INFORMATION TO USERS

EFFECTS OF GAZE ON SPEAKING BEHAVIOR

IN

TRIADIC CONVERSATIONS

by

YAPING DING

A thesis submitted to the

Department of Computing and Information Science

in conformity with the requirements for

the degree of Master of Science

Queen's University

Kingston, Ontario, Canada

March, 2002

0-612-69298-1

Abstract

We designed an experiment investigating effects of gaze on speaking behavior during triadic conversations. Understanding the effects of such non-verbal cue on speech and turn taking is crucial for the design of teleconferencing systems and Collaborative Virtual Environments (CVEs). Earlier studies found that subjects take more turns when they experience more gaze. We evaluated whether this is because more gaze allowed them to better understand if they were being addressed. We compared speech activity, conversation participation, and gaze behavior between two conditions: (1) in which subjects experienced gaze synchronized with conversational attention, and (2) in which subjects experienced random gaze. The level of gaze experienced by subjects was a covariate. Results showed that subjects significantly speak more and contribute more ratio of speech when gaze behavior was synchronized with conversational attention. However, covariance analysis showed results were likely due to differences in gaze level rather than gaze synchronization, with correlational coefficients of .62 between gaze level and amount of subject speech, and .56 between the level of gaze and the ratio of subject speech to group speech. With the removal of the effect of gaze level, gaze synchronization did not in fact have significant effect on speech. The result validated the emotional function of gaze as a favorable explanation of gaze effect on human interaction in multiparty conversation.

Results also demonstrated that subjects in the synchronized condition completed significantly more correct sentences than those in the random condition. However, we did not find significant linear relationship between the level of gaze and task performance. Hence, the possible causal

relation among gaze, speech and task performance would be that the level of gaze affected speech, and speech affected task performance.

With regard to gaze behavior of subjects, neither gaze condition nor the level of gaze has significant effect on this measure. In general, we also did not find sex difference in the level of gaze at their partners. However male subjects gazed more at their female partner than they gazed at male partner. We did not find a similar pattern in female group. People in both conditions were satisfied with conversational systems, partners, and task.

Based on the finding that the level of gaze has main effect on speech, we conclude that although it would seem commendable to use synchronized gaze models when designing CVEs (Collaborative Virtual Environment), we cannot rule out that, depending on task situation, random gaze models would suffice.

Acknowledgements

I would like to thank Dr. Roel Vertegaal for his help in supervising my thesis project. This work would not be possible without his valuable suggestions in every stage of the project.

I would also like to acknowledge the financial support provided by Queen's University and the Ontario Student Assistance Program.

I greatly appreciate every member in our CISC community who brings me fun during my study period, Alexander, Amber, Andrew, Chris, Debby, Freeman, Hogan, Jeff, Jeremy, Michelle, Richard, Ross, Sabine, Sam, Tiantong, Wenhu, Wenzhong, Xiao Du...... I will remember you forever.

Contents

Contents

List of Figures

List of Tables

List of Appendices

1 Introduction

1.1 Introduction

Eye gaze is defined as the act of looking at other humans (Argyle, 1967). Since the facial region is typically the focal point in this act, the term is virtually synonymous with gaze at the facial region. The function of looking behavior in human conversations has inspired many investigations. It is very important to recognize that vision is the major information source for human beings. About 70% of all the receptors in a human body are in the eyes (Campbell, 1993). Vision thus provides the most significant source of information to sighted human beings and plays a major role in information seeking tasks. On the other hand, eyes are also senders of information. Eyes can convey different meanings through gaze. A gaze may signal concern, worry or fear. People who have high status may show power by prolonged gazes or by denying others their attention (Kalma, 1992). As such, the eyes have long been recognized as mirrors of the soul. In human communication, eyes not only provide a way of seeking information from partners, but express emotions that provide additional cues to verbal communication. As such, gaze is one of the most powerful non-verbal expressions (Vertegaal, 1998).

Eye gaze should therefore also be considered as a critical element of computer-mediated communication. For example, in multi-agent, multi-user communication systems, both users and agents should know who is talking to whom, and when they are addressed and expected to speak (Vertegaal, 2001). In multi-user communication systems such as video-conferencing systems, in order to have effective and efficient communication and collaboration, users should be aware of the orientation of their partners and where or what they are focusing on. Non-verbal cues such as eye gaze may provide this kind of information. To provide non-verbal cues in multi-agent, multi-user communication systems, we need to know more about communicative functions of the eyes.

Mediated systems such as video-conferencing have long been introduced to the market. However, because of technical restrictions, current commercial systems do not support the communication of eye gaze. The eye gaze in video image is not appropriately conveyed through the system. During communication, users observe images with gaze lower when their co-workers actually gaze at them on the screen. The reason is that the video-conferencing cameras are usually mounted on top of the desktop, which makes it hard to convey images with right gaze direction. We consider a lack of support for proper gaze communication as one of the possible reasons that current commercial systems have not been successful.

It has been suggested that, in multi-user systems, the participants need to use gaze to determine who is talking to whom and who will take the next turn in order to obtain a smooth conversation. Vertegaal et al. (1999) suggested that eye gaze is an excellent predicator of conversational attention. When someone is listening to a person, there is 88% chance that the person gazed at is the person listened to. When some one is speaking to an individual person, there is a 77% chance that the person looked at is the person whom is being addressed. In human interaction, the duration and frequency of gaze of the first person are strongly associated with those of the second: the more frequently one person gazed, and the longer his gaze is, the longer and more frequent are the gazes of the second person (Kendon, 1967). With respect to turn taking, Kendon (1967) and Duncan (1972) revealed a general gaze pattern during opening and closing of an utterance, which signals attempts to obtain, hold or yield turns. As such, eye gaze may be seen as a signal of conversational attention for turn-taking purposes.

However, the signaling of conversational attention for turn taking is not the only communicative function of eye gaze. Argyle & Dean (1965) suggested that gaze is also used as a means for regulating joint arousal, a function associated with intimacy and awareness. Gaze is a salient stimulus of arousal that is hard for receivers to ignore. In a study on the effect of gaze on multiparty conversations Vertegaal et al. (2000) found that people took more speaking turns when there was more visible gaze at their eyes. This more speaking turn might be attributed to the fact that subjects felt more comfortable to speak when they experienced more gaze (See Explanation 1 on page 5). However, the result that more visible gaze generates more speaking turns could also be explained by a regulatory function of gaze. Subjects who experienced more gaze could better recognize their partners' conversational attention, and determine whom they were speaking or listening to (See Explanation 2 on page 6). This would allow them to obtain and release the floor more easily. In this thesis, we will try to unravel the individual contributions to the observed effects of eye gaze on attributions of the above explanations.

The need to support gaze communication in mediated systems, and the lack of fundamental knowledge on gaze function in multiparty conversations has spawned several related issues. What kind of models should we use to animate eye gaze behavior in multi-agent, multi-user systems (Vertegaal, 2000, 2001)? Do we need to present gaze that is synchronized with conversational attention, or is it sufficient to use random gaze models that generate certain levels of gaze at users? This thesis project was designed to answer these questions as well.

1.2 Research motivations

According to earlier studies (Exline, 1963; Argyle & Dean, 1965; Kendon, 1967; Kalma, 1992; Vertegaal & et al., 1999, 2001), gaze serves as one of several non-verbal cues for yielding the floor, attempting to obtain a speaking turn, or signaling who should be the next speaker. People spend more than half of the time gazing at each other during conversations. During two-party conversations, people spend 75% of time gazing at speaker while listening, and 41% of time gazing at the listener while speaking (Argyle, 1967). In multiparty situations, because of the introduction of a third or fourth person, turn taking patterns become more complicated. Unlike dyadic conversations where the second person always takes the next turn, in multiparty conversations the next turn is not guaranteed to be a specific non-speaker.

In a multiparty setting, any turn taking cue might function effectively only if information about the addressee is included. This poses problems for the regulation of turn taking. When we consider a speaker yielding the floor in a multiparty situation, the question rises to whom he or she would yield the floor, i.e., who would take the next turn. The speaking pattern needs to be associated with the attentive information to signal the yielding or taking of the floor. It has long been presumed that gaze provides critical information in resolving this process (Kalma, 1992; Kendon, 1967).

To verify the potential relationship between gaze and turn taking, Vertegaal et al. (2000b) measured the number of turns taken by subjects during triadic mediated conversations with different levels of gaze. They found that subjects took fewer turns when the level of gaze was below normal. Also, subjects indicated that they were less able to perceive who was talking to whom when the level of perceived gaze was lower than normal. Vertegaal suggested that two gaze functions might explain the research findings:

1) Gaze encodes conversational attention

The speaker or listener used gaze to signal conversational attention, thereby regulating conversations. When someone is listening or speaking to an individual, there is a high probability that the person he looks at is the person he listens to (88% chance) or speaks to (77% chance). This information does not seem to be encoded by other non-verbal means.

2) Gaze affects level of intimacy

Intimacy is a joint function of gaze, physical proximity and other components. Low levels of gaze made it difficult for subjects to keep the equilibrium of intimacy. It makes them feel uncomfortable and less inclined to take the floor.

In order to validate these explanations for the effects of gaze on speaking behavior in multiparty conversations, we designed an experiment that compared behavior of subjects on a number of conversational variables between two triadic mediated conditions: 1) one in which gaze by their conversational partners was synchronized through time with the conversational attention of those partners, and 2) one in which gaze was randomized. If one would find more speaking turns in the condition in which gaze was synchronized with turn taking this would of course be in support of Explanation 1 (See page 3), otherwise Explanation 2 (See page 3) might be supported if we could find a significant relationship between the level of gaze and speech activities.

1.3 Research hypotheses

The above leads to the following hypotheses for this research:

Hypothesis 1: Subjects are significantly more likely to speak when gaze behavior of conversational partners is synchronized through time with their conversational attention.

Hypothesis 2: Subjects are significantly more likely to speak when their conversational partners gaze more.

Hypothesis 3: Subjects are more likely to participate in the conversation when gaze behavior of conversational partners is synchronized with their conversational attention.

Hypothesis 4: Subjects in the synchronized gaze condition gaze significantly more at conversational partners than those in the random gaze condition.

Hypothesis 5: Subjects gaze significantly less at their conversational partners when their conversational partners gaze more.

Hypothesis 6: Subjects in the synchronized condition have more positive evaluation of the conversation and their conversational partners than those in the random condition.

Aims and goals

The present research has following goals:

1) To understand the function of gaze in group conversation If conveyance of gaze is a fundamental requirement for a group communication system, what would be its temporal precision?

2) To obtain the fundamental knowledge of video mediated system design The current video mediated systems lack support of gaze communication (Vertegaal, 1999, 2001). The participants do not establish eye contact like they do in a normal face-to-face conversation. It is presented as the parallax problem. This research aims to provide clues as to the solution of the parallax problem that causes this issue (Vertegaal, 1999, 2001).

3) To try a new methodology for gaze research The errors in the measurement and observation of gaze and conversational behavior may lead to inappropriate interpretation of gaze functions in empirical studies of group communication (Sellen, 1992). Unlike previous researchers who observed the natural occurrence of gaze, we manipulate the level and synchronization of gaze itself, something that is very difficult to achieve in natural observation.

In this thesis, we will review the relevant literature firstly, after which we will present our experimental design and implementation. After presenting our results, we will discuss our findings and conclusions about effects of gaze on speaking behavior in triadic conversations.

2 Review of Gaze and Conversation Research

2.1 Normal gaze patterns in human interaction

In 1967, Kendon published a classic paper that discussed gaze behavior in face-to-face conversations. He designed an exploratory study to investigate the regularity in gaze and eye contact. Kendon found that subjects spent 50% of their time gazing at their partners, ranging from 28% to 70% of time. In general, subjects gazed more while they were listening than while they were speaking. The average duration of individual glance was also greater during listening than during speaking. He found that the proportion of time spent in eye contact ranged from 10% to almost 40%, with an average of 23%. Eye contact was typically 1 second on average, with a range of .7 seconds to 1.4 seconds. Nielsen (1962) reported that people spent about 50% of time gazing at their partner during dyadic conversation, with 62% of time spent gazing while listening and 38% of time spent gazing while speaking. In multiparty conversations, Vertegaal et al. (1998) found that subjects gazed about 7.3 times more at the person listened to (62.4%) than at others (8.5%), and gazed about 3.3 times more at the addressed person (39.7%) than at others (11.9%).

When Kendon examined the individual gaze behavior within the conversation context, he found that both duration and frequency of gaze of the first person were strongly associated with those of the second: the more frequently one person gazed, and the longer his gaze was, the longer and more frequent were the gazes of the second person.

Besides the frequency and duration of gaze, researchers studied the gaze pattern during utterances, and at turn taking boundaries. Through examination of gaze behavior at the beginnings and endings of long utterances, Kendon found a clear pattern: the speaker generally gazed at the listener when he came to the end of his utterance; a new speaker generally looked away after he began to speak and continued to do so until he was well under way. In his experiment, Kendon found that more than 70% of the utterances began with the speaker looking away at first while 70% of the utterances ended with the speaker gazed at listener (Kendon, 1967). He explained that, through averting gaze, the speaker could concentrate on the organization of the utterance, and signal his intention to hold the floor. This, however, does not mean that speakers do not look more at their interlocutors than at anything else. At the end of utterance, in gazing at listener, the speaker intends to seek responses from his interlocutor, and signal to him that he is ready to yield the floor.

Previous studies had demonstrated that, in human conversation, gaze signals conversational attention in a way that is synchronized with the speech pattern. However, it is not clear how gaze affects conversational behavior.

As we discussed, in human interaction, the frequency and duration of gaze varies considerably among individuals. As Kendon & Cook (1969) showed, each individual behaved consistently from situation to situation. Some people simply gaze more and others gaze less. Sex and personality traits such introversion-extraversion are amongst the factors affecting predispositions in gaze behavior.

Sex

With regard to sex differences, earlier studies (Kendon & Cook, 1969; Argyle & Williams', 1969; (Argyle & Ingham, 1972; Levine & Sutton-Smith, 1973; Coutts & Schneider, 1975; Russo, 1975) suggested that, in multi-party interaction, women tend to gaze more than men. Both males and females, gazed more at female partners than male partners and women felt that they were observed more than men.

Two interpretations were suggested for the sex difference. Firstly, gaze is largely determined by motivation. Women may have a stronger need for affiliation than men, and that need for affiliation promotes gaze. Also, eye contact seems to be less threatening to women than to men. Hence, women gaze longer than men. Secondly, people tend to gaze more while listening than speaking. When we listen to someone else, we have more resources available for the necessary information than when we are speaking. Women are characterized as more dependent on social cue and feedback. Hence, they tend to listen more, and thus gaze more (Exline, 1963).

Extraversion/introversion

Most studies on extraversion and gaze support an effect of extraversion on gaze. Kendon & Cook (1969) demonstrated a positive correlation between extraversion and frequency of gaze, while Argyle & Ingham (1972) did not find a significant effect of extraversion. In a between-group approach, Mobbs (1968) used three groups of subjects with extravert, neutral and introvert personality scores and found that extravert people gazed longer. Rutter, Morley & Graham (1972) suggested that extravert gazed more frequently than introvert and had more eye contact while they were speaking. Vertegaal (1998) did not find a significant positive relationship between extraversion and time spent gazing at the addressed individual, but a significant negative relationship between extraversion and time spent gazing at others rather than the individual listened to.

As discussed above, although the research findings did not reveal a clear and consistent picture, personality does seem to affect gaze behavior. Similar to the explanations for sex differences, researchers used motivation and information processing characteristics to interpret the effect of personality. Extraverts and highly affiliative people need more social interaction than others, and they gaze more to encourage others to interact with them. Personality influences the attitude toward feedback and outside information. Extravert and highly affiliative people depend more on the feedback and outside information than others, hence they gaze more. However, a much more basic explanation is that of arousal mechanisms – introverts have a high natural arousal and extraverts have a low arousal. We will further discuss this explanation in section 2.3.

2.2 Functions of gaze: regulating and monitoring conversations

Regulatory functions

The function of gaze in regulating conversation flow was widely investigated. Kendon (1967) designed an experiment to explore the gaze pattern during the beginning and ending of utterance, and found that people tended to look away from the person before they began utterance. He attributed this gaze pattern to the avoidance of information overload. In this pattern, people could shut out information and avoided distraction while organizing their thought, and signaled their partners that they intended to hold the speaking turn. He also found that, at the end of their speaking turn, people tended to gaze at their conversation partner with a "prolonged" gaze. Kendon concluded that, at the end of speaking, the speaker was ready for a response and willing to yield the turn to the partner. Kalma (1992) introduced a personal influence factor into a series of triadic conversation experiments to investigate the "prolonged gaze" pattern. He revealed that persons with a high position of influence, either obtained naturally during the conversation or through an assigned role, were more likely to show a "prolonged gaze" pattern at the end of speaking. He also found that a person who showed "prolonged gaze" would yield the floor. Subjects who received "prolonged gaze" were more likely to take the floor.

Rutter et al (1978) investigated the gaze pattern during floor changes and found that, overall, 65.6% of floor changes began with the speaker gazing, 75.9% with the listener gazing, and 51.3% with both gazing; and 68.7% ended with the new speaker gazing, 66.1% with the new listener gazing, and 48.4% with both gazing. Vertegaal et al (2001) measured subject gaze at the faces of conversational partners during four-person conversations, and found that, when someone was listening or speaking to individuals, there was indeed a high probability that the person looked at was the person listened to (p=88%) or spoken to (p=77%). They concluded that gaze was an excellent predictor of conversational attention in multiparty conversations. In an earlier study, Vertegaal et al. (2000) designed three conditions: motion video only, motion video with gaze direction, and still image with gaze direction. The subject amount of speech and subject number of turns was measured as dependent variables. They found differences in the subject number of turns across conditions. Subjects in the still image with gaze direction condition took more speaking turns (7.7) than those in motion video only condition (5.9). This was because the still image condition showed parallax free gaze. Vertegaal used both conversational awareness – awareness of who is speaking to whom and who is listening to whom, and social intimacy as possible explanations to his findings. In the still image with gaze direction condition, gaze direction signaled conversational attention, which improved conversational awareness of subjects. In this way, gaze regulated the turn taking and stimulated subjects to take more turns. In addition, when subjects experienced more gaze in a still image condition, they felt more involved in the communication process, and then would be more comfortable to speak.

