Research on Eye Gaze Activities in Conversations under Influence of Communicative Insufficiency

Koki Ijuin

2019 March

Contents

List of Figures

List of Tables

Abstract

The rapid development of technologies has increased the opportunities to communicate with people who have different cultural background and different native-language. In those communications, the low communicative efficiencies of interlocutors, caused by low linguistic proficiencies or different background knowledge, prevent from having decent conversations. Previous studies suggest that nonverbal behavior plays an important role to establish smooth conversations, especially when the communicative abilities of the participants are not sufficient. However, the function of the nonverbal behavior in conversations under influence of communicative insufficiency is still not clear. This dissertation explores to verify how the eye gaze activities, which is one of the most important nonverbal behavior, function in such conversations.

The multimodal three-party conversational corpus was created to compare the eye gaze activities of the same interlocutors between native-language and second-language conversations, where they have massive difference of linguistic proficiency. Forty conversations in each language were collected and their eye gaze activities were manually annotated. The comparative and quantitative analyses of eye gaze activities were conducted in order to explore how the eye gaze activities function for floor apportionment, which is essential for having smooth conversations.

The results showed that the speaker gazes more at the next speaker in the second-language conversations than in the native-language conversations. The results also revealed that there is a significant positive correlation between duration of a speaker gazing toward a listener and ratios of that listener taking the floor in the second-language

conversations. The results also showed that the next speaker's gaze toward the current speaker lead the other listener's gaze toward the current speaker, due to the joint attention effect. That tendency is stronger in the second-language conversations than in the native-language conversations.

These findings showed that the eye gaze activities function efficiently in case where the interlocutors have difficulties in communication.

These findings will contribute to improve the robot's engagement and interaction capabilities by estimating the participant's intention from his/her eye gaze activities and coordinating the conversations with robot's eye gaze activities.

List of Abbreviations

L1	Native language or mother tongue of speakers
L2	Second language of speakers
CS	Current Speaker
NS	Next Speaker
OP	Other Participant
CStoNS	The eye gaze of the Current Speaker toward the Next Speaker
CStoOP	The eye gaze of the Current Speaker toward the Other Participant
NStoCS	The eye gaze of the Next Speaker toward the Current Speaker
OPtoCS	The eye gaze of the Other Participant toward the Current Speaker
SP	Speaker
GL	Mainly-gazed-at Listener
NGL	Not-mainly-gazed-at Listener
SPtoGL	The eye gaze of the speaker toward the Mainly-gazed-at Listener
SPtoNGL	The eye gaze of the speaker toward the Not-mainly-gazed-at Listener
GLtoSP	The eye gaze of the Mainly-gazed-at Listener toward the speaker
NGLtoSP	The eye gaze of the Not-mainly-gazed-at Listener toward the speaker

Chapter 1

Introduction

1.1 Background

Globalization has given us many more opportunities to communicate with people all over the world, and the development of transportation and information technologies has led to the growth of global collaborative research and business [1]. The Japanese government has reported that international trade is expanding and thus increasing opportunities for international cooperation [2]. They also reported that the number of researchers coming from abroad as well as those leaving Japan is increasing year by year [3]. Furthermore, these reports argue that the need to communicate and discuss various issues with research or business partners having different native languages and backgrounds is increasing even for those Japanese professionals who do not go abroad. In this situation, it is difficult to establish smooth conversation due to the varying levels of the interlocutors' communicative efficiency, including linguistic proficiency [4, 5]. A previous study revealed that conversations in a second language typically involve participants with various levels of linguistic proficiency and that such unbalanced skills may disrupt collaboration for both native and non-native speakers in face-to-face communication [6], thus preventing participants from fully utilizing their collaborative abilities.

These problems caused by unbalanced linguistic proficiencies can be observed in other conversational situations. One such example is conversation with humanoid robots. The rapid development of technology has given us opportunities to verbally communicate with humanoid robots [7, 8], although the current human-robot interactions are not yet sufficiently smooth or conducive to easy communication. In those situations, there is a huge difference in communicative efficiency between humans and robots due to the low level of the robot's linguistic proficiency and its lack of background knowledge. Several studies have attempted to establish smooth conversations with humanoid robots by using nonverbal behaviors that are less dependent on linguistic proficiency. They suggest that the nonverbal behaviors of humanoid robots help to share and understand the interlocutor's attention [9, 10, 11]. However, no study has elucidated the general mechanism of how nonverbal behavior contributes to unbalanced conversations. Consequently, there is urgent issue need to understand how such conversations are constructed in order to help make them more efficient and smoother.

