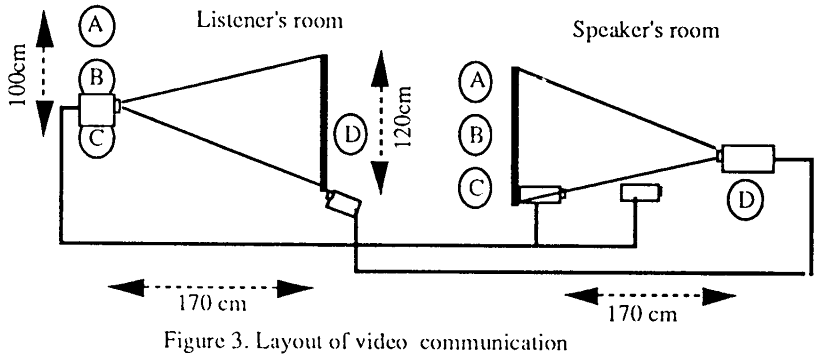

Figure 3. Layout of video communication

of little movement and fixed gaze, so there was only a standard distance.

6. Assumptions

The fixed gaze in the face-to-face experiment is actually used in people's daily lives, so good results were expected. Errors which could be expected were careless ones, such as subjects feeling they could be seen but were not being seen, and vice versa. In the latter type of error, subjects mistakenly believed that they were being seen, and they directed their attention toward the speaker. After they paid their attention toward the speaker, they realized they were not being seen. Then, this type of error caused subjects to waste time with needless processing. On the other hand, in the former type of error there was a possibility that in some cases, personal relationship played a major role. Therefore, under conditions in which it was difficult to distinguish gaze, the former type of error probably occurred more often than the latter type.

From the results of Anstis et al, during the fixed-gaze topic in the video experiment , the subjects on the far right(seat C), who were directly in front of the camera, tended to show the same degree of response accuracy as in the face-to-face experiment, while the others(seat A,B) felt that the people to their right were farther, albeit slightly, to the right. However, unlike Anstis' experiment in which there was a curved monitor screen, the screen here was flat with little image distortion, meaning that there was a possibility that constancy error was not observed. In such a case, uniform responses could be expected of the listeners regardless of their location.

In the interactive conversation experiment, the main purpose was not identification of the gaze direction but rather understanding the content of what was being communicated. Therefore, a decline in attentiveness caused by expending energy to distinguish gaze information may have led to an increase in error. On the other hand, the results might have been enhanced by the speaker's reaction to information such as agreeable responses ("uh-huh," "yes, you're right," etc.) coming from listeners which caused the results to improve. In other words, differences between interactive conversation and fixed gazes under face-to-face conditions could not be estimated, but under video conditions, as the quality of the image worsened, face-to-face became more effective than fixed gaze, which may have caused discrepancies in the results.

When the speaker spoke to a listener, the direction of the speaker became clear due to body movement, so during face-to-face communication, the results were expected better than those of fixed gaze; the accuracy of responses during the video communication was expected to be the highest for the person directly in front of the speaker. Furthermore, when the camera was at a close distance, forward and backward movement was exaggerated, so this trend was very evident, but the unnaturally of the image information increased. Therefore, even though the results improved, we should not necessarily assume that we have found a method to increase realism.

7. Results

In the face-to-face communication experiments, about 90% of the responses were correct, as we had expected (Table 1). Although the results of interactive conversation were usually better than those of the fixed gaze, there was no significant difference in either. The results of errors showed that there were twice as many careless errors (thinking that one was not being seen when the opposite was true) as the opposite kind of error. In addition, neither total reached 100% because of the evaluations from three levels of responses ("speaker probably looks at me", "can't say whether speaker looks at me or not", and "speaker probably does not look at me") which were not included in the calculations. According to a questionnaire distributed after the experiments,

Table 1 Resuls of face-to-face communicationexperiments

condition		seen	not seen
O	correct	88.10%	87.80%
	error	7.30%	2.50%
S	correct	81.60%	84.50%
	error	16.60%	3.10%
I	correct	93.00%	87.70%
	error	2.50%	1.20%

Table 2 Resulus of Video communicationexperiments

condition		seen	not seen
O	correct	28.90%	71.20%
	error	57.80%	14.90%
S	correct	30.30%	72.50%
	error	53.90%	15.80%
I	correct	28.10%	68.80%
	error	55.90%	16.70%
S (near)	correct	31.50%	73.40%
	error	54.10%	13.60%
I (near)	correct	31.30%	70.00%
	error	48.60%	14.40%