Monitoring functions

Of course, people predominantly use gaze to collect information about the person being gazed at and to seek feedback from their interlocutor. This is considered the monitoring function of gaze. As discussed, people gaze more when listening than when speaking (Nielsen, 1962; Kendon, 1967). In gazing at the speaker, the listener can gather information about how the speaker is behaving. Several studies were designed to investigate the monitoring function. Argyle et al. (1973) designed a paired communication setting where one subject sat in the darkness and the other sat in the normal lighting room. The first subject could see the second one but the second subject could not see the first. The result was that subjects who could see others gazed more than those who could not see. This finding provided some, albeit fairly obvious, evidence for the monitoring function of gaze. When people can see others, they are capable of gathering information and will gaze more. Rutter & Stephenson (1979) manipulated relationships between two people and also found evidence for the monitoring function of gaze. They assumed that, if gaze served as an emotional expression, friends should gaze at each other more than strangers. On the other hand, if gaze was more concerned with gathering information, friends should gaze less than strangers since they were already familiar with each other's responses and did not need to check them frequently. Since they expected the monitoring function to be more significant, they predicted that friends gazed at each other less than strangers did. They found indeed, friends gazed less than strangers in terms of time spent in gaze and eye contact, mean duration of gazes and the proportion of time spent in gaze when listening.

2.3 Functions of gaze in conversation: emotional expression

A gaze may signal fear, concern or worry. People may use gaze to present their willingness to establish a closer relationship. In human interaction, gaze is used to convey emotional expression.

Intimacy model

There is a large amount of literature on the emotional functions of gaze. While most papers analyzed this function in isolation, a few researchers considered gaze in context and made significant theoretical contributions. Argyle, Ellsworth and Patterson are among these.

Argyle (1965) was the first person to contribute significantly to a theory of emotional gaze behavior. He developed an intimacy model to describe gaze function in context, and investigated its relationships with a range of other variables. In his intimacy model, Argyle states:

> "There are both approach and avoidance forces behind eye contact. The approach forces include the need for feedback, affiliation, etc. The avoidance components include fear of being seen, fear of revealing inner states, etc. It would be expected that there should be an equilibrium level of eye contact for a person coming into social contact with some second person. (Two persons may have different equilibrium amount of eye contact.) The increase or decrease of eye contact to

certain amount will break the balance and cause anxiety. The similar consideration is applicable to other type of behavior. There is equilibrium point of physical closeness, of amount of smiling, and of affiliation. The equilibrium of intimacy is a joint function of eye contact, physical closeness, intimacy of topic, smiling, etc. For any pair of people, there is an equilibrium point of intimacy. If one component of intimacy changes, one or more other components will change in the reverse direction in order to maintain the equilibrium."

Many studies have been conducted to examine Argyle's intimacy model. People typically manipulated one or more of the components mentioned in the model and investigated the changes of gaze pattern that would provide. Conversation topics, the amount of smiling and physical distance were the most manipulated variables in these studies. For example, when the topic is personal or private, subjects tended to gaze less during their replies (Exline, Gray & Schuette, 1965). Carr & Dabbs (1974) found that the intimacy of topic did not affected the duration of eye contact. However, it did have a negative effect on the frequency of gaze. The relationship among three components of the intimacy model -- intimacy, smiling and gaze was also investigated. It was found that innocuous topics produced more gaze while speaking, but did not have effects on either gaze while listening or while smiling (Schulz & Barefoot, 1974). The intimacy of topic had no effect on smiling behavior, but did have effect on gaze (Goldberg & Wellens1979).

In Argyle's model, physical distance is a component of the equilibrium of intimacy. The change of physical distance may change gaze behavior and other components of intimacy. It could be expected that, to maintain the equilibrium of intimacy, subjects gaze more with an increase of physical distance, in terms of mean duration of gazes, duration of eye contact and duration of gaze (Hall, 1966, Argyle & Dean, 1965, Schulz & Barefoot, 1974, Patterson, 1977). The general finding was that there was a positive correlation between physical distance and eye contact, so the greater the distance, the longer the duration and the higher the frequency of eye contact.

The earlier studies produced mixed results. Some supported the intimacy model while some contradicted it. The inconsistency among results could be attributed to different approaches and methodological problems. Investigators typically used an observational approach to score gaze behavior. With the change of subject's position, the direction and distance of observation was also changed. It was possible that this led to inaccurate recordings. The other arguable point was the relationship between subjects involved in the experiment. The closer relationship the subjects had, the shorter distance they expected. These two factors might affect the results.

Some studies (Exline, Gray & Schuette, 1965, 1968; Strongman & Champness, 1968; Fromme & Schmidt, 1972; Rubin, 1978) used typical encoding approaches to analyze the relationship between gaze and emotional variables, such as attraction, approval, embarrassment, shame and sorrow. They found attraction and approval had positive relationships with gaze; while embarrassed subjects looked less than others did. When subjects were asked to tell a sad or angry story, they looked down more than when they were asked to tell about a recent happy event.

Attribution model

In human communication, gaze has three key properties: salience, arousal and involvement. Gaze is a salient stimulus, which arouses the receivers and is difficult to ignore. However, it is not a specific emotional activator. The receiver typically interprets it in context, and responds accordingly. Ellsworth suggested an arousal concept to interpret this process. A gazes at B → B becomes aroused → B tries to interpret the arousal → B infers A's attention → B interprets A's attention by means of contextual cues. Patterson (1976) adapted the arousal concept and suggested an attribution model, a reason behind this arousal stimulus. The receivers of gaze interpret this stimulus in the context, and make attributions accordingly. Patterson was concerned with interpersonal change, and argued that a change in person A's intimacy behavior would sometimes arouse B and sometimes not. If B were aroused, he would make an attribution, either positive or negative. There was no attribution and no change in behavior if there was no arousal.

As studies revealed, the intimacy model did not provide conclusive explanations for the emotional function of gaze. The model was imprecise in what it predicted. For example, as Carr & Dabbs (1974) found, an intimacy of topic did not lead to a shorter duration of eye contact. However, the attribution model makes the explanation to the results from both encoding and decoding approaches simple. When we become aroused, we make an attribution. If the attribution is positive we look more, and vice versa.

2.4 Gaze in mediated conversation

There have been few experiments that studied gaze in mediated conversation. These studies typically used decoding approaches with a between-group design. Vertegaal et al. (1998) designed a multiparty conversation experiment with three conditions: motion video only, motion video with gaze direction, and still image with gaze direction. It was found that subjects in the still image with gaze direction condition took more speaking turns than those in motion video only condition. Since subjects in the still image with gaze direction condition experienced four times more gaze than those in the motion video with gaze direction condition, they performed a covariance analysis, adjusting the subject turns for differences between conditions in the mean percentage of gaze. With the effect of gaze removed, analysis of covariance no longer showed significant differences across conditions with regard to the number of subject turns. However, there was a significant linear relationship between the percentage of gaze and the observed number of subject turns ($r = .34$, $p<.05$). They argued that the level of gaze has a direct effect on the number of subject turns.

Novick et al. (1996) found a "mutual break" as the most frequent gaze pattern during turn taking in dyadic conversations. As one conversant completes an utterance, he or she gazes at the other. Gaze is momentarily mutual, after which the other conversant breaks mutual gaze and begins to speak. They formalized the "mutual break" pattern as a computational model of turn taking. The simulations of this conversational model demonstrated that computational agents could successfully coordinate their conversation using gaze actions to signal their intention to give or keep the turn.

Colburn et al. (2000) compared looking behavior of subjects between two dyadic mediated communication conditions. In one of the conditions an avatar produced a fixed gaze at the subject. In another condition an avatar produced gaze synchronized with speaking turns, using a probabilistic model. They found subjects themselves gazed slightly more when they were listening to the synchronized avatar. In a related experiment, Garau et al. (2001) compared subjects' evaluations of dyadic conversations in a number of mediated conditions: video, audio-only, and two conditions in which an avatar was used. The two avatar conditions differed only in their treatment of eye gaze. In the informed gaze condition the avatar's head and eye animations were synchronized with the turn taking process; in the random-gaze condition they were not. They found that the inferred-gaze avatar significantly outperformed both the random-gaze model and audio-only conditions on several response measures. They concluded that an avatar whose gaze behavior is related to the conversation is a marked improvement upon an avatar that merely exhibits liveliness. Their experiment did not provide measurements of speaking behavior.

2.5 Summary

Studies on gaze in human communication went through several stages. At the first stage, people concerned themselves with the several general patterns and individual difference of gaze in social interaction. Then the focus turned to context-specified gaze behavior. Three functions of gaze – emotional expression, regulatory and monitoring functions – were widely investigated in different context. Several theoretical models were proposed to interpret these gaze functions. In this thesis, we focus on the function of gaze in mediated conversation, a relatively new stage in gaze studies.

3 Research Design

In this chapter, we will present the rationale for our experiment design, research methodology, the stimulus and experimental task, and independent and dependent variables.

3.1 The rationale for our experiment design

Controlled experiment vs. observation

Both observation and controlled experiment approaches have been used in studies on gaze functions in group conversations. In our study, we used a controlled experiment in which gaze was manipulated.

In the observation approach, the behavioral ethogram, which includes gaze behaviors, facial expressions, head movements and positions, hand and arm movements and positions, trunk movements and positions, and vocalizations, are recorded either by a video camera set behind a one-way mirror or by human observers. The observation approach has some limitations. Firstly, it limits the research topic. Gaze cannot be separated from other variables in the observational method, which is critical to the related investigation of gaze function in group conversation. Secondly, one cannot manipulate gaze patterns in an observation setting. The third limitation falls in the measurement of experimental variables. In an observational approach, the accuracy and objectiveness of observation can never been guaranteed.

Problems are not just technical, but also contain reliability of measurement. Hence, we used a controlled experiment rather than an observational approach in this study.

Decoding vs. encoding

People have used both encoding and decoding approaches in gaze studies. The encoding approach focused on the sender of gaze. Gaze was scored by an observer as a dependent variable, and its relationship with conversational variables was analyzed. The decoding approach focused on the interpretation of gaze. It was concerned with the receiver of gaze, and examined whether variances in the sender's gaze – now the independent variable – had measurable effects on the behavior of receiver. Researchers typically used a between-group design, in which confederates were used to present different gazes and subjects were exposed to the conditions accordingly. The advantage of using confederates was its suitability to manipulate gaze as independent variable. We were concerned with the effect of gaze on conversation and therefore applied the decoding approach in this study.

Triadic conversation

Since we wanted to study group conversations, we needed at least three participants. However, the more people we used, the harder control would be. In a group conversation with more than three people, side conversations in a subgroup are inevitable. When a side conversation occurs, it is hard to register speech activity of each individual properly. Therefore, a triadic conversation was the most appropriate setup for our study.

To implement a controlled triadic conversation experiment, we used two actors. Besides the advantage of allowing us to manipulate the independent variable, the use of actors requires significantly fewer subjects, and would simplify experimental control.

Controlling gaze presentation through still images

In our experiment, gaze was controlled during a conversational session. Neither real people nor video images are capable of being a gaze agent. As we discussed before, video images do not provide proper gaze direction because of the parallax problem. The problem with using real people as gaze agents is obvious – it is impossible to systematically manipulate the gaze behavior of humans.

We decided to use carefully manipulated still image presentation instead. Through selection of different images, gaze direction can be presented to the receiver very precisely. The other advantage of using picture images is the capability of separating gaze from other non-verbal conversational cues. Through this separation, we are able to analyze the causal relation between gaze and conversational behavior. If the only variance is gaze, we can attribute the effect solely to gaze functions. Through selection of different images, gaze can be presented in accordance with conversational attention of actors. The demonstration of possible gaze selection was as follows:

Figure 3-1 The four possible gaze selections

a) actor 1 looking at the subject
b) actor 1 looking at actor 2
c) actor 1 looking away
d) actor 1 looking down

Experimental conditions

Two conditions were used in our experiment:

- Synchronized condition. The gaze direction of actors was synchronized with their conversational attention. In this condition, the "gaze at subject" image was selected with a high probability when the corresponding actor was speaking or listening to the subject. The actual gaze pattern was driven by a gaze model, which took speech signals and score of the actors' conversational attention as input and generated corresponding image selections. We will discuss this gaze model in section 4.2.

- Random condition. In this condition, the picture selection, and thus the gaze experienced by the subject had no relation with the conversational attention of actors. Gaze was randomly presented to the subject through selection of different picture images, while ensuring that the level of gaze the subject experienced in this condition approximated that of the synchronized condition.

Two groups, which were matched on personality score, language usage skill, and the proportion of sex, were assigned to our two gaze conditions. As such, a between-group design was used to compare the conversational behavior of subjects between the two conditions.

3.2 Subjects

Most subjects were students from Queen's University. All participants were paid volunteers. The two actors were also students.

We recruited subjects on the basis of following criteria.

Conversational capability and vision ability

We used a registration questionnaire to screen out subjects whose oral English was not fluent, and subjects with vision problems. People requiring hard contact lenses were also screened out since they could not go through calibration. The registration questionnaire also screened out people who had experience with video conferencing tools. The previous experience of subjects might affect their conversation behavior as well as their conversational satisfaction. This effect might mix with the effect of our experimental manipulation.

Language skill and introversion/extraversion

A pre-experiment questionnaire was used to measure subjects' language skill and personality, and matched subjects to two groups. The questionnaire consisted of a language usage test and a personality test. The language usage test, which is a part of the DAT (Differentiated Aptitude Test) Canadian version, consists of forty sentences. Each sentence is divided into four parts. In many sentences, one of the four parts contains an error in punctuation, capitalization, or grammar. The subjects were required to find the error ones. The personality questionnaire was a "Big Five" Personality Test, and consisted of one hundred Likert scale questions measuring the following five personality traits:

- Scale I – Extraversion/Introversion
- Scale II -- Agreeableness
- Scale III -- Conscientiousness
- Scale IV -- Emotional Stability
- Scale V -- Intellect Autonomy

We used extraversion/introversion scores and language usage scores to assign subjects in two groups, in a way that matched groups on these variables.

Eye fixation calibration

In our study, subject's gaze would be measured as a dependent variable. An eye tracker system was used for this measurement. Persons who were not able to pass the calibration were filtered out from our subject pool. The principle of operation of the eye tracker system and operationalization of gaze measurement will be discussed in gaze registration section in 4.1.2.

After screening, 47 people were recruited for the experiment. Thirty-four subjects finished the experiment with complete data. The others were used for testing the experimental systems. Among 34 subjects who produced complete data, there were 18 males and 16 females aging from 18 to 45 years. Twenty-eight people were native

English speakers, and six were non-native speakers with a good command of English. The other information of subjects are described in the table below:

Item	N	Minimum	Maximum	Mean	S.D.
Age	34	18	45	25.91	6.61
Language skill	34	21	39	31.91	5.34
Introversion-extraversion	34	-12	32	6.71	12.00
Agreeability	34	-15	33	17.82	9.76
Conscientiousness	34	-15	32	8.18	11.64
Emotional Stability	34	-24	35	13.88	13.06
Intellect Autonomy	34	-20	35	11.71	12.44
Valid N	34				

Table 3-1 The descriptive statistics of subjects

Subjects were allocated to two experimental groups, which were matched in following aspects:

- Language skill
- Personality (extroversion-introversion)
- Proportion of sex
- Age

The distribution of subjects in two groups is shown in Table 3-2.

Item	Synchronized group	Random group
Age (year)	27.41 (8.56)	24.41 (3.47)
Male (n)	9	9
Female (n)	8	8
Native English speaker(n)	14	14
Non native speaker (n)	3	3
Language skill	32.76 (5.39)	31.06 (5.31)
Introversion/extraversion	5.71 (13.27)	7.71 (10.9)

Table 3-2 The distribution of subjects in experiment groups

A t-test did not show significant differences between two groups in age, language skill and score on introversion-extraversion (See Table 3-3). As such, there is no evidence that they were not matched.

Item	t	df	Sig. (2-tailed)
Age	1.340	32	n.s.
Language skill	.930	32	n.s.
Introversion-extraversion	-.480	32	n.s.

Table 3-3 T-test for difference between subjects in two groups

3.3 Experimental task

Earlier investigators used idea generating, problem solving, judgment making, popular topic discussion, and sentence puzzle solving in their conversational studies. In some situations, the experimental task might have direct effects on conversational behavior. Tasks that were highly personal or involve conflict were more likely to affect our conversational measures.

We had the following requirements for the task. Our experimental task should involve a certain amount of conversation and require cooperation among participants, but not affect experiment measures directly. The ideal task for our experiment would be non-personal, non-emotional, and easy to be presented, controlled and measured. It would be insensitive to the manipulation of our independent variable, yet stimulate group conversation. Sentence puzzle solving was a good candidate. The sentence puzzle-solving task is well suited to our requirement for its advantages: 1) it is simple and does not interfere with the presentation of gaze behavior (Vertegaal, 2000); 2) task performance is easy to control and measure; 3) it is a non-emotional conversational task that gives a clear measure of subject performance; 4) the task requires the cooperation and interaction among all participants.

In the experiment, the subject and two actors worked collaboratively to solve as many sentence puzzles as possible. The subject and actors were exposed to different segments from an identical sentence. To solve a puzzle, the subject and actors needed to exchange what sentence segment they received, and constructed as many meaningful and syntactically correct sentences as possible. After a puzzle was completed, a new sentence fragment was displayed for ten seconds on subject's screen, right below the actor images (See Figure 3-1). Subjects needed to memorize the sentence segment after it appeared on their screen and exchange this information with their partners. The actors had all fragments and correct solutions listed on paper instead, although they pretended to have the same setting as the subject.

Throughout the experimental session, it was the subject's responsibility to submit each complete sentence to Actor 1, who pretended to enter it for verification by computer. After this, Actor 2 reported its correctness.

3.4 Operationalization of variables

3.4.1 Independent variables

In our study, synchronization of gaze was treated as a factor and the level of gaze was measured as a covariate. The effects of synchronization of gaze, the level of gaze, and their interaction on conversation were investigated.

It was our hypothesis that gaze had a function in regulating conversation. In a multiparty setting, any turn taking cue might function effectively only if information about the addressee was included. Gaze played a role in providing information about whom the speaker was addressing. Vertegaal (2000) revealed that subjects took fewer turns when the level of gaze was below normal. Also, subjects indicated they were less able to perceive who was talking to whom when the level of gaze was lower.

In our study, two gaze conditions were implemented, one in which gaze by conversational partners was synchronized through time with the conversational attention of those partners, and one in which gaze was randomized.

Both synchronization of gaze and the level of gaze were controlled, but it was important to allow the level of gaze to diverge with conversational context in the synchronized condition. The actual level of gaze was produced by a gaze model in accordance with the conversational turn taking and varied from session to session. This design had two advantages. One was to remove noise more effectively. Since we did not control the actual level of gaze at the subject, the level of gaze presented to the subject was consistent with the conversational turn taking, in this way any possible noise due to this control could be removed. The other advantage was to allow testing of two factors in a simple design.

To allow control of our independent variable, each group consisted of two double-blind actors pretending to be subjects, and only one subject who was administered treatment. The actors were trained with the task before experiment so that the possible practices effect was removed.