In this study, we analyzed how nonverbal behaviors function in conversations under the influence of communicative insufficiency. We focused on multiparty second-language conversation as the experimental subject, since it is one of the most common types of unbalanced conversations and, in contrast to native-language conversations, it can clearly highlight the differences in behavior caused by linguistic proficiency. Comparative and quantitative analyses of the interlocutors' behaviors in native- and second-language conversations were conducted in order to grasp how nonverbal behaviors compensate for the lack of linguistic proficiency.

Previous research on human interaction was mainly conducted with native-language conversations, where the interlocutors have the same background and high level of linguistic proficiency. Sacks et al. showed that turns in conversations are constructed out of linguistic units that have recognizable structures, enabling the next speaker to project the structure in advance and anticipate the possible completion of the unit [12]. These research works also found that interlocutors use not only verbal but also nonverbal behaviors, such as eye-gaze activities, to construct conversations. Several studies have aimed to further understand human-human communication and multimodal signaling of social interactions [13, 14, 15]. Previous studies have also pointed out the role of gaze in regulating linguistic turn taking [16, 17, 18, 19, 20]. Quantitative studies have reported that eye gaze plays an important role in monitoring the understanding of the content of conversations by communication partners and in contributing to the performance of collaborative tasks [21, 22, 23]. Argyle also reported that participants gazed nearly twice as much while listening as while speaking. Speakers monitor addressees to ensure understanding and, when necessary, alter their utterances as they make them [23]. Eye-gaze activities contain multiple functions, and people use eye-gaze activities consciously or unconsciously to establish smooth conversations, even when the interlocutors have high levels of linguistic proficiencies.

As mentioned above, the opportunities to have second-language conversations are increasing year by year, and unbalanced proficiencies in the second language may also lead to uneven opportunities for participation in conversations. Although difficulties were observed in those conversations, few studies have investigated how such conversations are constructed. One previous research work on second-language dyad conversations suggests that language proficiency may affect the function performed by eye gaze [24]. Veinott et al. also found that gaze activities function differently between native- and second-language conversations [25]. These results suggest that visual information and nonverbal behaviors compensate for low levels of communicative efficiencies such as low linguistic proficiency or the lack of common background. A multiparty second-language conversation consists of "ratified participants" [26], while participants with poorer proficiency might be relegated to "side participant" status regardless of their level of proficiency in the tasks they are tackling collaboratively. However, very little research has been conducted on the function of nonverbal behaviors in such situations from the perspective of qualitative analysis. Accordingly, it is still unknown how interlocutors use their nonverbal behaviors for smooth conversations in such conditions.

1.2 Thesis Statement

We focused on eye-gaze activities as one of the most important nonverbal behaviors for smooth conversation, based on the findings of previous studies. The main purpose of this research is to understand the function of eye-gaze activities in compensating for the low communicative efficiency in human-human conversations. We assume that quantitative and comparative analyses of eye-gaze activities between L1 and L2 could highlight that function of eye-gaze activities, since there is a massive difference in linguistic proficiency between L1 and L2.

This dissertation presents quantitative and comparative analyses of the functions of eye-gaze activities between native-language (L1) and second-language (L2) conversations from the perspective of linguistic proficiency. A new multimodal conversational corpus was created by collecting both native- and second-language conversations from each conversational group in order to compare the eye-gaze activities in those conversations. To reduce the effect of each participant's characteristics, the participants have both L1 and L2 conversations in order to compare eye-gaze activities of the same participant. Then, analyses of eye-gaze activities were conducted from the perspective of floor apportionment, which is one of the most important structures used to establish smooth conversations. Floor apportionment in those conversations was studied in order to grasp how eye-gaze activities contribute to L2 conversations, where a smooth speaker change is one of the keys to establishing smooth conversations.

The findings obtained from this study can be utilized to develop technologies for monitoring the understanding and the contributions of all participants and for supporting smooth interactions in L2 conversations. They can also be categorized into eye-gaze activities in conversations under the influence of communicative insufficiency.