Table 3 Resuls of Video communication experiments
(position of each seat)

		L (30°)		L (15°)		center	
condition		seen	not seen	seen	not seen	seen	not seen
O	correct	2.90%	64.70%	1.40%	67.60%	82.30%	81.20%
	error	89.70%	18.30%	77.90%	22.70%	5.80%	3.60%
S	correct	1.40%	63.60%	2.90%	69.10%	86.70%	84.80%
	error	79.40%	22.70%	79.40%	21.30%	2.90%	3.30%
I	correct	0.00%	59.30%	0.00%	63.40%	81.50%	84.00%
	error	83.30%	22.30%	79.10%	25.00%	7.80%	2.60%
S	correct	4.40%	66.50%	5.80%	69.70%	83.80%	84.10%
(near)	error	83.50%	18.70%	70.50%	20.20%	8.80%	1.80%
I	correct	4.10%	59.70%	9.20%	66.60%	81.90%	83.50%
(near)	error	75.00%	20.60%	64.40%	20.40%	5.50%	2.20%

nearly all of the subjects paid more attention to gaze during the experiment than they usually do in their normal lives. Given this situation, the response accuracy of actual face-to-face communication would probably be lower. In other words, even if gaze recognition error occurs at a 10% frequency, it does not pose a problem in face-to-face communications in people's daily lives.

In the video communication experiment, the response accuracy when not being seen was about 30%, and about 70% when the subject was being seen (Table 2). Looking at this in terms of the position of each seat (Table 3), correct responses, whether the subject was seen or not, were at the 80% level when subjects were directly in front of the camera. In all other cases, the response accuracy was 60% when not being seen, but was almost non-existent when being seen, falling to around 2%. When the camera was close, there was no change when the camera was directly in front, but at a 15 degree position there was a 7% accuracy rate, which fell to 5% when the angle to the camera increased to 30 degrees.

Since the video images were not of a high quality, it had been expected that the results would have been (from worst to best), fixed gaze, starting a conversation, and interactive conversation. In actuality, however, no significant difference was found.

Table 4 shows an arrangement based upon Anstis et al's concepts of video image recognition. According to the results of this arrangement, recognition accuracy rates were high even at oblique angles from the screen: 70% for a 15 degree angle, and 60% for a 30 degree angle. There is some question as to whether these results can be interpreted constancy error caused by the visual angle from the screen, but it is certain that the accuracy rate declines as the angle from the screen increases.

8. Feeling that there is an alter ego

Among the post-experiment impressions of some of the subjects was the following: "I, who was having a sideways conversation, felt that there was another part of me which the speaker was speaking to. Therefore, when you asked me whether or not I was seen, I had trouble deciding which of my lines of sight [of the ego or alter ego] to use in the answer" . However, upon examining the average values of all subjects' data, we could not find any evidence to support such a feeling. Therefore, we calculated the scores into the 5-level evaluations to derive a total score for each subject (Scoring: correct, 2 pts.; mostly correct, 1 pt.; could not determine, 0 pts.; mostly incorrect, -1 pts.; and incorrect, -2 pts.) . Analysis of the responses of the top 9 scores showed a significant difference among recognition tendencies, depending on the experience the speaker had with video communication. There were also significant differences between the overall the judgment and judgment based solely on video information. Therefore, when response trends were analyzed in terms of the experience of the speaker, the general trend regarding the distribution of listeners' impressions could be read from the results (Fig.4). We can take it that there were two peaks-- immediately in front of the screen, and at a 30 degree angle. This phenomenon can be interpreted as follows:

Gaze direction is determined from video images. This is similar recognition to the C-seat position. At the same time, they can use the experience of the speaker to understand the relationship between the speaker and the camera. And, through I-condition conversation, they can understand eye contact when they are being seen. Our reasoning tells us that they are being seen,