3.4.2 Dependent variables

We measured subject's speech, task performance, subject gaze and conversational satisfaction as dependent variables.

3.4.2.1 Speech

Speech could be measured at both structural level and semantic/syntactic level. The structural measures include number of speaking turns and amount of speech.

The speech energy produced by the subject and actors was individually registered. It is not evident that the energy produced during speech activity is a good indicator of what we consider to be a meaningful conversational contribution or *talkspurt*: a string of words produced by the same speaker. This is because throughout the articulation of speech, a speaker may introduce various moments of silence: between phonemes, between words, and between strings of words. We needed to filter these pauses to recognize speaking turns and talkspurts. The technical details of this process will be discussed in section 4.3.2.

Speaking turns

A turn was bounded by the act of exchanging the role of speaker and listener. It occurs when a new speaker has a talkspurt of one or more phonemic clauses, with others (including previous speaker) being silent for at least one phonemic clause. It consists of "a series of talkspurts bounded by a speaker switch, including the silence that may occur before the speaker switch". The number of turns has been used in studies comparing face-to-face and various mediated conversations. Researchers (Argyle, et. al. 1968, Rutter & Stephenson, 1977; Vertegaal, 1998) found the number of turns was sensitive to the way of mediation. In mediated condition in which more gaze cue was provided, subjects produced more speaking turns.

The number of speaking turns was measured for the subject as well as the group. Our hypothesis was that subjects were more likely to know who was speaking to whom in the condition in which the gaze direction was synchronized with the conversational attention. Therefore this condition would facilitate speaker switching and would produce more subject number of turns than in the random condition.

Talkspurts

Talkspurts are the smallest meaningful strings of speech. Like the number of turns, the number of talkspurts was measured for subjects as well as the group.

The number of turns reflects the general conversational structure while the number of talkspurts provides more accurate information about the contribution of each participant.

Hypothesis 1 predicted that, because of the regulatory function of gaze, subjects in the synchronized condition might produce more talkspurts than those in the random condition.

Amount of speech

Amount of speech is defined as the distribution of speech signals throughout a conversational session. It was measured by the ratio of the number of speech samples that indicated energy and the total number of speech samples throughout a conversational session. The amount of speech was measured for subjects as well as the group.

Subject participation

The subject participation is defined the ratio of subject's contribution to total speech. Hypothesis 3 suggested significantly more participation from the subject in the synchronized condition than in the random condition as a result of the regulatory function of gaze. The ratio of subject contribution to total speech was predicted to be higher in the synchronization condition.

3.4.2.2 Subject gaze

The relationship between gaze and "being gazed at" has been investigated in different contexts. In the condition in which gaze was synchronized with conversational attention, subjects experienced more natural gaze than in the random condition, possibly leading to different gaze behavior of subjects. To investigate this, we measured subject's gaze behavior as a dependent variable.

Coordinates of gaze fixation throughout an experiment session were divided into three categories: gaze at actor1, gaze at actor 2, and gaze away. These variables were calculated as a proportion of total session time.

3.4.2.3 Task performance

The subjects and two actors were asked to finish as many permutations as possible based on the sentence segments provided. Permutations were to be semantically and syntactically correct. The number of correct permutations produced by the subject was recorded as task performance and the contribution of subject. Since Vertegaal, et al. (2000) demonstrated this task to be insensitive to manipulations in gaze level, we expected no differences in performance.

3.4.2.4 Qualitative measures

After each experiment session, subjects were administrated a post-experiment questionnaire concerned with conversational satisfaction. The questionnaire included eight structured and two open questions, which involved evaluation of the conversation, the partner, the task performance and the mediated systems. The full questionnaire is provided in Appendix 6.

4 Experiment Implementation

In the previous chapter, we discussed our research design. According to this design, we will discuss the implementation of the experiment, which includes experimental setup, gaze behavior models, and measurement of dependent variables.

4.1 Experimental setup

4.1.1 Prototype

The principal purpose of this research was to explore the effect of gaze on speaking behavior in multiparty conversations. To make gaze behavior and the experiment controllable, two actors were used for all conversation sessions. Gaze behavior of the actors was simulated through the selection of picture images, driven by a probabilistic gaze model. This model took speech signals of three parties and real-time attention scores from the two actors as input, and then selected the proper picture image in accordance with the conversational flow. The actor images and task stimuli were presented on the subject's screen. (See Figure 4-5)

Experiment Implementation

Experiment rooms

Three experiment rooms were used, one for the subject, one for the two actors, and one for the experimenters. They will be referred to as the subject room, actor's room and control room respectively. The audio cables connected three rooms and made group communication possible. The video cables connected the control room and the subject room, which transferred the experimental stimuli from control room to subject's screen, and transferred subject's video image and eye positioning image to control room as well.

Audio system

The audio systems had several functions: facilitating the audio conversation between group members; recording audio signals from participants to three separate channels; providing input to the gaze model; and generating experimental instruction.

Video system

The video system consisted of a stimulus display, which presented both picture images and task stimuli (sentence fragments) to subjects, and experiment monitoring, which monitored subject's movement and activities throughout the session. The presentation of task stimuli was pre-programmed and controlled by the investigator according to the progress of task performance.

Control system

The experiment monitoring system received signals from a video camera mounted in front of subject and eye image of subject from an eye gaze system, and allowed the investigator to monitor the gestures and eye movements of subjects in the experiment session.

The control system consisted of a gaze model and experiment control model. In order to work properly, the gaze behavior model needed knowledge about who was speaking to whom and who was listening to whom. The speech signals did provide conversational information about who was speaking. However, it did not provide information about whom the speaker addressed to. In a case that the subject spoke or the group fell into silence, the system did not know the auditory attention of actors. Therefore, it was necessary for actors to explicitly score their conversational attention while speaking, listening or while they were silent. With sufficient knowledge about conversational attention, the gaze model was able to present gaze that is synchronized with conversational flow in real time.

Recording system

The recording systems registered eye gaze, speech and video signals from subjects, and speech signals and attention scores from the two actors. The eye tracker registered the gaze fixation of subjects throughout the experiment session. The gaze behavior of actors was simulated by a gaze model, and was recorded according to the actual selection of pictures. The speech signals from the subject and two actors were converted to MIDI data through a MIDI (musical instrument digit interface) interface, and recorded with a G4 Macintosh running Performer. MIDI real-time is a local area network for musical applications. We used MIDI because it records all events with associated timestamps.

All experimental data, including gaze fixations, speech, and video signals were backed up using a video tape recorder. The attention scores of the actors were recorded by MIDI and used for selection of picture images. Task performance was recorded manually.

The diagram of the experiment control system is shown below.

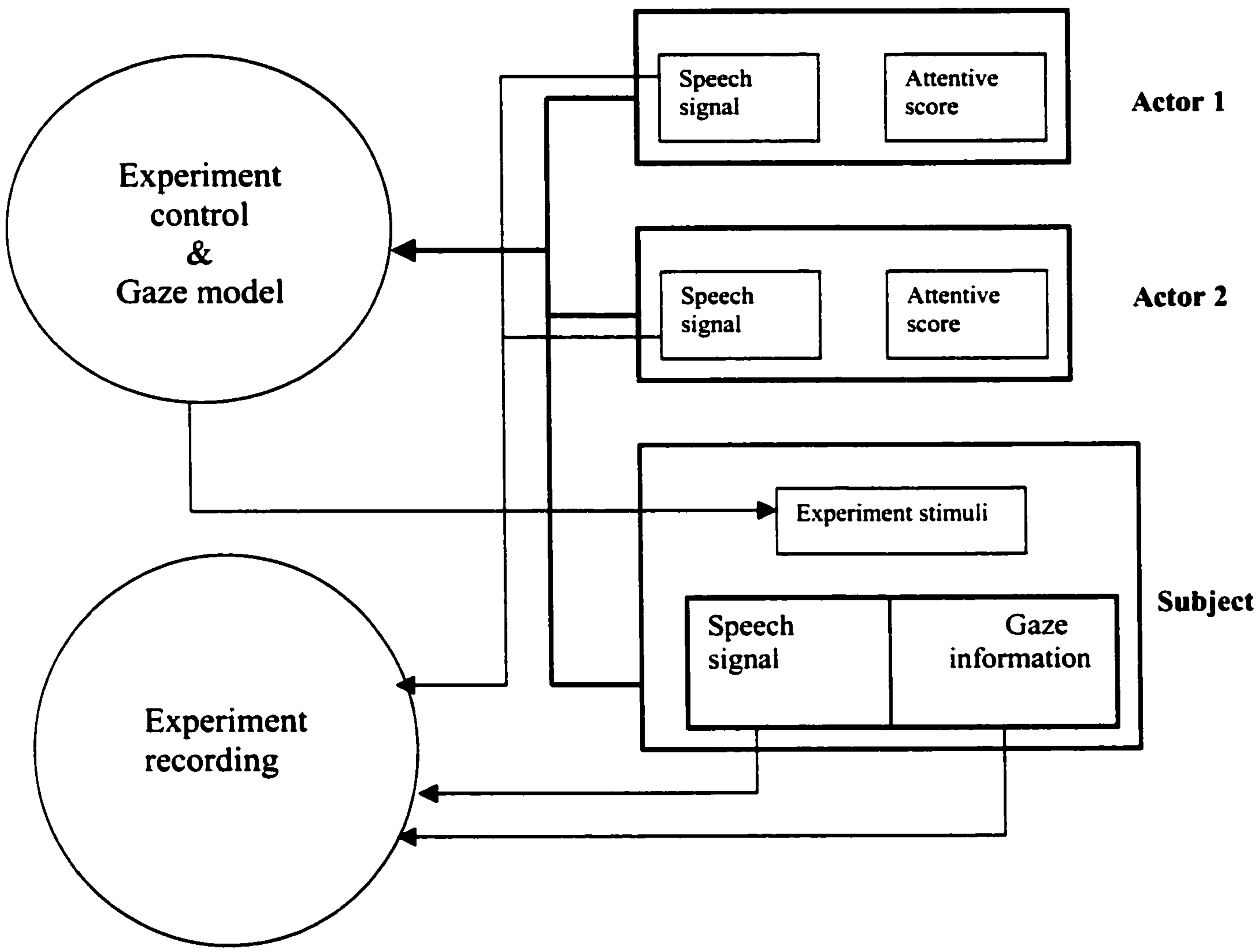


Figure 4-1 The experiment control system

4.1.2 Physical setup

The experiment system was implemented with three computers -- two Macintosh G4s and one iMac, one LC Technologies' eye tracker system, and several supplementary audio and video devices. One G4 was used to record all experimental data that had been converted to MIDI format as well as the audio data. Another G4 was used for experiment control while the iMac served for stimulus presentation to subjects, and for displaying picture images of actors and sentence segments (See Figure 3-1). The computers communicated through USB, MIDI, and serial connections. Each participant communicated with the others through audio connections. The speech signals were recorded separately for each participant, and were routed to our gaze model for real-time processing. Figure 4-2 and Figure 4-3 demonstrate the audio and video connections in the experiment setup.

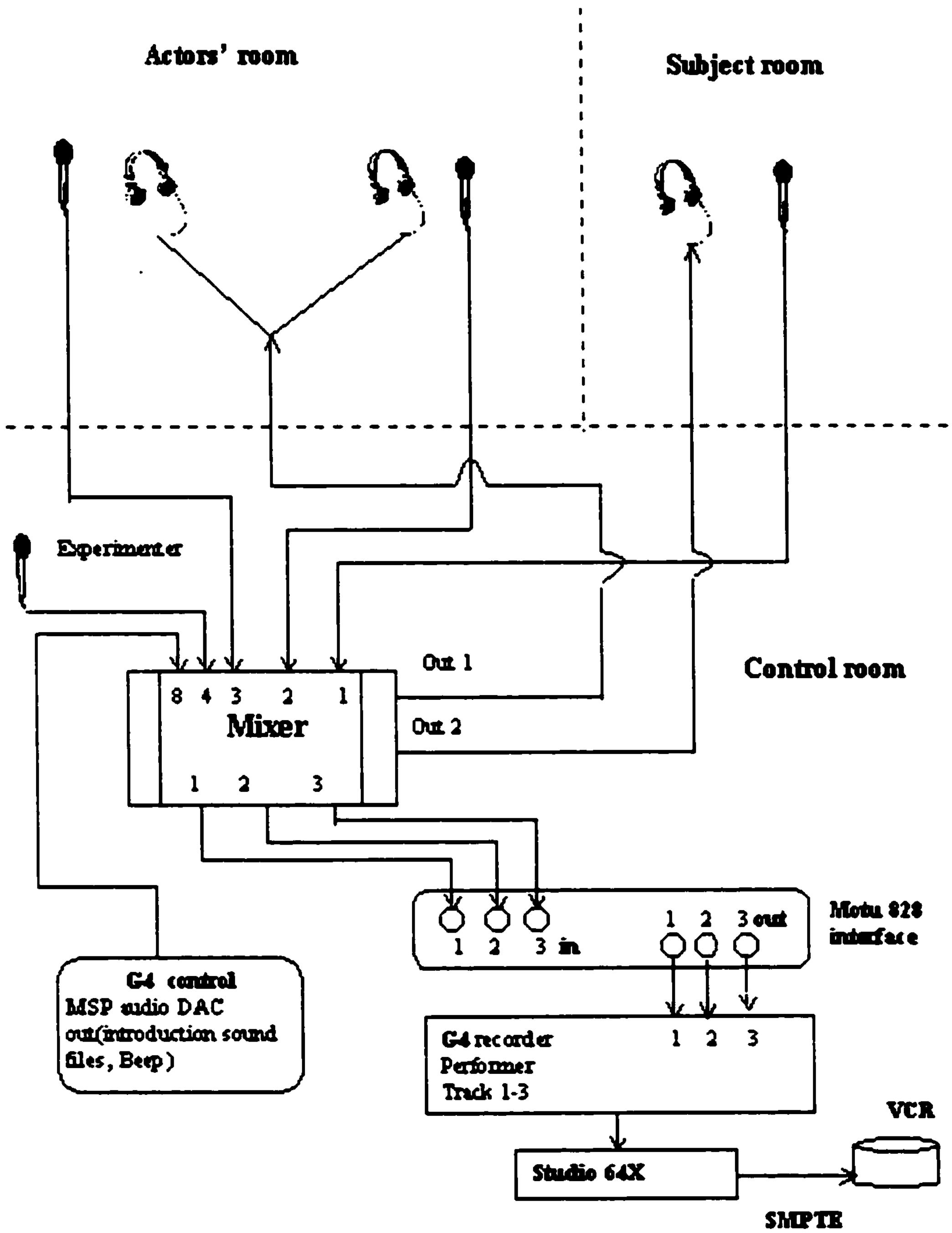


Figure 4-2 The audio connections in experiment setup

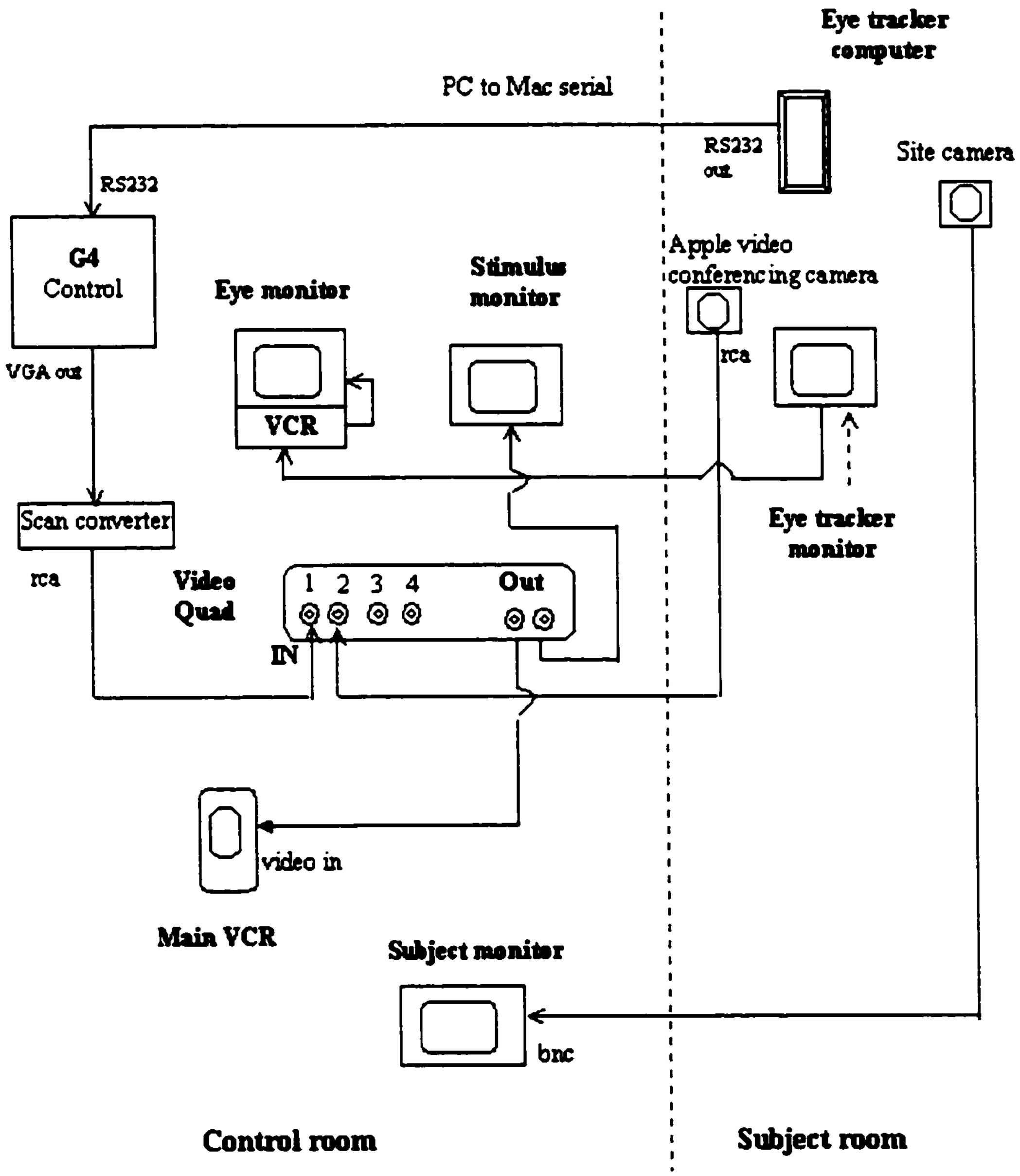


Figure 4-3 The video connections in experiment set-up

Gaze registration

The eye tracker system from LC Technologies was used to capture, store and process the gaze of subjects, and sent fixation coordinates to the recording computer through a serial cable. The system used the corneal/pupil reflection relationship to determine fixation position. On the top of the lens of infrared camera, an infrared light source is mounted which projects invisible light into the eye. This infrared light is reflected by the retina, causing a bright pupil effect on the camera image. The light is also reflected by the cornea of the eye, causing a small glint to appear on the camera image. Because the cornea is approximately spherical, the corneal reflection does not move with the eye when eye moves. However, the reflection of the retina through the pupil will move with the rotation of the eye. By processing the image on the computer unit, the vector between the center of pupil and the corneal reflections can be determined. In order to correctly translate this vector into real-world coordinates, the system needs to be calibrated. Then, based on the relative positions of the reflections of cornea and retina, the eye tracker system can determine the eye fixations of subjects on the screen. When the subject's eye did not move for 120ms, the system concluded that the subject fixated, and the coordinates of this fixation were sent to the recording sub-system.