This work is expected to contribute to the creation of eye-gaze behavior models of humanoid robots by applying these findings to improve the robot's engagement and interaction capabilities from the perspective of nonverbal behavior, which is less dependent on conversational topic or domain.

1.3 Thesis Structure

This dissertation presents comparative and quantitative analyses of eye-gaze activities, which is one of the most important types of nonverbal information, between L1 conversations and L2 conversations. Three analyses of eye-gaze activities were done as follows: (1) basic statistics of eye-gaze activities for entire conversations, (2) eye-gaze activities from the perspective of floor apportionment, and (3) eye-gaze activities from the perspective of the effect of speaker's eye gaze on floor apportionment.

The dissertation is organized as follows:

■ **Chapter 2 - Related Work**

This chapter describes the previous research related to eye-gaze activities. We introduce the general functions of eye-gaze activities and then focus on the function of floor apportionment in multi-party conversations. The previous work of our research group is also presented at the end of this chapter.

■ **Chapter 3 - Multimodal conversational Corpus**

This chapter describes the multimodal conversational corpus we use to analyze the relationship between eye-gaze activities and communicative efficiency.

■ **Chapter 4 - Analyses of Utterances and Eye-gaze Activities**

This chapter overviews the analyses of how much and how long participants uttered or gazed at the other participants. We formulated a gazing ratio to measure the eye-gaze activities of participants quantitatively.

■ **Chapter 5 - Eye-gaze Activities for Floor Apportionment**

Based on the results in Chapter 4, this chapter gives detailed analyses of the relationship between eye-gaze activities and floor apportionment. We classified the participants into three roles, i.e., current speaker, next speaker, and other participant, to analyze how each participant uses eye gaze for floor apportionment.

■ **Chapter 6 - Effect of Speaker's Eye-gaze Activities for Floor Apportionment**

We analyzed the effect of the speaker's eye-gaze activities on floor apportionment. Listeners were classified into two groups according to their status of being targets of the speaker's eye gaze: mainly-gazed-at listeners and not-mainly-gazed-at listeners. Then, we analyzed the relationship between the gazing ratio and the ratio of a listener taking the floor.

■ **Chapter 7 - Overall Discussion**

This chapter provides an overall discussion covering all of the analyses.

■ **Chapter 8 - Conclusion**

This chapter summarizes the main contributions of this dissertation and discusses future work.

Chapter 2

Related Works

2.1 Human's Eye Gaze Activities

In typical human-human face-to-face interactions, the interlocutors use not only speech and language but also a wide variety of paralinguistic means and nonverbal behaviors to signal their speaking intentions to the partner [16], to express intimacy [27] [28], and to coordinate their conversation [29]. Eye-gaze activity is one of the strongest and most extensively studied visual cues in face-to-face interaction. A human's gaze behavior is a highly evolved system that allows us to detect an opponent's gaze direction at a distance by gazing at his/her eyes, and thus eye gaze can be expected to have communicative functions [30, 31]. Eye-gaze activities have dual functions: they can both signal and perceive. Consequently, eye-gaze activities provide a remarkable tool for social interaction [32]. Seeing others and being seen by others has a special significance in human interactions, which goes beyond the mere perceptual or communicative functions of the eye. Bavelas et al. showed that the listener tends to respond when the speaker gazes at him/her, and the speaker tends to gaze away soon after the listener responds; in this way, the speakers and listeners create and use mutual gaze to coordinate their actions

[33]. Previous studies have associated a variety of functions with eye gaze activities, such as managing the attention of interlocutors [34], expressing intimacy and exercising social control, highlighting the information structure of the propositional content of speech, and coordinating turn-taking [16, 17]. Other studies have established that joint attention is important from a developmental point of view because it helps in understanding the others' thoughts and intentions [35]. These findings show that eye-gaze activity has multiple functions not only in perceiving visual information but also in the process of becoming social beings.