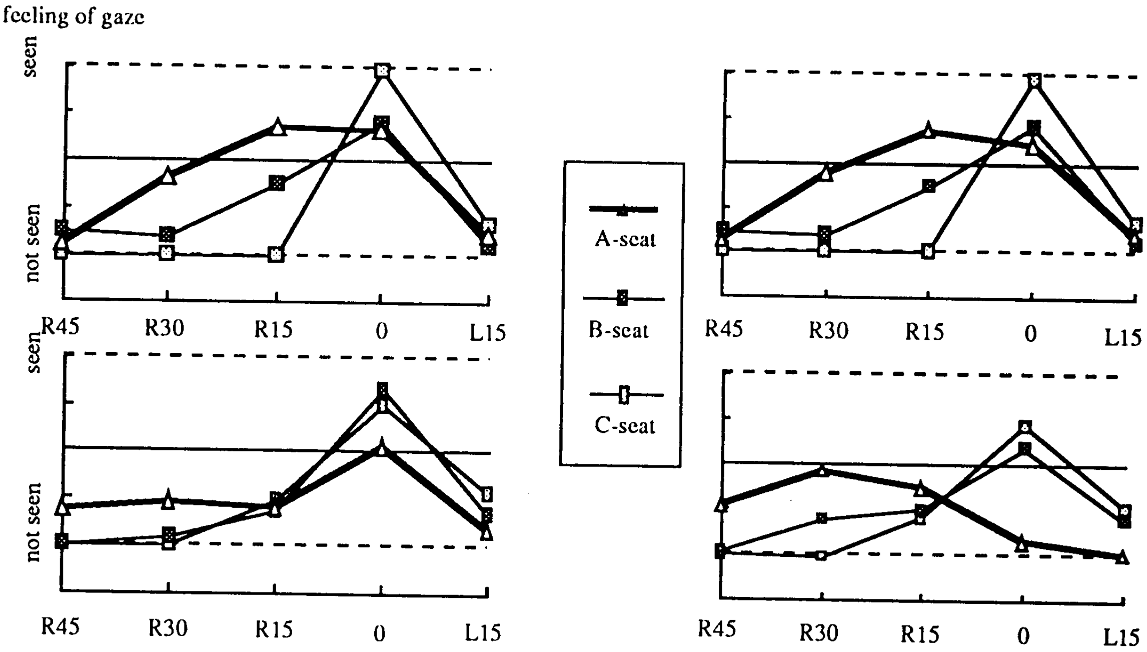

Figure 4. Mean of best 9 subjects' rating.
Intuotial judgement (left)/synthetic judgement(right)
before the experience a speaker(upper)/after the experience a speaker(lower)

but contrarily they feel as if someone else is the object of the video image's gaze. Therefore, the subjects of this experiment develop a new information processing strategy to eliminate this contradictory condition. Methods for this elimination may have included a method to correct gaze direction calculated from image information(Fig. 5 left), and a method which changed the overall judgment while keeping the values for gaze direction which were calculated from image information(Fig. 5 right). Subjects who reported feeling an alter ego began to develop the latter processing strategy. In other words, to eliminate these contradictions, these subjects put themselves in the gaze direction determined from video information, interpreting this to mean that they were being spoken to.

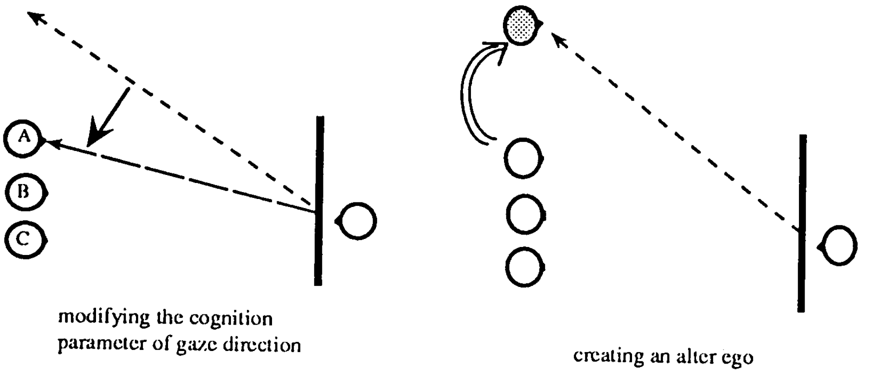

Figure 5. strategies for elimination of video image's gaze conflict.

9. Discussion

Evaluation of the information media from the information processing model provided the following properties:

A. In multipath-induced conclusions, if some of the information which leads to the conclusion is missing, only the reliability declines; the conclusion itself remains unchanged.
B. When multipath-induced conclusions contradict one another, there are many ways of dealing with the situation, such as ignoring either speaker or listener, putting oneself in the place of both speaker and listener, and devising new strategies for eliminating contradictions.
C. The effect on intermediate processing is greater on subjects who are closer to the point of input.
D. Information which causes misunderstanding is detrimental to the communication.
E. The exhibition of information not used in the communication is one factor reducing the

effectiveness and comfort of the communication.