MIDI and audio interfaces were used to convert signals from different sources into MIDI data that could be recorded in different tracks of G4 performer. Figure 4-4 shows the MIDI connections in our experiment setup.

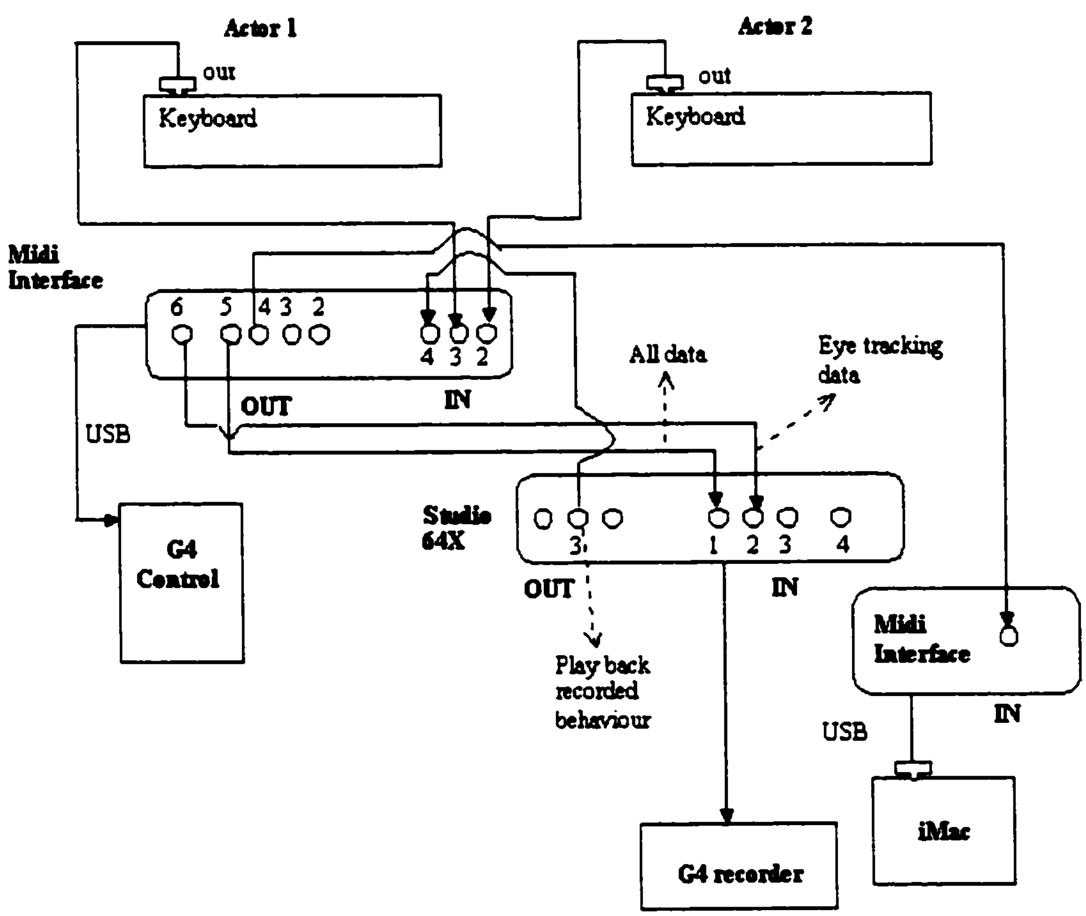


Figure 4-4 The MIDI connections in experiment set-up

Setup for the subject room

The subject room had one eye tracker system and a monitor used for eye tracker calibration and stimuli display. In each experimental session, after the calibration of eye fixation, the monitor was switched to display picture images and task stimuli. The experimental setup in the subject room is demonstrated in Figure 4-5.

Setup for the actors' room

The setup for the actors' room was comparatively simple. The devices that each actor used in the experiment were a headset, microphone, and MIDI keyboard.

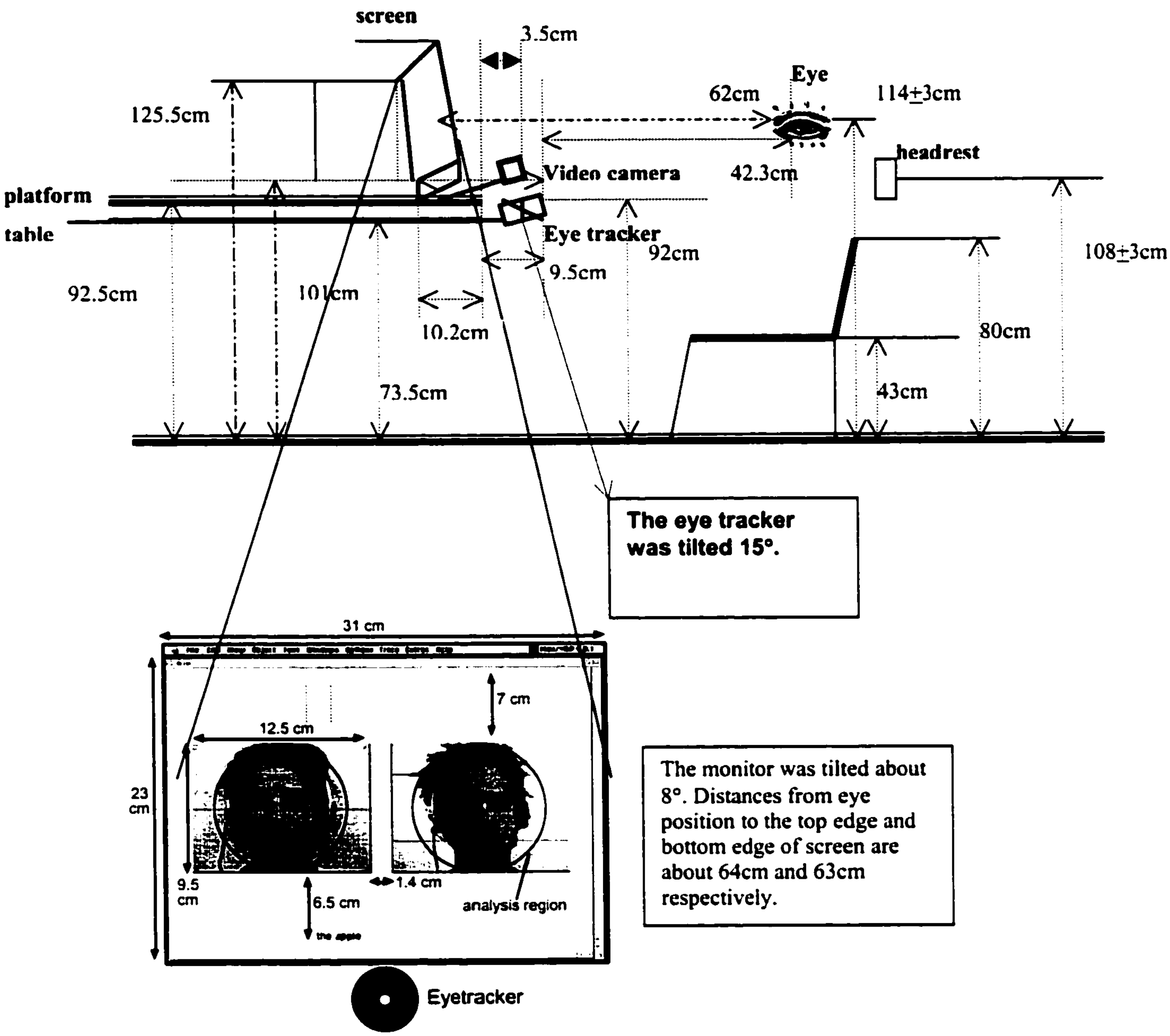


Figure 4-5 The experimental set-up and screen display in subject room

4.2 Gaze behavior models

Conversational states

In a triadic conversation involving one subject and two actors, the conversational state of each actor generally falls into one of the following states at any time:

a. Speaking or listening to the subject. The actor scored conversational attention for the subject while he or the subject was speaking.
b. Speaking or listening to the other actor. The actor scored conversational attention for the other actor while he or the other actor was speaking.
c. Speaking or listening to both participants. The actor scored no conversational attention or scored conversational attention for both participants.

During each session the actors scored, using an accord keyboard, whether they were speaking or listening to the other actor or to the subject. At the same time, speech activity by all participants was registered and processed by a real-time analysis algorithm described below. This algorithm effectively constituted a low-pass filter that determined whether a participant was in fact speaking. Using a simple set of binary logic rules, real-time conversational attention scores from the actors were combined with the outcome of speech activity analysis to determine the current conversational state for each actor.

Image selection using gaze behavior models

Although subjects believed they were communicating with the actors using a video conferencing system, they were being presented with images of the actors that were generated by the software algorithm that constituted our experimental treatment. Before the experiment, we captured four images of each actor (See Figure 3-1): (1) gaze at the subject; (2) gaze at the other actor; (3) looking away and (4) looking down.

For each conversational state a simple model of human gaze behavior determined, for each actor, which of the four images to select for presentation to the subject. In the synchronized condition the model always selected *gaze at the subject* as the first image when entering state (1), and *gaze at the actor* when entering state (2). 500 ms after entering these states the model would continue selecting subsequent images probabilistically. In state (3) the model would always use probabilistic image selection. For each image, **Table 4-2** shows the selection probabilities (*p*) of the model associated with each state and the duration with which that image appeared on the subject screen before a new selection was made. As long as the state remained the same, the system would continue to select images with the probabilities and durations associated with that state. When entering a new conversational state, however, the system would immediately select a new image. A simple low-pass filter ensured image transitions on the subject screen would never be shorter than 500 ms. In the randomized condition all data was processed the same way, but the conversational state selection output was ignored. As such, the system would remain in state (3).

Internal logic of gaze model

The operation of our gaze model is based on the transition between conversational states and between gaze states within each conversational state. At each conversational state, there are four possible gaze states; each gaze state has a certain probability to be selected. Associated with each gaze state, there is a time interval and a probability matrix that determines the selection of the next gaze state. Determining the probability matrix of gaze selection within each conversational state requires an in-depth literature study.

Probability of gaze in dyadic conversation

Argyle and Cook (1976) found that people in dyadic conversations spent 75% of time gazing at their conversation partner while listening and 41% of time while speaking, with an average gaze duration of 2.95 seconds. In MM (male-male) groups, people spent 74% of time gazing their partner while listening and 31% of time while speaking. The average gaze duration was 2.45 seconds. In FF groups, people spent a bit more time gazing at their partner while speaking (48%). In MF groups, males spent 52% of time gazing at the female partner, while the female spent only 36.5% of time gazing at her partner. In his research on visual interaction in a group, Rutter (1972) revealed differences in gaze patterns between introvert and extravert people, and between male and female subjects. Introvert females have the longest mean gaze durations (5.1-5.6 seconds), followed by introvert males (4.4-6.0) and extravert females (3.9-4.8). Extravert males have the shortest gaze duration.

Colburn, Cohen and Drucker (2000) developed a model simulating the eye gaze patterns for avatars used in virtual environments. Based on their literature review and their informal observation, they developed a state transition and state duration matrix that automatically generated gaze patterns. In their model, the avatar always gazed at the subject when the subject began to speak. It gazed at the subject 30% of the time when the avatar began to speak. The duration of gaze depended on who is speaking and what the current gaze status was, ranging from 2 to 6 seconds.

Gaze probability in group conversation

People in triadic conversations gaze at others differently than those in dyadic conversations do. Argyle and Cooker (1976) designed a triadic conversation experiment and indicated that people spent 29.8% of the time looking at the speaker and 25.6% of the time looking at others while talking in an all male group. In an all female group, people spent 42.4% of time looking at the speaker and 36.9% of the time looking at listeners. In a four-party experiment, Vertegaal et al. (2001) found percentages of gaze by a single addressed individual in a four-person group similar to those found by Argyle and others in dyadic conversations (62%). They also found that speakers in a four-person conversation gaze at others who were not addressed about 12% of time, and that the rest of gaze behavior is spent looking away (mostly sideways while thinking).

Our probability matrix

According to the earlier findings, we summarized the percentage of gaze in conversation as in **Table 4-1**.

	During speaking		During listening		Sources
	Probability (%)	Duration (s)	Probability (%)	Duration (s)	
MM	31	2.45	74	2.45	Argyle, 1972
FF	48	3.12	78	3.12	
MF(M)	52	3.61	76	3.61	
MF(F)	36.5	2.98	69	2.98	
All	41	2.95	75	2.95	
M in dyadic	–	.97-3.66	–	–	Kendon, 1967
F in dyadic	–	1.15-1.68	–	–	
MMM	25.6	–	29.8	–	Exline, 1963
FFF	36.9	–	42.4	–	
Averted, dyadic	37.8	1.61	68	2.87	Cook & Smith, 1972
Normal, dyadic	45.3	2.09	71.6	2.73	
Continuous, dyadic	59.7	2.5	81.8	3.85	
Four-party conversation	39.7	–	62.4	–	Vertegaal, et al 2001
Average[1]	43.4		72.9		

Table 4-1 The percentage of gaze in different conversational context

[1] The data from Exline (1963) are excluded.

There is a difference in the measurement of the above probabilities between studies. Most researchers (Vertegaal, 2001; Argyle, 1972; Cook & Smith, 1972) measured gaze directed at a specific individual while speaking or listening, and while others (Exline, 1963) measured the amount of gaze target at a group. We needed to distinguish between these two measures. Based on earlier findings, we calculated that the probability of gazing at a specific conversational partner while listening is 72.9%, the probability of gazing at a person addressed to is 43.4%, and the probability of gaze at a group while listening or speaking is 34%.

We decided to group measures of gaze at individuals while listening and while speaking. If we provided different probabilities for listening and speaking to conversational partners, the differences in probability might introduce an extra turn taking cue, and confound the experiment. Based on above consideration, we decided to adopt the following probability matrix for our gaze model. (See Table 4-2)

In Table 4-1, p means the probability of a gaze pattern being selected when entering a conversational state or after a selection of gaze pattern within a conversational state. The percentage means overall time of a gaze pattern being selected within a conversational state.

When an actor spoke or listened to another individual participant, there was a 67% of chance of gaze at this person. The total time of gazing at another participant, looking away, and looking down was 13%, 13%, and 7% in this conversational state.

State	Gaze at subject	Gaze at actor	Look away	Look down
(1) Speaking or listening to subject	67% p=.51	13% p=.10	13% p=.25	7% p=.14
(2) Speaking or listening to actor	13% p=.10	67% p=.51	13% p=.25	7% p=.14
(3) Speaking or listening to both	35% p=.24	35% p=.24	20% p=.35	10% p=.17
Duration of image selection (s)	2.0	2.0	.8	.8

Table 4-2 The probabilities of gaze behavior and timing in accordance with conversational states

Our model selected the *gaze at the subject* image in units of 2.0 seconds (See Table 4-2). Given the high probability of successive selections, this yielded a theoretical mean length of 2.4 s while speaking or listening to two persons, and 5.1 s while listening or speaking to one person. We arrived at the statistics presented in Table 4-2 after fine-tuning our models such that the observed average percentage of gaze at the subject in the synchronized condition was within an error margin of about 10% of that in the random condition.

In our experiment, the gaze behavior model was used only for the synchronized condition. In the random condition, the gaze presentation of two actors was randomized with the constraint that the average level of gaze approximated to that of the synchronized condition.

In each conversational state, the gaze behavior model would drive the selection of picture images, and the gaze direction would be changed among four states with the probabilities shown in Table 4-2. The figure below shows the transitions of gaze states when the conversation entering the "*speaking or listening to subject*". T1 to T4 represent the duration for each selection. Ps indicate the probability of transition from other gaze state to this state. In this conversational state, the model would always select "gaze at subject" firstly to show the conversational attention of actors. The low-pass filter was applied for any transition between two gaze states to guarantee each selection of gaze direction lasted at least 500ms.

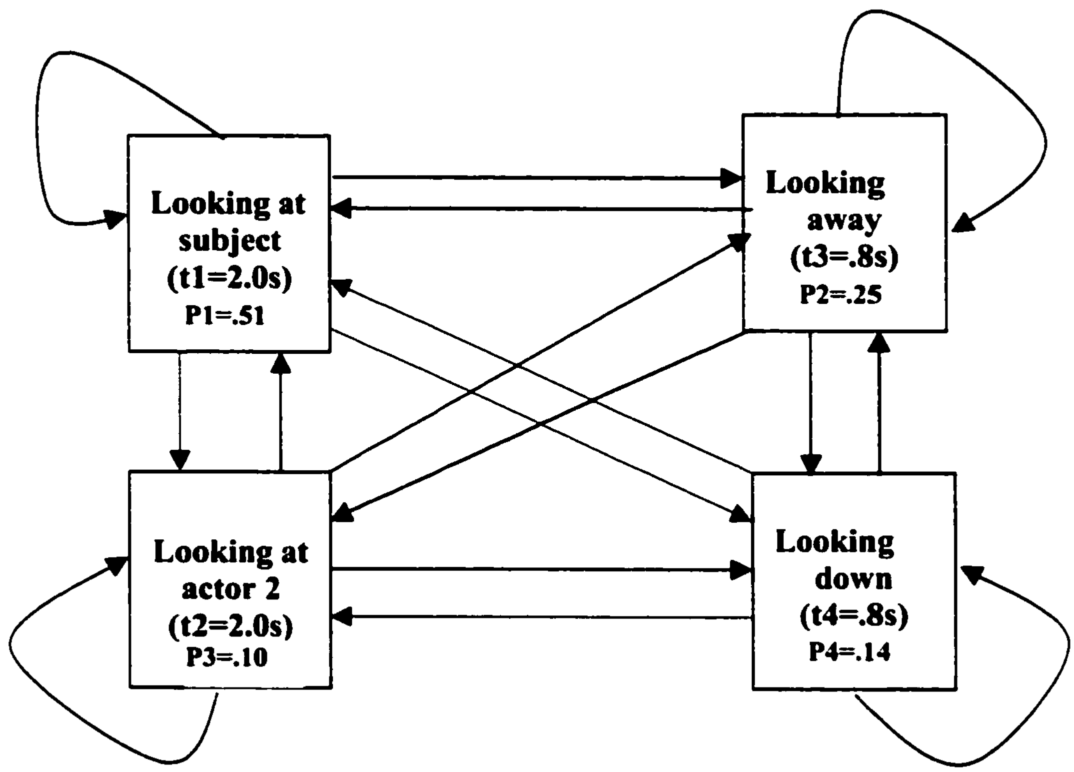


Figure 4-6 Gaze transitions while actor1 is speaking or listening to subject (p1 to p4 mean the probabilities of gaze transition from other state to this state)

Reliability and validity of conversational attention scores

The actors' task of scoring conversational attention was very similar to the one applied successfully by untrained subjects in Vertegaal (2001). Before scoring, actors were carefully instructed on the task. They practiced for more than 3 hours with 20 subjects in 20 training sessions. For each actor, training scores were compared with those made by the experimenters. The agreement of scores was calculated as a percentage of time in which scores overlapped (with a resolution of 120 ms). Pooled between actors, the average overlap was better than 87.5% of time. The scoring process was also monitored continuously by the experimenters throughout each session. The lag in the registration and image selection process was less than .01 s.

4.3 Measurement and analysis

One independent variable – the level of gaze perceived by subjects and four types of dependent variables were measured throughout the experiment sessions. The dependent variables included speech, subject gaze, task performance, and conversational satisfaction.

4.3.1 Measuring the independent variable

The level of gaze perceived by subjects was measured on the basis of the actual selection of actor images and the duration of each selection in real time. It was calculated as the ratio of time in which "gaze at subject" was selected with total session time. Due to the probability nature of the gaze selection process, subjects in different sessions might experience different levels of gaze. Image selection was based on the probability matrix shown in Table 4-2.

The length of gaze

Our model selected the *gaze at the subject* image in units of 2.0 seconds. Given the high probability of successive selections and the fact that first selection of image was always *gaze at the subject* while speaking or listening to subject, this yielded a theoretical mean length of 2.4 s while speaking or listening to both, and 5.1 s while listening or speaking to the subject.

The level of gaze

According to Table 4-2, assuming that probabilities of entering conversational state 1 to 3 are same, the level of gaze could be estimated as 38.3% ((.67 + .13 + .35)/3).
Since 1) the actual probabilities entering conversational state 1 and 2 are higher than the probability entering state 3, and 2) when entering state 1, the first selection is always "gaze at the subject", the actual level of gaze in synchronized conditions would be slightly higher than 38.3%.

4.3.2 Measuring speech activities

The number of turns, number of talkspurts, and amount of speech were measured as speech indicators.

In each experiment session, speech signals from subject and two actors were sampled with a rate of every 120 ms, recorded in separate channels, and converted to MIDI data. Throughout each experiment session, the speech data was processed to determine who was speaking at the time. During the articulation of speech, a speaker may introduce various moments of silence: between phonemes, between words, and between strings of words. In some situations, the short pause or silent period between words could be considered as a part of meaningful speech while in other cases, i.e. between talkspurts, the silent period has no contribution to the conversation (Vertegaal, 2001).

The pauses throughout the articulation of speech might occur within words, between words, and between phonemic clauses, and should be processed in accordance with conversational context.

Pauses within words

Words consist of a string of basic vocalization units, or phonemes (Trager & Smith, 1957). In between phonemes, pauses may occur. For example, the word "good" consists of three phonemes: g -/o/-d. The first two are typically uttered consecutively while there might be a small pause before the last phoneme "d". The length of this type of pause is usually less than 200ms. In our processing, this type of pause was considered a part of speech (Vertegaal, 2001).

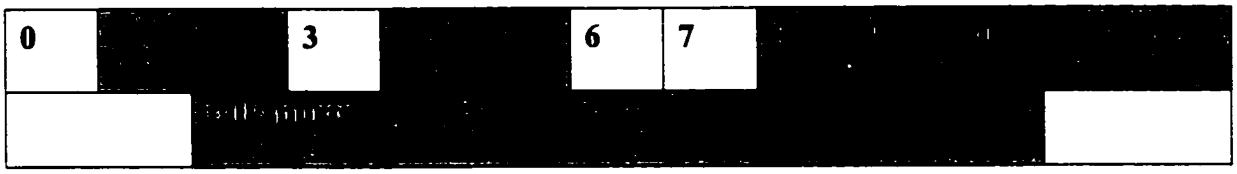
Pauses between words

Pauses can also occur between words. Whether the pauses should be considered as a part of the utterance depends on whether they belong to a phonemic clause (Sellen, 1992). The phonemic clause is recognized as the basic syntactic unit of speech. It is a string of 2-10 words (typically 5 words with a duration of 1.5 seconds), in which there is one primary stress and which is terminated by a juncture, a slight slowing of speech, with slight tone changes at the very end (Jaffe & Feldstein, 1970). The phonemic clause could be considered as the smallest string of words necessary for expressing a proposition. Hence, the pauses within a phonemic clause should be considered a part of speech.

Pauses between phonemic clauses

Talkspurts consist of a string of phonemic clauses. The speaker could pause between phonemic clauses. Whether the pauses should be considered a part of speech depends on whether they belong to a set of talkspurt by the same speaker (Sacks, Schegloff, and Jefferson, 1974).

Based on the analysis described above, we used a real-time fuzzy algorithm (used in Vertegaal's previous study in 1998) to filter pauses throughout session. The algorithm processed the speech activity of actors and subject, designating moments of silence and moments of talkspurts activity by each conversational partner. First, our algorithm filled in 240ms pauses to account for stop consonants, effectively removing pauses within words. Then, the algorithm removed



The talkspurt filter

The algorithm counts the number of samples with a value 1 (shown as gray cells in the figure above, indicated that the speech energy at this time is above the threshold) in a 13- sample window of speech activity signal. If the total is less than 7, the algorithm does nothing but shifts one position ahead in time. If the count is bigger than 7, the mean position of samples with a value 1 is calculated. For the example above, the calculation is (1+2+4+5+8+9+10+11+12)/9 = 6.89. if the calculated value is between 5 to 7 inclusive, it means that these samples are evenly distributed over the window. Then it sets all samples between 2 to 10 to 1 (shown in the large gray cell in the figure above), and shifts one position ahead in time.

Figure 4-7 A graphical demonstration of the speech analysis algorithm

(Adopted from Vertegaal, 1998)

pauses between consecutively spoken words and phonemic clauses to identify *talkspurts*, a series of phonemic clauses uttered by the same speaker. As we mentioned before, the phonemic clause is a basic syntactic unit of speech with duration of approximately 1.5 s. Considering our sampling rate of 120ms, the number of samples in 1.5 s is approximately 13. Hence, to identify talkspurts, a 13-sample (1.5 s) window moved over the speech data, filling samples within a 70% confidence interval around its mean position with speech energy if more than half of the samples in the window indicated speech activity, and if this speech activity was balanced within the window. If fewer than 7 samples in the window indicate speech activity, the algorithm does nothing and simply shift one position ahead. (See Figure 4-7 for a graphical illustration).

Measuring speaking turn

A speaking turn occurs when a new speaker has a talkspurt of one or more phonemic clauses, with others (including the previous speaker) being silent for at least one phonemic clause. To identify the speaker switch, we needed to locate the boundary silence before a speaker switch. When a new talkspurt was found, the algorithm first located the start of the talkspurt. Then it would look back in time to find speech activity by a previous speaker. This was designated the end of previous turn. It would then look forward in time again to locate the first speech activity after this position by the new speaker. This was designated the start of new turn. In this way, the number of turns could be computed on the basis of registered speaker switches.

The number of turns were measured for each participant and for whole group.

Measuring number of talkspurts

The talkspurt is constituted by a series of phonemic clauses uttered by the same speaker. Our speech-processing algorithm can identify talkspurts for each participant; therefore, the number of talkspurts can be counted automatically for all three parties.

Measuring amount of speech

Amount of speech was measured on the basis of speech signal registration. It was measured by the ratio of the number of speech samples in which the energy of the signal was above a threshold to the total number of speech samples throughout a conversational session.

Processing speech for our gaze model

Speech activity was used as an input to the gaze model. To identify the presence of speech, the algorithm would only treat talkspurts longer than an average phonemic clause (i.e., 1.5 s) as speech. To prevent an initial lag of 1.5 s before talkspurt identification, we used a version of the above algorithm with a 0-sample window that would identify the presence of speech activity within a few samples. Talkspurts would end 1.5 seconds after the current speaker fell silent or, if the speaker was an actor, immediately when the actor stopped scoring conversational attention.

Accuracy of speech registration

Speech data was recorded synchronously with timestamps. Due to the differences in the processing speed of subsystems that measured and registered speech data, the timestamps would have to be corrected to indicate the time the event happened rather than the time at which it was recorded. To correct the timestamps on the speech record, we analyzed latency by sending timed noise bursts through the system. After comparing the recorded timestamps with the times at which the bursts occurred, we did not find any measurable latency in transmitting and processing of the speech signal (error < .01s).

4.3.3 Measuring gaze behavior

The gaze behavior of subjects throughout the experimental session was measured with an eye tracker system. Before each section, the subject was calibrated with this system and with our experimental software. The eye tracker system uses calibration data to register and correct the gaze data of the subjects. We also calibrated subjects' fixation on the facial area of images of two actors. The images of actors were displayed in two rectangles at the center of screen. Before each session, two boxes bounding the images were shown to the subject. Subjects were asked to fixate on the top-left and bottom-right corners of each bounding box. This calibration data would be used to determine subject's gaze direction throughout conversations.

Accuracy of measurement

During the experimental session, the x, y coordinates of the subject's fixations were recorded with timestamps. The fixation and calibration data was used to determine whether the subject gazed at facial areas of the image. When subjects were asked to fixate on the top-left and bottom-right corner of each bounding box, the coordinates of these points were recorded. In each box, we then fitted a circle that constituted the facial region boundaries (See Figure 4-8). Our subsequent automated analysis procedure was straightforward: it registered *gaze* for a conversational partner whenever the tracked subject fixated within the circle around the facial region of that partner. All saccades were skipped.

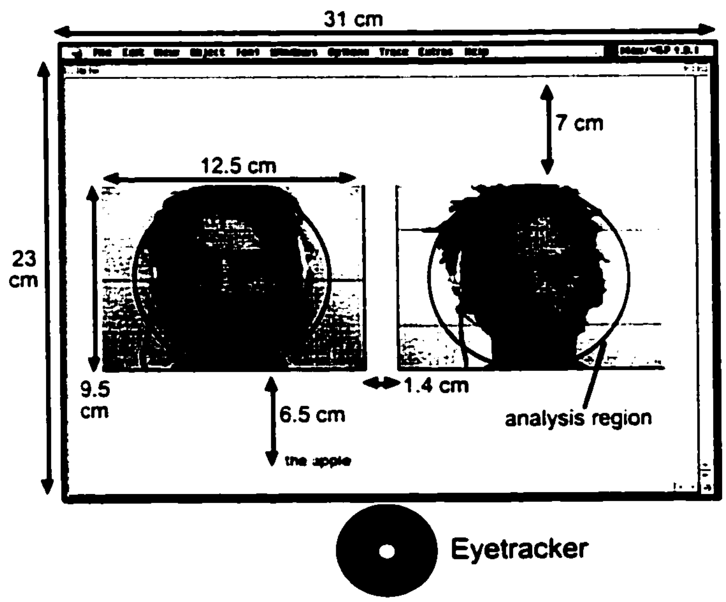


Figure 4-8 Measuring subject's gaze

Thus, the subject's gaze direction could be divided into three categories: gaze at actor 1, gaze at actor 2 and look away. Whenever the subject gazed at an actor while that actor gazed at him, this was regarded as mutual gaze.

The latency of the eye tracker and the gaze data processing system was .36s on average. Lag was determined by comparing video registrations of eye movements with the recorded gaze data superimposed. A volunteer was asked to fixate on specific points on the screen. His eye image was recorded with a timestamp. Through comparing the time on the frames of his eye image with the timestamps of gaze data, the latency was determined, and subsequently corrected to within one video frame (<.04s).

4.3.4 Measuring task performance

The number of sentences uttered by the subject was measured as task performance. Experimenters recorded the number of correct sentences produced by the subject during each trial.

Task performance was recorded by two independent investigators, and double-checked with our speech recordings.

4.3.5 Measuring conversational satisfaction

Conversational satisfaction was measured using a post-experiment questionnaire. The questionnaire consisted of eight Likert-style questions, measuring the subject's satisfaction with the task, satisfaction with his conversation partners, and satisfaction with the experiment set-up. The questionnaire used a 5-point scale, in which 1 indicated a strong disagreement and 5 indicated a strong agreement (See Appendix 6 for details).

5 Research Results

The experiment produced four categories of measures covering conversational activities, gaze, task performance, and participation satisfaction. For the first three categories of measures, the first 5 minutes of 34 sessions were analyzed. To compile our statistics we averaged all measures over the analyzed 34 sessions. All data was normally distributed (Kolmogorov-Smirnov test, $p>.05$) and of equal variances (Levene's test, $p>.05$). All planned comparisons were carried out using 1-tailed t-tests, evaluated at $\alpha = .05$. All other comparisons were carried out using two-tailed tests. Covariance analyses were carried out using both multivariate (Pillai's Trace) and univariate tests. The correlational analyses were used to analyze the linear relationship between the level of gaze and our dependent measures with the significant level of .05 or .01.

5.1 Gaze and speech

Gaze synchronization and speech

We used three measures -- the amount of speech, number of talkspurts, and number of turns -- for speech activity.

T-tests showed a significant difference in the subject amount of speech and subject number of talkspurts between two conditions. Subjects spent about 22% more time speaking to the actors in the synchronized condition (18.5%) than they did in the random condition (15.2%) ($t = -1.87$, $df = 32$, $p < .05$, 1-tailed). They used about 26% more talkspurts in the synchronized condition (27.9) than they did in the random condition (22.1) ($t = -2.52$, $df = 32$, $p < .01$, 1-tailed). (See Table 5-1)

There was no significant effect of gaze condition on the number of turns ($t = -.57$, $df = 32$, $p > .05$ 1-tailed). However, on average, actors took about 17% more turns in the random condition (7.7) than they did in the synchronized condition (6.6) ($t = 2.29$, $df = 32$, $p < .05$, 2-tailed). With respect to group number of turns, there was no significant difference between two gaze conditions. (See Table 5-1)

	Sync. condition (mean & δ)	Random condition (mean & δ)	t-test
Subject amount of speech (% of time)	18.5 (4.9)	15.2 (5.3)	p<.05.
Subject number of talkspurts	27.9 (6.9)	22.1 (6.5)	p<.05
Subject number of turns	9.35 (2.89)	8.82 (2.53)	n.s.
Actor number of turn	6.6	7.7	p<.05
Group amount of speech (%time)	44.9 (6.0)	45.6 (8.5)	n.s.
Group number of talkspurts	66.2 (8.3)	65.9 (9.1)	n.s.
Group number of turn	22.53 (6.16)	24.24 (5.30)	n.s.

Table 5-1 **Means and δ of the speech measures between two conditions**

The level of gaze and speech

The correlational analysis revealed a linear relationship between the level of gaze experienced by subjects and the amount of speech they produced($r = .62$, $p < .01$, 1-tailed), and between the level of gaze and subject number of talkspurts ($r=.54$, $p<.01$, 1-tailed). The effect of our covariate gaze level was particularly strong in the synchronized condition where it accounted for up to 49% of the variance in subject speech activity (Synchronized: $r = .70$, $p<.01$ 1-tailed; Random: $r=.53$, $p<.05$ 1-tailed). We did not find significant linear relationship between the level of gaze and group amount of speech ($r= .112$, $p> .05$, 1-tailed) or group number of talkspurts ($r=.212$, $p>.05$, 1-tailed).

We also found a significant partial correlation between the level of gaze and the subject number of turns ($r = .34$, $P<.05$, 1-tailed). As per condition, there was also significant linear relationship between the level of gaze and subject number of turns (Synchronized: $r = .393$, $P < .05$, 1-tailed; Random: $r = .216$, $p < .05$, 1-tailed). We did not find a similar relationship between the level of gaze and the group number of turns.

Covariate analysis

Table 5-2 shows the results of covariate analysis. The level of gaze had a significant effect on our conversational measures (Pillai's Trace; $F(3, 29) = 6.5$, $p<.01$). Subjects spoke significantly more often with higher levels of gaze ($F(1,31) = 19.3$, $p<.001$). Subjects also used significantly more talkspurts with higher levels of gaze ($F(1,31) = 12.9$, $p<.001$). We found a similar trend in the number of turns taken by subjects ($F(1,31) = 3.9$, $p=.05$).

With the effect of differences in levels of gaze taken out of the equation, differences between conditions in subject conversational measures were in fact not significant (Pillai's Trace, $F(3, 29) = .735$, $p>.05$; Subject amount of speech $F(1, 31) = .285$ $p>.05$; Subject number of talkspurts $F(1, 31) = .175$, $p>.05$; Subject number of turns $F(1, 31) = .392$, $p>.05$). Our test did not show a significant interaction between gaze condition and the level of gaze.

Source	Dependent variable	df	F-test
Gaze condition	Subject amount of speech	1	n.s.
	Subject number of talkspurts	1	n.s.
	Subject number of turn	1	n.s.
	Group amount of speech	1	n.s.
	Group number of talkspurts	1	n.s.
	Group number of turns	1	n.s.
Gaze level	Subject amount of speech	1	p<.01
	Subject number of talkspurts	1	p<.01
	Subject number of turns	1	p = .05
	Group amount of speech	1	n.s.
	Group number of talkspurts	1	n.s.
	Group number of turns	1	n.s.
Condition * gaze level	Subject amount of speech	1	n.s.
	Subject number of talkspurts	1	n.s.
	Subject number of turns	1	n.s.
	Group amount of speech	1	n.s.
	Group number of talkspurts	1	n.s.
	Group number of turns	1	n.s.

Table 5-2 Covariance analyses of the effects of gaze condition and gaze level

5.2 Gaze and participation of subjects

Participants were not likely to contribute speech evenly throughout our group conversation. In addition to an analysis on the amount of speech, number of talkspurts, and number of turns, we analyzed the participation of subjects. The participation of subjects was determined by the ratio of subject's contribution to the group's contribution with respect to the amount of speech, number of talkspurts, and number of turns. For example, assuming that the group number of talkspurts was 30, and the subject number of talkspurts was 12, the participation of subject would be .4 in terms of the number of talkspurts.

We had three measures of subject participation: 1) ratio of speech; 2) ratio of talkspurts; and 3) ratio of turns. The participation of subjects in two conditions is shown in Table 5-3.