F. Input information which is especially demanding on mental resources is another factor reducing the effectiveness and comfort of the communication.

One of the manifestations of property B is a method in which an alter ego is created to correctly recognize gaze direction in video communication. However, at the present time it is unclear whether such a method is actually suitable for humans to use as a strategy. If it is suitable as a strategy, the intensive use of this property may give rise to new video communication formats and/or methods. We would like to structure these formats so that the alter ego can easily be distinguished from the self and that no confusion occurs.

In fact, when "cyber-hero" of TV video games confront his "cyber-enemies", the player feels as if (s)he is actually facing these enemies. At such times, the movement of the enemy on the screen is toward the hero, yet there is no actual physical movement of the enemy toward the player. Nevertheless, the player feels as if the enemy on the screen is actually facing his/her self. In this respect, there was no disorientation among the subjects of the present experiment. We may be able to attribute this to the fact that there are no common elements between the make-believe world of the video game and the real world as seen on a video display.

On the other hand, the existence of the self and the alter ego in the same space in the experimental environment differs greatly from the case of video games. Therefore, at the time when we perceive the gaze direction to be directed at our own self, the conclusions may differ depending on which "self" is used. This causes confusion for us.

Therefore, when the results of gaze cognition contradict other types of information and when an environment is created in which these results cannot be interpreted as gaze toward one's self, we can expect the number of common elements between the alter ego and the self to decline and confusion to diminish.

Subjects who could not understand the nature of gaze information of video communication could not resolve the multipath contradictions, or else there were no contradictions but they made the wrong interpretations. This means that image information satisfies property D. From the results of the present experiments, we can conclude that for the majority of subjects, image information was detrimental to their cognition of gaze direction. It is necessary to instruct such people that their "cognition of the gaze direction obtained from image information is of no use."

In the case of the telephone, which has until recently used only verbal communication, it is clear to users that only voice can be transmitted through this medium. Therefore, during the communication, users automatically select a strategy which does not use perceived visual information.(Fig. 6). This allows them to use the telephone without becoming confused. However, even when contradictions appear at the stage of final judgment despite the effective use of the telephone ad subsequent information processing, processing results as they relate to visual information will be discarded.

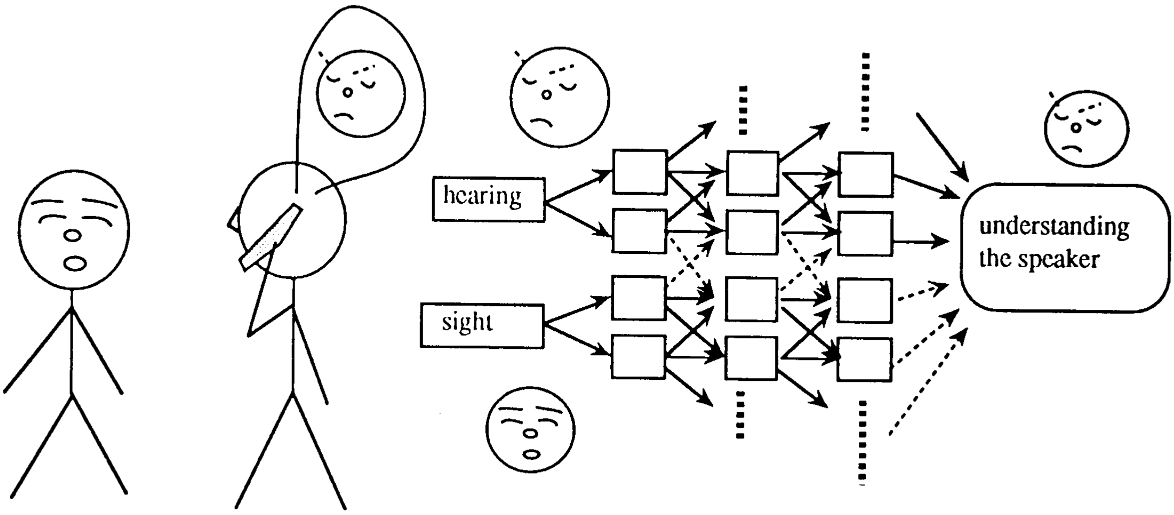

Figure 6. A strategy of telephone communication

On the other hand in video communication, effective visual information is received mixed with inaccurate information (forcing an interpretation which differs from that made during face-to-face communication). The pick-and-choose selection of only effective information is presently hard to do, even when one is conscious of the effort. Furthermore, even at the stage of final judgment, this can lead to a non-contradictory, interpretable solution. In other words, one of the problems with video communication occurs when people make the wrong interpretation and stick with it.