	Ratio of speech (mean & δ)	**Ratio of talkspurts** (mean & δ)	**Ratio of turns** (mean & δ)
Synchronized condition	.412 (.087)	.419 (.079)	.415 (.052)
Random condition	.334 (.095)	.334 (.081)	.362 (.068)
t-test	p<.05	p<.05	P<.001

Table 5-3 Mean and δ of subject contributions in terms of ratio of speech, and ratio of talkspurts, ratio of turns

T-test showed a significant difference in subject participation across conditions. Subjects contributed higher ratio of speech in the synchronized condition (41.2%) than they did in the random condition (33.4%) ($t = 2.506$, $df = 32$, $p < .05$, 2-tailed), and higher ratio of talkspurts in the synchronized condition (41.9%) than in the random condition (33.4%) ($t = 3.086$, $df = 32$, $p < .01$, 2-tailed). We found a similar trend in the ratio of turns, 41.5% in the synchronized condition vs. 36.2% in the random condition ($t = 2.502$, $df = 32$, $p < .05$, 2-tailed).

However, variance in subject participation could be attributed to gaze synchronization as well as the level of gaze. Statistics demonstrated a significant relationship between the level of gaze and the subject's ratio of contribution (Ratio of speech: $r = .56$, $p < .01$, 2-tailed; Ratio of talkspurt: $r = .56$, $p < .01$, 2-tailed; Ratio of turns: $r = .50$, $p < .01$, 2-tailed;). A covariate analysis could be appropriately applied to reveal the effect of gaze synchronization and the level of gaze.

A covariate analysis demonstrated that the effect of our covariate on the ratio of subject's conversational contributions was highly significant (Pillai's Trace, $F(3,29) = 5.361$, $p < .01$). Subjects contributed more ratio of speech ($F(1,31) = 13.26$, $p < .01$), talkspurts ($F(1,31) = 13.25$, $p < .01$), and turns ($F(1,31) = 10.18$, $p < .01$) with higher levels of gaze. With the effect of differences in levels of gaze taken out of the equation, differences between conditions in the ratio of subject's contribution to group contribution were in fact not significant (Pillai's Trace, $F(3, 29) = .683$, $p > .05$; Ratio of speech: $F(1, 31) = .207$ $p > .05$; Ratio of talkspurts: $F(1, 31) = .206$, $p > .05$; Ratio of turns: $F(1, 31) = .386$, $p > .05$).

5.3 Gaze and being gazed at

5.3.1 Gaze, subject gaze and mutual gaze

Effect of gaze synchronization

We conducted a t-test to analyze the effect of gaze conditions on subject gaze and mutual gaze and did not find the effect of gaze condition on the level of gaze by subject ($t = -.851$, $df = 32$, $p > .05$, 2-tailed), and the amount of mutual gaze ($t = .512$, $df = 32$, $p > .05$, 2-tailed). (See Table 5-4).

	Synchronized Condition (%)	Random Condition (%)	T-test
The level of subject gaze	19.3 (7.4)	18.2 (8)	n.s.
The amount of mutual gaze	6.9 (2.7)	7.2 (3.1)	n.s.

Table 5-4 Mean and δ of the level of gaze produced by subjects, and the amount of mutual gaze in two gaze conditions

Effect of the level of gaze

The statistics also did not reveal a significant linear relationship between the level of gaze experienced by subjects and the level of gaze produced by subjects ($r = -.222$, $p > .05$, 2-

tailed) and amount of mutual gaze ($r = .145$, $p > .05$, 2-tailed). As per condition, a similar pattern was found. The level of gaze experienced by subjects did not have effect on either the level of subject gaze (Synchronized: $r = -.093$, $p > .05$, 2-tailed; Random: $r = -.293$, $p > .05$, 2-tailed) or the amount of mutual gaze (Synchronized: $r = .149$, $p > .05$, 2-tailed; Random $r = .054$, $p > .05$, 2-tailed). (See Table 5-5)

	Overall		Sync.		Random	
	r	sign.	r	sign.	r	sign.
The level of subject gaze	-.222	n.s.	-.093	n.s.	-.293	n.s.
The amount of mutual gaze	.145	n.s.	.149	n.s.	.054	n.s.

Table 5-5 Relationship between the level of gaze experienced by subjects and subject gaze and mutual gaze

5.3.2 Sex differences in gaze behavior

We did not find a significant difference in the amount of subject gaze and mutual gaze with actors between male and female groups (Gaze: $t = -.528$, $df = 32$, $p > .05$, 2-tailed; Mutual gaze: $t = -.356$, $df = 32$, $p > .05$, 2-tailed). (See **Table 5-6**)

	Male subjects (%)	Female subjects	T-test
Gaze at actors	18.66 (1.64)	20 (1.99)	n.s.
Mutual gaze with actors	6.77 (2.58)	7.11 (2.99)	n.s.

Table 5-6 Mean and δ of gaze by male and female subjects

In general, there was also no significant difference between the levels of subject gaze at an actor vs. the level of subject gaze at actress ($t = -1.74$, $df = 33$, $P > .05$, paired, 2-tailed,). However, as per sex, we did find that male subjects gazed at the actress significantly more than at the actor ($t = -2.502$, $df = 17$, $P < .05$, paired, 2-tailed). Also, male subjects had more mutual gaze with the actress than with the actor ($t = -2.078$, $df = 17$, $P < .05$, paired, 2-tailed).

The female subjects did not demonstrate a similar gaze pattern. We found no significant difference between the level of subject gaze at an actor and at an actress in the female group ($t = .213$, $df = 15$, $P > .05$, paired, 2-tailed). With regard to mutual gaze with the actor and actress, the female group also did not show a significant difference on these two measures ($t = .184$, $df = 15$, $p > .05$, paired, 2-tailed).

5.4 Gaze and task performance

The number of correct permutations generated by the subject was used as the measure of task performance. Task performance in the two gaze conditions is shown in Table 5-7. A t-test demonstrated a significant difference in task performance between the two conditions ($t = 3.195$, $df = 32$, $p < .01$, 2-tailed).

	Sync. condition	Random condition	t-test
Task performance	18.88 (6.13)	12.88 (4.73)	P<.01

Table 5-7 Mean and δ of task performance in two gaze conditions

However, correlational analysis did not find a significant linear relationship between gaze level and task performance ($r = .25$, $P = .082$, 2-tailed) across conditions. Likewise, we did not find a significant linear relationship between the level of gaze and the task performance per condition (synchronized $r = .171$, $p > .05$, 2-tailed; random $r = .382$, $p > .05$, 2-tailed). This disqualified task performance for covariance analysis.

5.5 Gaze and conversational satisfaction

The conversational satisfaction was measured by post-experiment questionnaire. Each participant was asked to answer questions in three aspects: evaluation on conversational partner, experimental setup, and experimental task. We merged the raw scores into two categories which are positive evaluation with score larger than or equal to 3 and negative evaluation with the score less than 2. The distribution of the number of positive and negative evaluation on Q1 to Q8 in two conditions is shown in Table 5-8. We used Kruskal-Wallis non-parametric test to analyze the variance on the evaluation on Q1 to Q8 in two conditions and did not find a significant difference in the answers to any of 8 questions. The detailed statistics will be discussed in the following sections.

	Synchronized condition		Random condition		χ^2 Test
	Positive	Negative	Positive	Negative	
Q1	4	13	5	12	n.s.
Q2	1	16	0	17	n.s.
Q3	17	0	16	1	n.s.
Q4	3	14	2	15	n.s.
Q5	16	1	17	0	n.s.
Q6	1	16	0	17	n.s.
Q7	16	1	14	3	n.s.
Q8	15	2	16	1	n.s.

Table 5-8 The distribution of positive and negative evaluations on Q1 to Q8 in two conditions

Q1: Hard to do task with partners

The distribution of evaluation on Q1 in two gaze conditions is demonstrated in Figure 5-1.There was no significant difference between two conditions (χ^2 = .147, df = 1, p > .05). (In Figure 5-1, "s" means synchronized condition, "r" means random condition, same for Q2 – Q8).

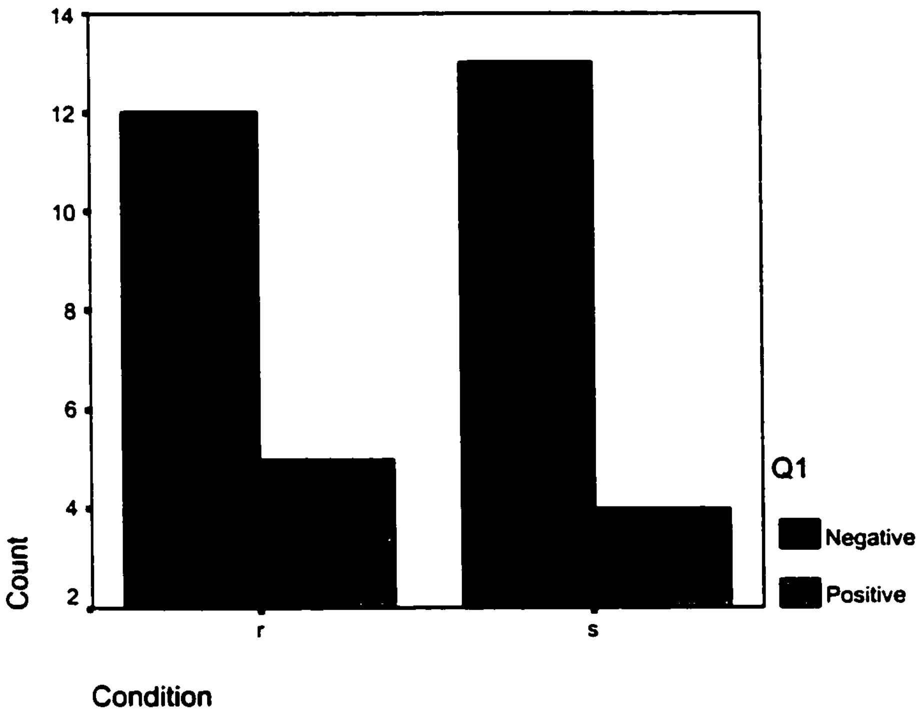


Figure 5-1 The distribution of positive and negative responses on Q1 in two conditions

Q2: Have worked with these people before

The distribution of positive and negative evaluation on Q2 in two gaze conditions is demonstrated in Figure 5-2. There was no significant difference between two conditions ($\chi^2 = 1$, df = 1, $p > .05$).

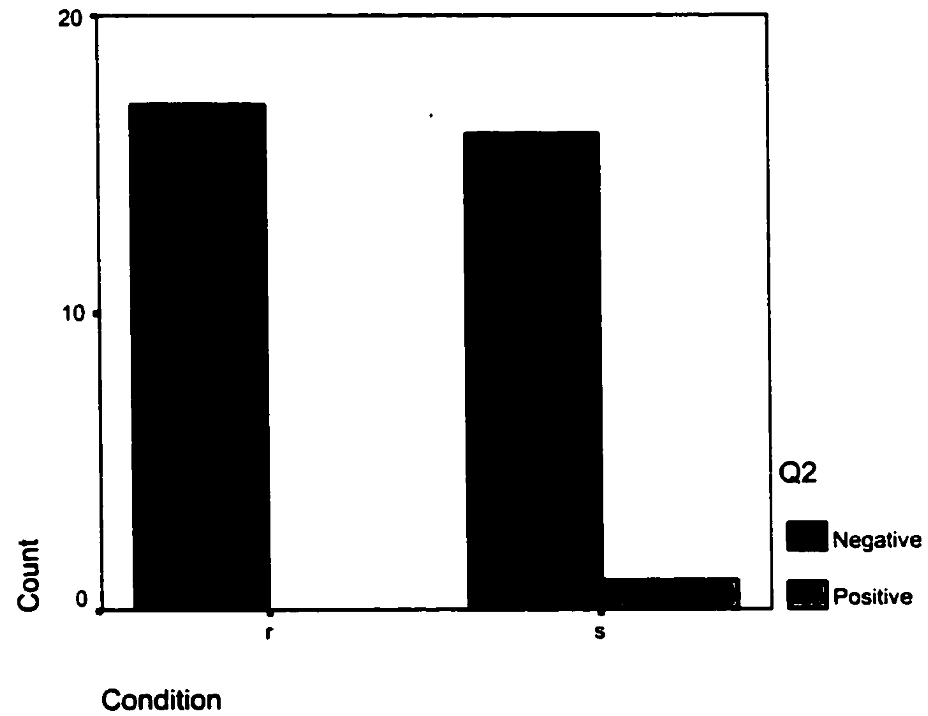


Figure 5-2 The distribution of positive and negative responses on Q2 in two conditions

Q3: *Enjoyed collaborating with the partners*

The distribution of evaluation on Q3 in two gaze conditions is shown in Figure 5-3. There was no significant difference between two conditions ($\chi^2 = 1$, df = 1, $p > .05$).

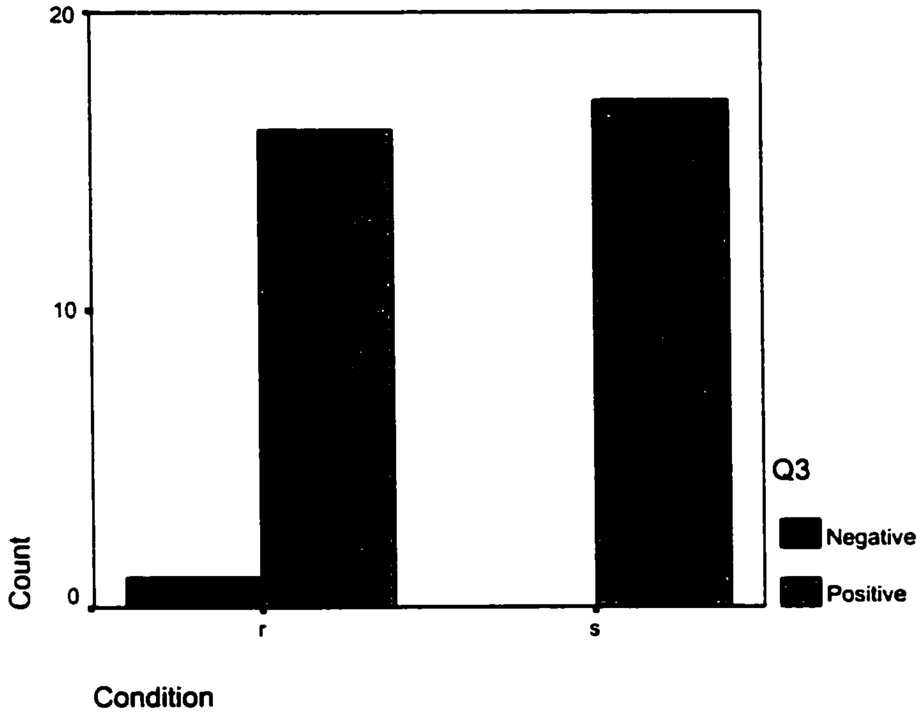


Figure 5-3 The distribution of positive and negative responses on Q3 in two conditions

Q4: *Hard to take speaking turn*

The distribution of evaluation on Q4 in two gaze conditions is demonstrated in Figure 5-4.There was no significant difference between two conditions (χ^2 = .228, df = 1, p > .05).

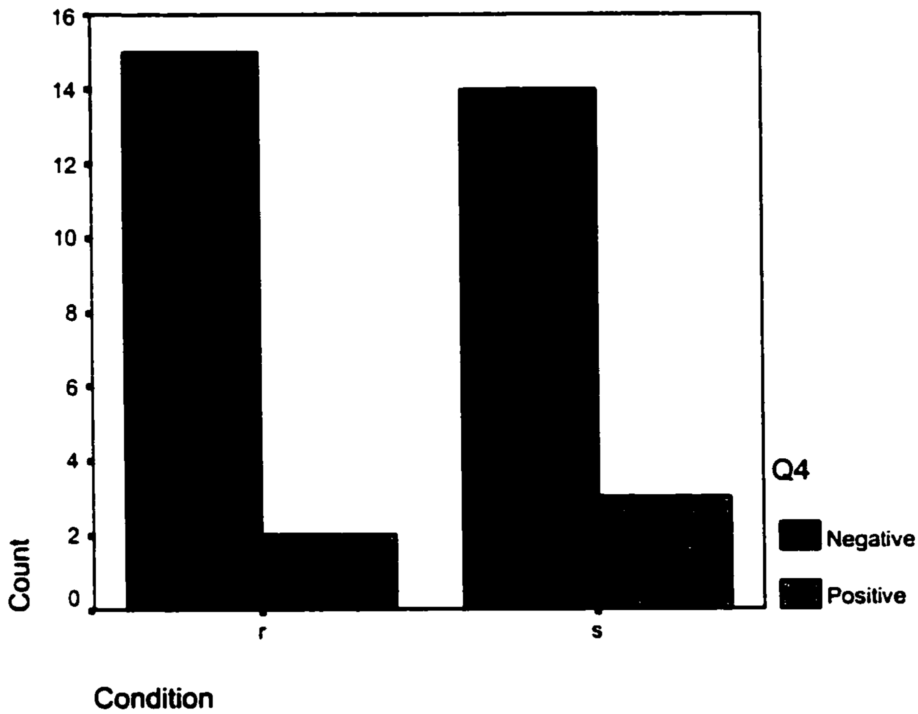


Figure 5-4 The distribution of positive and negative responses on Q4 in two conditions

Q5: *Being paid attention by others*

The distribution of evaluation on Q5 in two gaze conditions is demonstrated in Figure 5-5.There was no significant difference between two conditions ($\chi^2 = 1$, $df = 1$, $p > .05$).