Therefore, the next topic for research should be how to avoid such problems. There are four possible methods. Although it is difficult to attain, conceptually the simplest solution would be to provide image information which is identical to face-to-face communication. The next simplest method would be not to transmit visual information. In this case, facial images would not be transmitted.

Depending on the situation, such clear-cut solutions would be the best. In fact, Gaver[9] has reported that facial images are used only about 10% of the time in face-to-face communication, and eye contact is made only 2% of the time when watching the screen.

Although visual information can be detrimental information, facial expressions, timing, the appearance of the partner's room, etc., are usable information which we want to use. A possible solution from this perspective would be to cause processing results from a different flow to be used during general judgment. For example, this might entail lowering the temporal, spatial, and color resolutions, which are properties of image information. By doing so, we would no longer attempt to read visual information. Or, even if we did read it, the reliability of the recognition results would decline. During general judgment, this would, in effect, not be used but instead discarded. Human information processing is one way of dealing with such a scenario. In fact, even in face-to-face communication, at distances greater than the psychological "social distance" of 4 meters, judgment of gaze direction from image information becomes uncertain. Moreover, humans naturally switch to a strategy which overlaps other types of information.

An example of another type of method we use is the incorporation of some kind of framework to indicate where the our communication partner is looking. By doing so, we reinforce the cognition path in which the other person's gaze direction is judged by backward calculating from the position of the object he is watching. This effectively discards the results of gaze cognition from unreliable facial images. Such a scenario is also being considered.

The final solution method involves the creation of cognition results which contradict other data due to environmental setting, along the lines of property B.

10. Future research

This report has examined the way that eye contact is transmitted in video phones and video teleconferencing systems. Our results showed that face-to-face communication resulted in an accuracy rate of 90%, while the rate for positioning directly in front of a camera was 80%. However, as the visual angle increased from the camera, the accuracy rate fell dramatically to about 5%.

As some speakers gained experience, they could understand the characteristics of the video image, which affected their gaze recognition. This result may have increased the level of information processing by making an image of the alter ego.

We have reported on four methods which provide for efficient use of video images through the human information processing model. Future research will focus on the third and fourth methods, that is, (3) using the results of processing other flows during the general judgment, and (4) creating recognition results which are contradictory to other types of information due to environmental setting. We would like to create and investigate scenarios for these two methods.

As an example, to know the object of the other person's interests, we would like to conduct an experiment which assumes that gaze information from the communication partner is being read. For this, we plan use a remote-controlled camera, and set up an environment such that every participant in the communication would know how every other participant is acquiring which kind of image information. We are planning research to measure the frequency of trying to read gaze information and the recognition accuracy of such information under such conditions.

reference

1. Yoshida, A., Kakuda, J.Research study on the communication by telephone in Japanese university students(in Japanese). 8th symposium on Human Interface, 1992,pp643-650,
2. "Ergonomic requirements for office work with visual display terminals (VDTs), Part 11: Guidance on usability " , ISO DIS9241-11, ISO (1995)
3. Morikawa. O.: A human processing model for explaining the effects of media in communication (In Japanese). Information Processing Society of Japan SIG Notes, 1995,HI63-6, pp.41-47
4. Kendon, A.Some functions of gaze direction in social interaction. Acta Psychologica,

1967,vol.26,pp.22-63

5. Gibson, J. J. and Pick, A. D. Perception of another person's looking behavior. American Journal of Psychology,1963, Vol.76, pp.386-394.
6. Anstis, S.M., Mayhew, J. W. and Morley,T. The perception of where a face or television 'portrait' is looking. American Journal of Psychology,1969,Vol. 82,pp.474-489.
7. Hall,E.T.: The Hidden Dimension,Doubleday & Company, Inc., New York,1966
8. Kurosu, H. Yamadera,H. Motomiya,Y. Mimura, I.The Optimal Size of the Human Body on the Screen in the Visual Communication (In Japanese), Information Processing Society of Japan SIG Notes, 1995,GW13-8, pp.43-48
9. Gaver,W., Sellen, A., Heath,C.and Luff P. One is not enough: Multiple views in a media space. Proc. INTERCHI' 93. ACM, pp.335-341.