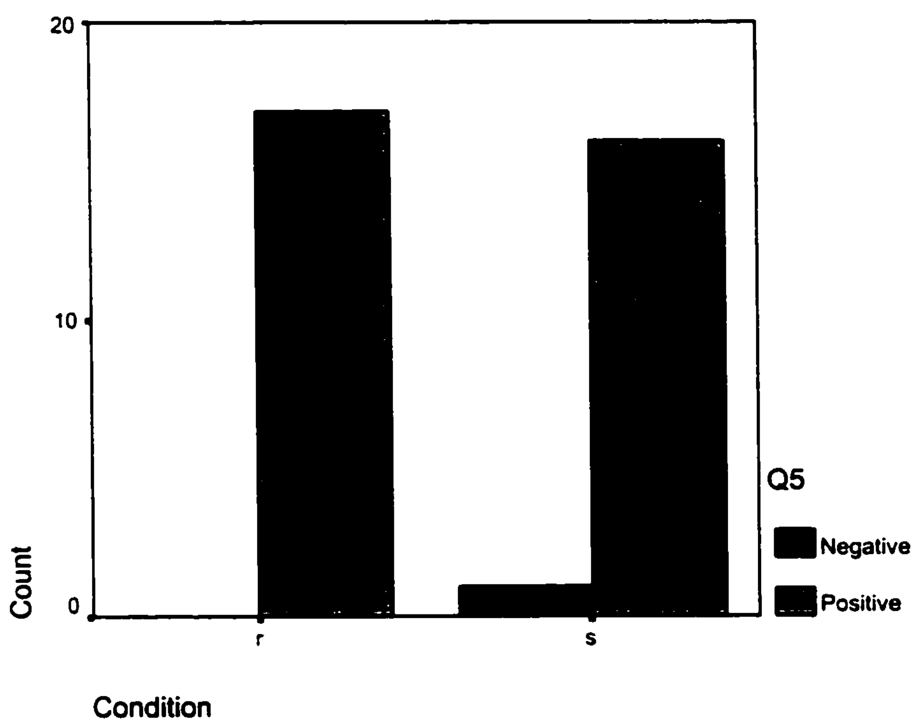


Figure 5-5 The distribution of positive and negative responses on Q5 in two conditions

Q6: *Sometimes hard to identify who is speaking*

The distribution of evaluation on Q6 in two gaze conditions is demonstrated in Figure 5-6.There was no significant difference between two conditions ($\chi^2 = 1$, df = 1, $p > .05$).

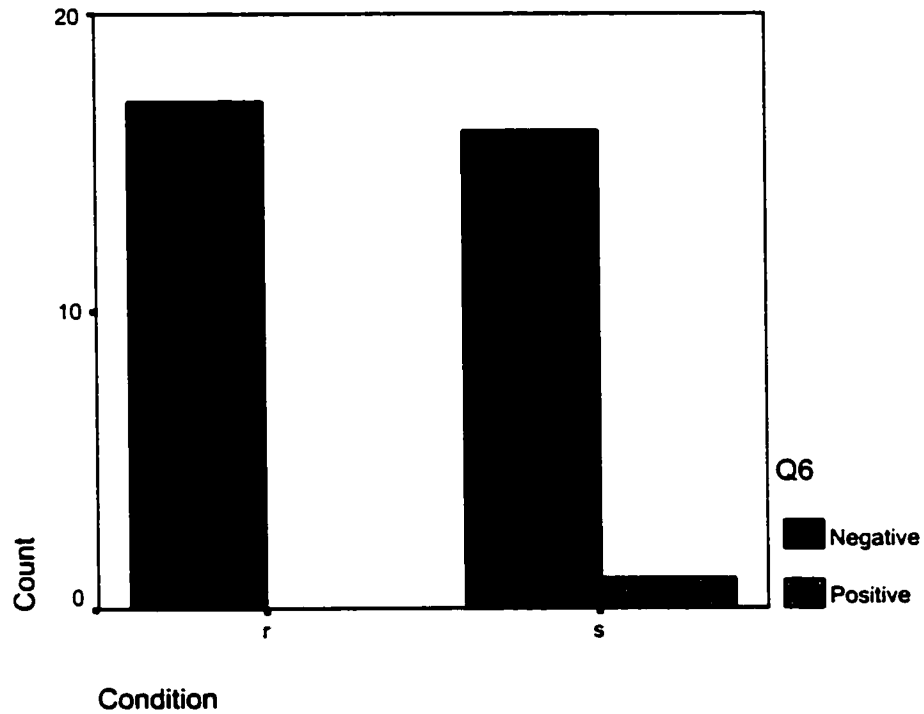


Figure 5-6 The distribution of positive and negative responses on Q6 in two conditions

Q7: *The system is easy to work with*

The distribution of evaluation on Q7 in two gaze conditions is demonstrated in Figure 5-7. There was no significant difference between two conditions ($\chi^2 = 1.1$, $df = 1$, $p > .05$).

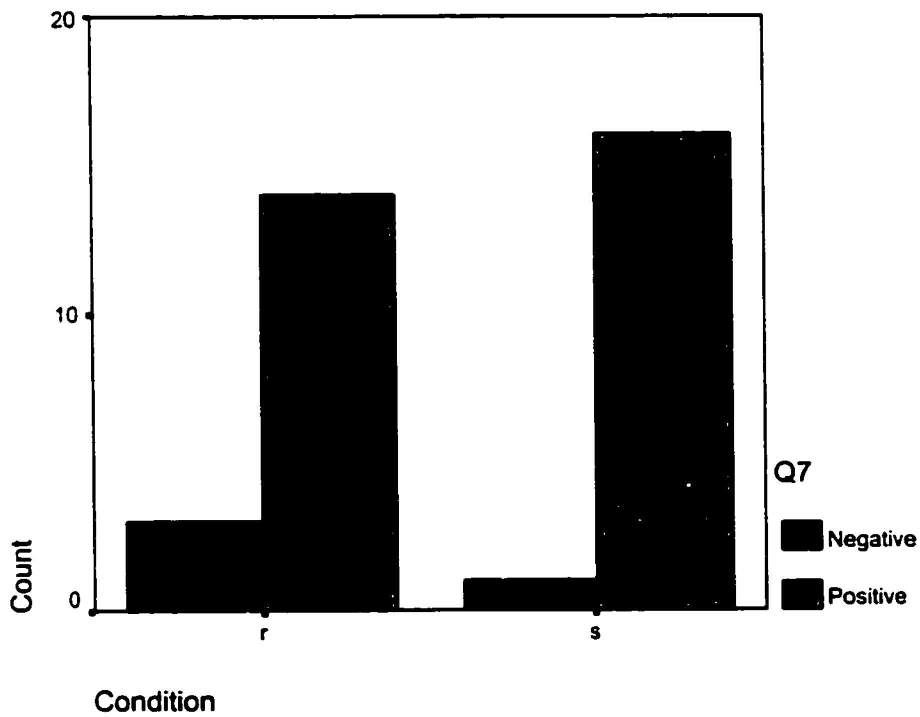


Figure 5-7 The distribution of positive and negative responses on Q7 in two conditions

Q8: *Easy to identify to whom the others were talking*

The distribution of evaluation on Q8 in two gaze conditions is demonstrated in Figure 5-8. There was no significant difference between two conditions ($\chi^2 = .355$, df = 1, $p > .05$).

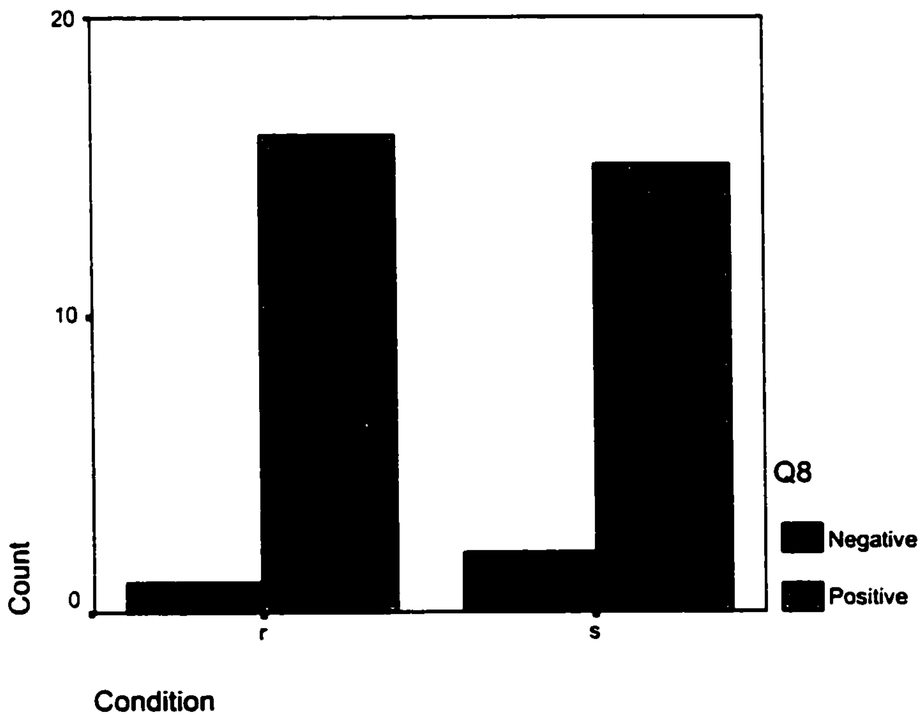


Figure 5-8 The distribution of positive and negative responses on Q8 in two conditions

6 Discussion

In chapter 1, we proposed six hypotheses. We will now discuss our results with regard to each hypothesis.

6.1 Validation of the hypotheses

Hypothesis 1: People are more likely to speak when gaze conveys conversational attention.

Our preliminary results would seem to provide clear support for this hypothesis, stating that people are significantly more likely to speak when gaze behavior of conversational partners is synchronized through time with their conversational attention. Subjects produced 22% more speech in the synchronized condition, using 26% more talkspurts. Although differences between conditions in the number of subject turns were not significant, actors took 17% percent more turns in the random condition. We believe this was to compensate for what they perceived as a lack of subject-initiated turns. However, such results are clearly confounded by our covariate. In the synchronized condition, about 50% of the variance in speech behavior was accounted for by the variance in levels of gaze. This cannot be explained by the 46% difference in task performance between conditions. Similarly, in the random condition about 28% of the variance in speech activity was accounted for by the variance in levels of gaze. Although the variation in mean gaze level between conditions were kept to within 12% of each other, it appears likely that the observed effects were in fact caused by this difference. Our covariance analysis provides clear evidence for this line of thought, showing a significant effect of level of gaze on our conversational variables that was large enough to render effects of condition insignificant. This means that we do not in fact have enough evidence to support hypothesis 1. However we note that although our tests did not show any significant interaction between level of gaze and condition, it does appear that effects of gaze on speech activity were stronger when conversational attention was indeed conveyed. As expected, we did not find a significant partial correlation across conditions nor a significant correlation per condition between the levels of gaze and task performance. This disqualifies task performance for covariance analysis. It therefore seems likely that there was indeed an effect of synchronization on task performance. Although we did not observe the gaze-avoidant behavior that would support such explanation, subjects may have found it harder to focus on the task in the random condition. It appears, however, that we should not necessarily equate the quality of the conversation – in terms of the number of correct answers submitted – with the quantity of conversation. We therefore conclude that although the observed positive effects of condition on task performance may explain part of the variance in conversational behavior, effects of the level of gaze were sufficiently large to account for all differences between conditions.

Hypothesis 2: People are more likely to speak when they are being gazed at more

With regards to our second hypothesis, we verified whether people are significantly more likely to speak when their conversational partners gaze more. Results of our covariance analysis provide clear support for this hypothesis. We found a surprisingly high partial correlation across conditions of r=.62 between level of gaze and the amount of speech activity. We found significant correlations between level of gaze and the number of talkspurts (r=.54) and turns (r=.34). The observed effect of gaze on the number of turns is line with Vertegaal's findings (2000). In addition, our correlational studies provide little support for the hypothesis that gaze affected speech activity through increased task performance, leading us to accept this hypothesis.

Hypothesis 3: People are more likely to participate in the conversation when gaze conveys conversational attention.

Our results would seem to provide support for this hypothesis. We found a significant difference in ratio of subject's contribution to group contribution between two conditions. Subjects in the synchronized condition were more likely to participate in the conversation than those in the random condition with regard to the ratio of speech (41.2% vs. 33.4%), talkspurts (41.9% vs. 33.4%), and turns (41.5% vs. 36.2%).

However, such results are clearly confounded by our covariate, since we found a significant linear relationship between the level of gaze and ratio of subject's contribution (r= .56, .56, and .50, respectively for the ratio of speech, talkspurts, and turns). Our covariance analysis also showed a significant effect of the level of gaze on the subject participation measures, and did not find the effect of gaze condition on the subject participation measures with removal of the effect of differences in the levels of gaze. Therefore, we do not have enough evidence to support hypothesis 3. As such, we attributed the variance in subject participation to the change of the level of gaze.

Hypothesis 4: People are more likely to gaze at their conversational partner when gaze conveys conversational attention.

People in the synchronized condition spent 18.2% of time gazing at their conversational partners with a mutual gaze of 7.2%. In the random condition, subjects spent slightly more time (20.4%) gazing at their partners, with a mutual gaze of 6.7%. We did not find an effect of gaze condition on the level of subject gaze ($t = -.851$, $p < .05$) and mutual gaze ($t = .512$, $p > .05$), therefore this hypothesis was not validated.

According to Swain et al (1982) and Rutter & Stephenson (1979), if there were a regulatory function of gaze, people would gaze less in the synchronized condition since they did not need to gaze at their partners frequently to seek feedback or gather information about conversational attention. In this condition, gaze experienced by subjects was synchronized with the conversational attention of their partners. If presentation of this kind of gaze regulated the conversation, subjects might gaze less at their partners throughout conversation. Our results did not provide enough evidence to support this regulatory function of gaze.

Hypothesis 5: People are less likely to gaze at their conversational partners when they experience more gaze.

Although our results demonstrated a considerable negative correlation across conditions ($r = -.222$) and in the random condition ($r = -.293$) between the levels of gaze experienced by subjects and the level of subject gaze, there were no significant linear relationships between the level of gaze and subject gaze and mutual gaze across condition and per condition (See Table 5-5). Therefore, our results do not support this hypothesis.

Hypothesis 6: People feel more satisfied with conversation and conversational partners when gaze convey conversational attention.

Our results did not validate this hypothesis. χ^2 tests showed that subjects in two conditions did not differ significantly in the scores on eight questions measuring conversational satisfaction (See Table 5-8).

We attributed a consistency of satisfaction evaluation in two conditions to the expectation of subjects. Most subjects did not have prior experience in using video-conferencing systems. Throughout our experimental procedure, subjects were repeatedly told that it was a video-conferencing experiment. When they used the system, they might be attracted by the novelty of systems and the quality of audio and video communication. Some subjects mentioned the unreal gaze pattern in the random condition, but they considered it as latency in video signal transmission. The quality of communication through this video-conferencing system exceeded their expectation. Hence they gave high scores to the evaluation items.

6.2 Sex differences in gaze behavior

Although the females spent slightly more amount of time gazing at their partners (20%) than the males did (18.7%), and had slightly more mutual gaze with their partners (7.1%) than the males did (6.8%), there were no significant differences in the amount of subject gaze and mutual gaze between the male and female groups. The results were not fully consistent with earlier findings (Kendon & Cook, 1969; Argyle & Williams, 1969; Argyle & Ingham, 1972; Nevill, 1974; Coutts & Schneider, 1975; Russo, 1975) that women tend to look more than men.

Surprisingly, we found another sex difference in the gaze pattern. Our results showed that male subjects spent significant more time gazing at the female (21.6%) than at the male (15.7%), and had more mutual gaze with the female (7.88%) than with the male (5.67%). Female subjects did not show similar pattern (Gaze: 20.2% at the male vs. 19.8% at the female; Mutual gaze: 7.21% with the male vs. 7.0% with the female). Both men and women usually pay more visual attention to the opposite sex than to the same sex. However, women are often characterized as those who are more dependent on social evaluation. When they recognized that they were observed, they might suppress gaze at their male partners in order to keep a socially acceptable pattern.

7 Conclusions

7.1 Conclusions

In this research, we presented an experiment examining effects of gaze – looking at a person's eyes and face – on speech activity during triadic conversations. Our experiment aimed to explain the relationship between the level of gaze and speech, and if subjects in group conversations take more speaking turns and speak more with more gaze. Would more gazes allow subjects to better observe whether their partner was speaking or listening to them, or did it make them feel more comfortable and therefore more inclined to take the floor? To evaluate this, we designed two gaze conditions: 1) in which subjects experienced gaze synchronized with the conversational attention of their partners, and 2) in which gaze was randomly presented. The level of gaze experienced by subjects was treated as a covariate and was controlled. The amount of speech, number of talkspurts, and number of turn were measures of speech activities. Task performance, gaze behavior and conversational satisfaction were also measured as additional dependent variables. Furthermore, subject participation was determined as the ratio of subject contribution to group contribution in speech.

We used speech, talkspurts, turns, and subject participation to measure the speech activities, and found a considerable correlation of .62 between the level of gaze and amount of subject speech ($r = .62$), number of talkspurts ($r = .54$), number of turns ($r = .34$), and subject participation ($r = .56$). Although the t-test showed an effect of gaze condition on speech activity, the covariant analysis demonstrated that gaze condition did not have significant effect on our speech measures after a removal of the effect of gaze level. The level of gaze had a significant effect on our speech measures. The effect of the level of gaze was large enough to render the effect of gaze condition insignificant. The result validated the emotional function of gaze as a favorable explanation of gaze effect on human interaction in multiparty conversation. More gaze would allow subjects to feel more comfortable to speak in our triadic conversation.

We found a significant effect of gaze synchronization on task performance. Subjects in the synchronized condition outperformed those in the random condition in terms of the number of sentences produced. However, we did not find a significant partial correlation across conditions or a significant correlation per condition between the levels of gaze and task performance. This provides little support for the hypothesis that the level of gaze affected speech activity through increased task performance. The possible causal relationship among gaze, speech and task performance would be that the level of gaze affected speech, and speech affected task performance.

We did not find an effect of gaze condition on the subject's gaze behavior. There was no significant difference in the amount of subject gaze and mutual gaze between the two conditions. We also did not find sex differences in both the level of subject gaze and amount of mutual gaze. However, as per sex, we found that male subjects gazed more at their female partner than at their male partner. Female subjects did not show this pattern. Women are often characterized as those who are more dependent on social evaluations.

When they recognized that they were observed, they might suppress gaze at their male partner in order to keep a socially acceptable level of gaze.

We did not find a significant difference in conversational satisfaction between the two conditions. Subjects in both conditions were satisfied with the experimental setup, task, and their conversational partners.

7.2 Problems and future work

Due to the dynamic nature of the gaze model, the level of gaze in each experimental session is impossible to be controlled accurately. The level of gaze experienced by subjects changed throughout the experiment. In our design, the synchronization of gaze was a factor and the amount of gaze a covariate. It is efficient to investigate the effects of both synchronization and the level of gaze in one experiment. However, the design limited the capability of investigating the independent effects of synchronization and the level of gaze separately. In addition, how to generalize the finding from triadic conversation to a larger group conversation is an open issue.

Developing precise gaze model

Because of its controllability and manipulability, gaze models can be continuously used to simulate the gaze behavior of participants in conversation research. Our gaze model takes speech signals and conversational attention scores as input. The precision is inevitably affected by the accuracy and latency of the actor's score. A gaze model that can directly take gaze data and speech signals from each participant would be an alternative source of input.

Investigating the independent effects of synchronization

When we separate the factor of gaze level from the factor of synchronization by controlling the variance of the level of gaze across conditions, we are able to investigate the independent effect of synchronization. A *Yoked Control Experiment* design can be used for this purpose. In a Yoked design, two groups are matched in all conversationally sensitive criteria, such as introversion/extraversion score and language skill. Two groups are exposed to the synchronized and random condition respectively. The difference from the current design is that, for each trial in the synchronized condition, there should be a Yoked trial in the random condition, in which the subject experiences approximately the same level of gaze as his counterpart did in the synchronized condition, except that the gaze is randomly given rather than synchronized with the conversational attention. For example, if a subject in the synchronized condition experienced 30% of gaze during the conversation, his counterpart in the random treatment will be exposed to about the same level of gaze by adjusting the probability matrix of the gaze model. Through a Yoked control, the level of gaze subject experienced is about the same in the two conditions.

Increasing the size of group conversation

Our research investigated the gaze function in triadic conversation and produced several valuable findings. We must verify our results with groups larger than three. It will be interesting to generalize our findings to groups with different numbers of participants.

Replicating the experiment with different tasks

We used an auditory task in the current experiment. The task facilitated conversation flow, and probably influenced speech activities as well. We should verify the results with other type of tasks, such as negotiation and problem solving tasks.

Bibliography

1. Argyle, M. & Williams, M. (1969) Observer or observed? A reversible perspective in person perception. Sociometry, 32, 396-412.

2. Argyle, M. & Ingham, R. (1972) Gaze, mutual gaze, and proximity. Semiotica, 6, 32-49.

3. Argyle, M., Ingham, R., Alkema, F., & McCallin, M. (1973) The different functions of gaze. Semiotica, 7, 19-32.

4. Argyle, M., Lefebvre L. & Cook, M. (1974) The meaning of five patterns of gaze. European Journal of Social Psychology, 4(2), 125-36.

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

6. Beattie, G. W. (1979) Contextual constraints on the floor-apportionment function of speaker-gaze in dyadic conversations. British Journal of Social and Clinical Psychology, 18, 391-92.

7. Campbell, N. (1993) Biology 3rd Ed. Redwood City, CA: Benjamin/Cummings Publishing. P. 1022.

8. Carr, S. J. & Dabbs, J. M. (1974) The effects of lighting, distance and intimacy of topic on verbal and visual behavior. Sociometry, 37, 592-600.

9. Cary, M. S. (1978) The role of gaze in the initiation of conversation. Social Psychology, 41, 269-71.

10. Colburn, R. A., Cohen, M. F. & Drucker, S. M. (2000) The role of eye gaze in avatar mediated conversational interfaces. Technical Report, MSR-TR-2000-81. Microsoft Research.

11. Coutts, L. M. & Schneider, F. W. (1975) Visual behavior in an unfocused interaction as a function of sex and distance. Journal of Experimental Psychology, 11, 64-77.

12. Deaux, K & Wrightsman, L. S. (1983) Social Psychology in the 80s, 4th Edition, Brooks/Cole Publishing Company.

13. Duncan, S. (1972) Some signals and rules for taking speaking turns in conversations. Journal of Personality and Social Psychology, 23, 283-93.

14. Dovidio, J. F. & Ellyson, S. L. (1982) Decoding visual dominance: attributions of power based on relative percentages of looking while speaking and looking while listening. Social Psychology Quarterly, 45(2), 106-113.

15. Ellsworth, P.C. & Carlsmith, J.M (1973) Eye contact and gaze aversion in an aggressive encounter. Journal of Personality and Social Psychology, 28, 280-92.

16. Ellsworth, P.C., Carlsmith, J.M & Henson, A. (1972) The stare as a stimulus to flight in human subjects: a series of field experiments. Journal of Personality and Social Psychology, 21, 302-11.

17. Exline, R. V. (1963) Explorations in the process of person perception: visual interaction in relation to competition, sex, and need for affiliation. Journal of Personality, 31, 1-20.

18. Exline, R. V., Gray, D. & Schuette, D. (1965) Visual behavior in a dyad as affected by interview content as sex of respondent. Journal of Personality and Social Psychology, 1, 201-9.

19. Exline, R.V. (1971) Visual interaction: the glances of power and preference. Nebraska Symposium on Motivation, 163-206.

20. Fromme, D. K. & Schmidt, C. K. (1972) Affective role enactment and expressive behavior. Journal of Personality and Social Psychology, 29, 586-92.

21. Garau, M., Slater, M., Bee, S., & Sasse, M.A. (2001) The impact of eye gaze on communication using humanoid avatars. In Proceedings of CHI'01. Seattle, WA: ACM, 2001.

22. Jaffe, J. & Feldstein, S. (1970) Rhythms of Dialogue. New York, NY USA: Academic Press.

23. Kalma, A. (1992) Gazing in triads: a powerful signal in floor apportionment. British Journal of Social Psychology, 31, 21-39.

24. Kendon, A. (1967) Some Function of Gaze Direction in Social Interaction. Acta Psychologica, 32, 1-25.

25. Kendon, A. & Cook, M. (1969) The consistency of gaze patterns in social interaction. British Journal of Psychology, 69, 481-94.

26. Kleinke, C. L., Bustos, A. A., Meeker, F. B. & Staneski, R. A. (1973) Effects of self-attributed and other-attributed gaze on interpersonal evaluations between males and females. Journal of Experimental Social Psychology, 9, 154-163.

27. Lefebvre, L. M. (1975) Encoding and decoding of ingratiation in modes of smiling and gaze. British Journal of Social and Clinical Psychology, 14, 33-42.

28. Nevill, D. (1974) Experimental manipulation of dependency motivation and its effects on eye contact and measures of field dependency. Journal of Personality and Social Psychology, 29, 72-9.

29. Nielsen, G. (1962) Studies in Self Confrontation. Copenhagen: Monksgaard.

30. Novick, D., Hansen, B., & Ward, K. (1996) Coordinating turn-taking with gaze, Proceedings of ICSLP-96, Philadelphia, PA, October, 1996, 3, 1888-91.

31. Ross, Abraham S., Grant M. (1994) Experimental and Non-experimental Designs in Social Psychology. Wm. C. Brown Communications, Inc.

32. Rubin, A. (1970) Measurement of romantic love. Journal of Personality and Social Psychology, 16, 265-73.

33. Russo, N. F. (1975) Eye contact, interpersonal distance, and the equilibrium theory. Journal of Personality and Social Psychology, 31, 497-502.

34. Rutter, D.R. (1972) Visual Interaction in a Group of Introverts and Extroverts. European J. of Social Psychology, 2(4), 371-84.

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

36. Sellen, A. J. (1992) Speech patterns in video-mediated conversations. Proceedings of ACM CHI'92 Conference on Human factors in Computing Systems. Monterey, CA USA: ACM, 1992, 49-59.

37. Snyder, M., Grether, J. & Keller, C. (1974) Staring and compliance: a field experiment on hitchhiking. Journal of Applied Social Psychology, 4, 165-70.

38. Straus, Susan G., (1997) Technology, group process, and group outcomes: testing the connections in computer-mediated and face-to-face groups. Human-Computer Interaction, 12, 227-66.

39. Strongman, K. T. & Champness, B.G. (1968) Dominance hierarchies and conflict in eye contact. Acta Psychologica, 28, 376-86.

40. Swain, J., Stephenson, G. M. & Dewey, M.E. (1982) Seeing a stranger: Does eye-contact reflect intimacy? Semiotica, 42, 107-118.

41. Trager, G.L. & Smith Jr., H.L. (1957) An outline of English structure. Washington, DC USA: America Council of Learned Societies.

42. Vertegaal, R., (1997) Conversational awareness in multiparty VMC. Extended Abstracts of ACM CHI'97 Conference on Human factors in Computing Systems. Atlanta, GA USA: ACM, 6-7.

43. Vertegaal, R., (1998a) GAZE: Visual-spatial attention in communication. Video Contribution. Proceedings of ACM CSCW'98 Conference on Computer-Supported Cooperative Work. Seattle, WA USA, ACM

44. Vertegaal, R., (1998b) Look who's talking to whom: mediating joint attention multiparty communication & collaboration. University of Twente.

45. Vertegaal, R. (1999) The gaze groupware system: mediating joint attention in multiparty communication and collaboration. In Proceedings of CHI'99. Pittsburgh, PA: ACM.

46. Vertegaal, R. & et al. (2000a) Why conversational agents should catch the eye. In Extended Abstracts of CHI'00. The Hague, Netherlands: ACM, 2000.

47. Vertegaal, R., Van der Veer, G.C. & Vons, H. (2000b) Effects of gaze on multiparty mediated communication. In Proceedings of Graphics Interface 2000. Montreal, Canada.

48. Vertegaal, R. & et al. (2001) Eye gaze patterns in conversations: there is more to conversational agents than meets the eyes. In Proceedings of CHI'01. Seattle, WA: ACM.

Appendix

Appendix 1 Glossary of terms

Avatar	An icon or representation of a user in a shared virtual reality.
Conversational attention	Speaking or listening to one or more persons.
Conversational state	A description of someone's listening or speaking activities at any moment in time.
Cue	An act that conveys information from one human to another.
Double-blind	Neither the actors nor the subject know which of the two possible conditions they are treated.
Dyadic	Involving 2 persons.
Eye contact	Occurs when two person gaze at each other's eye region. Virtually synonymous with mutual gaze.
Eye tracker	A device for measuring the orientation of pupil(s), often relative to a visual scene.
Gaze	The act of looking at other human. Since the facial region is typically the focal point in this act, the term is virtually synonymous with gaze at the facial region.
Gaze direction	Absolute direction of looking, as constituted by the summation of body, head and eye orientation.
Gaze state	A description of someone's gaze direction at any moment in time.
Mediated conversation	Conversations carried out using email, mailing lists, news groups, bulletin board systems, video conferencing tools, structured conversation systems, etc.
Mutual gaze	Occurs when two persons gaze at each other. Virtually synonymous with eye contact.
Multiparty	Involving more than two persons.
Nonverbal cues	All cues expressed without words, including paralinguistic expression.
Parallax problem	A problem of inappropriate gaze communication caused by the parallax position of monitor and video camera in video conferencing tools.
Speaker switch	The act of exchanging the role of speaker and listener. Occurs when a new speaker has a talkspurt of one or more phonemic clauses, with others (including the previous speaker) being silent for at least one phonemic clause.
Talkspurt	A series of phonemic clauses by same speaker.
Triadic	Involving 3 persons.

Turn	A series of talkspurts, bounded by a speaker switch, including the silence that may occur before the speaker switch.
Turntaking	The continuous process of exchanging the role of speaker and listener. This allows, as a rule, only a single person to speaker at any moment in time.
Utterance	A series of talkspurts, bounded by a speaker switch, not including the silence that may occur before the speaker switch.
Verbal cues	All cues expressed with words, with the exception of paralinguistic expression.
Video mediated system	A system in which motion images, typically of humans, are conveyed, typically with sound, in order to communicate or collaborate across large distance.

Appendix 2 Registration Form

1. Can you attend the introduction meeting Tue 7th August 10:00 - 10:30? Yes / No

2. Can you attend the introduction meeting Tue 7th August 11:00 - 11:30? Yes / No
   Note: The introduction meeting will be in Walter Light Hall Room 210.

3. Do you have any difficulty reading from a computer screen? Yes / No

4. Do you require glasses or contact lenses? Yes / No
   If so,
   Can you see better at long or short distance? long/short
   Do you wear hard or soft lenses? hard/soft

5. How much experience do you have using videoconferencing tools?
   much/some/little/none

6. Are you native English speaker? .... Yes/No

7. Contact information
   Last Name: ________________ First Name :________________________

   Address:____

   City:____

   Phone number: ___________________ Email: _____________________

8. Are you a student? Yes / No
   Major: ________________________

9. Sex: Male / Female

10. Age: ______

Appendix 3 Personality Test

DIRECTIONS

Person to be rated: **Yourseif**

The enclosed list contains personality traits. You are asked to fill in behind each trait the amount to which this trait is applicable to yourself. Choices are:

1	=	not at all applicable
2	=	little applicable
3	=	moderately applicable
4	=	largely applicable
5	=	entirely applicable

For example:

Loves good food ___

When you find the trait "Loves good food" entirely applicable to the above mentioned person, then fill in the number 5 behind that trait.

Loves good food 5

Please be frank. It is essential for good results. Your answer will be treated confidentially. You will not be judged on the basis of your answer.

Don't take too much time to think about your answer. If you are in doubt, please compare the person to be rated with others you know well. Please don't skip traits.

1	2	3	4	5
not at all applicable				entirely applicable

Takes time out ____
Gets angry easily ____
Acts without planning ____
Feels desperate ____
Does as he/she told ____

Gets the party going ____
Goes out of his/her way for others ____
Loves order and regularity ____
Thinks that all will be well ____
Echoes what others say ____

Keeps in background ____
Inquires about others' well-being ____
Likes to follow a regular schedule ____
Invents problems for himself/herself ____
Decides things on his/her own ____

Feels uncomfortable ____
Takes others' interests into account ____
Leaves his/her belongings around ____
Is able to see the best in a situation ____
Takes charge ____

Speaks softly ____
Waits for his/her turn ____
Does things that others find strange ____
Is afraid that he/she will do the wrong things ____
Will believe anything ____

1	2	3	4	5
not at all applicable				entirely applicable

Has a lot of fun ____
Accepts people as they are ____
Does crazy things ____
Keeps his/her emotions under control ____
Is easily deterred ____

Avoids company ____
Sees to his/her own need first ____
Does things at the last minute ____
Has a dark outlook on the future ____
Takes risks ____

Loves large parties ____
Tries to prevent quarrels ____
Accomplishes his/her work on time ____
Fears for the worst ____
Lets others make the decisions ____

Loves to chat ____
Starts fights ____
Behaves properly ____
Readily overcomes setbacks ____
Looks at things from different angles ____

Is apprehensive about new encounters ____
Respects others' feelings ____
Works according to a routine ____
Can stand a great deal of stress ____
Takes the initiative ____

	1	2	3	4	5	
not at all applicable						entirely applicable

Cheers people up ____
Uses others for his/her own ends ____
Seeks danger ____
Bursts into tears ____
Does what others want him/her to do ____

Keeps apart from others ____
Takes an interest in other people's lives ____
Wants everything to add up perfectly ____
Looks at the bright side of life ____
Is easily intimidated ____

Starts conversations ____
Tells tall stories about himself/herself ____
Leaves his/her work undone ____
Can take his/her mind off his/her problems ____
Waits for others to lead the way ____

Keeps others at a distance ____
Likes to do things for others ____
Does things by the book ____
Knows how to control himself/herself ____
Follows the crowd ____

Prefers to be alone ____
Imposes his/her will on others ____
Works hard ____
Is sure of his/her ground ____
Reacts quickly ____

not at all applicable	1	2	3	4	5	entirely applicable

Acts comfortably with others ____
Wants to be in charge ____
Is easily distracted ____
Panic easily ____
Meets challenges ____

Lives in a world of his/her own ____
Does most of the talking . ____
Keeps his/her appointment ____
Is down in dumps ____
Knows what he/she wants ____

Falls silent when strangers are around ____
Respects the opinions of others ____
Likes to be well prepared ____
Loses his/her temper ____
Is easy to fool ____

Makes friends easily ____
Orders people around ____
Does unexpected things ____
Worries about things ____
Is full of ideas ____

Radiates joy ____
Demands to be the center of interest ____
Neglects his/her duties ____
Keeps a cool head ____
Engages in discussions ____

Appendix 4 Consent Form: Video-Conferencing Experiment

I agree to participate in a study of behavior in multi-party conversations that is conducted by Prof. Roel Vertegaal of the Department of Computer and Information Science, Queen's University.

I understand the following concerning my participation:

1. The experimental task involves participation of a multi-party conversation through a computer-mediated video-audio system. The eye gaze behavior will be measured.
2. A full description of purpose and value of the research will be provided following my participation.
3. All record of my individual participation will be treated as confidential, and only group performance will be reported in any publication.
4. I may terminate my participation at any time.
5. I will receive $10.00 for completing my participation in the study.

The faculty member conducting this research is Prof. Roel Vertegaal. Any complaints or additional questions that I have regarding the study may be expressed to Prof. Roel Vertegaal at 533-3070.

SIGNED: ______________________________ DATED: ________________

Appendix 5 Experimental Instruction

Please keep the comfortable gesture and position. After seated, please don't move the chair or adjust your body position in the chair during the conversation procedure. Although it's not necessary that you remain absolutely still, please don't move your head too much. Get familiar with the screen display. The two avatars on the screen represent your two conversation partners. Get familiar with the microphone and earphone and adjust the volume to the comfortable level.

Your two conversation partners are experiment participants just like you. They are seated in the same setting as you. Be ensured to participate the conversations as usual.

During experiment, three of you will be exposed to three different segments of a sentence. You are required to talk about what segment is on your screen, and discuss with your partner for the solution. You are also required to speak out the solution to other partners. One of your partners is required to type in the solution, which will be displayed on another's screen. The last one will double-check the solution to see if there is any error.

There will be an example before the formal experiment.

Before the formal experiment, there will be a calibration procedure and small test that will calibrate your gaze point on the screen, and may take 2 minutes.

The formal experiment will last about 15 minutes. Please don't touch either the camera or the screen.

Thank you very much for your cooperation!

Appendix 6 Post-experimental Questionnaire

Name:________________________________ Number: _________

Indicate your agreement or disagreement with the following statements by circling the numbers.

(1 indicates complete disagreement; and 5 indicates complete agreement)

1. I found it hard to form sentences with the three of us.

completely disagree *1 2 3 4 5* *completely agree*

2. I have worked together with these people before.

completely disagree *1 2 3 4 5* *completely agree*

3. I enjoyed collaborating with my partners.

completely disagree *1 2 3 4 5* *completely agree*

4. It was hard for me to take a speaking turn.

completely disagree *1 2 3 4 5* *completely agree*

5. My partners paid attention to me during the conversation.

completely disagree *1 2 3 4 5* *completely agree*

6. Sometimes I could not identify which partner was speaking.

completely disagree *1 2 3 4 5* *completely agree*

7a. I found this communication system easy to work with.

completely disagree *1 2 3 4 5* *completely agree*

7b. Could you please explain why?

....

8. It was always clear to whom the others were talking.

completely disagree *1 2 3 4 5* *completely agree*

9. Would you like to work more often using such a communication system? Please explain.

....

....

10. Do you have any further comments?

....

Vita

Name	Yaping Ding
Place and year of birth	Jiangsu, China, 1964
Education	East China Normal University, Shanghai, China, 1987-1990
	Wuhan University, Wuhan, China, 1980-1987
Experience	Senior Consultant, Unisono Ltd., Beijing, China, 1995-2000
	Researcher, Institute of Aviation Medicine, Beijing, China, 1991-1995
	Lecturer, Shanghai University, Shanghai, China, 1990-1991
Awards	Queen's Graduate Award, 2000-2001
	National Award for Development of Science and Technology, China, 1995