2.2 Eye-gaze Activities for Regulating Conversations

The fundamental gaze patterns related to turn negotiation were discussed in Kendon [16]. Kendon stated that eye-gaze activity has at least four such functions: (1) to provide information; (2) to organize the conversation's flow; (3) to interpret emotions and relationships; and (4) to concentrate on understanding the utterance by shutting out visual information [16]. For the function of providing information, the previous research suggests that eye-gaze activities provide many kinds of information such as liking and attraction [36], attentiveness, [37], and competence [38]. The eye-gaze activities can be used to evaluate those communicative behaviors. There is another important function for establishing smooth conversations, which is the eye-gaze function of organizing a conversation's flow. Kendon [16] suggested that eye-gaze activities such as gazing at or avoiding conversational partners might be used for some functions of turn organization during two-person conversations. Duncan et al. [39] argued that the speaker's eye-gaze activities have a floor-apportionment function in conversations and function as a turn-yielding cue. They reported that speakers tend to gaze away at the beginning of turns but tend to gaze toward the recipient when approaching turn completion in order to signal that they are ready to turn the floor to the other person. Kendon also stated not only the relevance of the speaker's eye-gaze activities to floor apportionment but also how they affect the timing of floor apportionment [16]. He suggested that the speaker gazes at the listener at the end of his/her turn for giving a turn, while the listener gazes away from the speaker at that time in order to concentrate on generating his/her next utterance. The other previous studies also suggest that listeners gaze away during the speaker's utterances or during the gap between turns in order to provide clear evidence of anticipation of the next turn [40, 41, 42, 43]. Argyle and Cook also found that participants gaze nearly twice as much while listening as they do while speaking [44]. While those studies reported that gaze plays an important role in floor apportionment, one study observed that there was no relation between eye gaze and floor apportionment in a conversational setting where two participants play different social positions [45]. Another study reported that the importance of visual communication in synchronizing conversation is not that it permits eye contact but that it enables the speaker and listener to simply see and be seen by each other [46]. Those results suggest that eye-gaze activity combines many functions, not only coordinating the conversations but also sociality, and that the condition of the conversational setup might change the relative importance of these functions [47].

2.3 Eye-gaze Activities in Multi-party Conversations

Those studies mainly dealt with two-party dialogues, not multiparty conversations where the features used for managing turn control may be different from those used in two-party dialogues. In such multiparty conversations as a group of people informally chatting with each other or people attending a more formal meeting, it is obvious that coordination and interaction cannot be managed in a similar way to how this is done in dialogues between two speakers who share the responsibility for coordination. Goodwin claimed that the gaze of the speaker during utterances should identify the party being gazed at as the addressee of this speaker's utterances [48]. For turn management with eye gaze, the speaker signals the assumed next speaker with her/his gaze, thus requiring the gazed-at participant to see that gaze while the other participant also grasps the expectation that someone else will speak next. Kalma reported that the recipient of a prolonged gaze, i.e., the participant who is gazed at by the speaker during the silence after his/her utterance, tended to take the floor [49]. Leaner also reported that the speaker anticipates the next speaker explicitly in many ways, including eye gaze [50]. The previous studies on eye-gaze activities of a third listener, or observer, suggest that the listeners normally gaze at the current speakers and shift their gaze toward the next speaker just before he/she begins to speak [51, 43]. On the other hand, Holler and Kendrick observed that unaddressed participants were able to anticipate the next turns, and that they often shifted their gaze toward the next speaker before the current turn ended [52]. Vertegaal discussed the importance of gaze in multiparty conversations for signaling conversational attention [34], and Jokinen showed that the speaker's gaze is important for coordinating turn taking in multiparty conversations and that partners pay attention to the speaker's gaze behavior [53]. Some researchers observed that gaze behavior is affected by the type of communication (social actions like requesting vs. complaining) and the length of turns as projected by the social action [54, 55, 56, 52, 57]. They also showed that the planning of these gaze shifts virtually coincided with the points at which the turns first became recognizable as possibly complete. These studies showed that the eye gaze of the speaker in the conversations has a relation to floor apportionment in multi-party conversations.

2.4 Eye-gaze Activities in Second-language Conversations

Compared to the wide coverage of gaze studies in native-language (L1) conversations, only a limited number of studies on gaze have been conducted in the condition where interlocutors have low levels of communicative efficiency, such as second-language (L2) conversations. In L2 conversations, the interlocutors have difficulty in delivering and understanding utterances due to their low linguistic proficiency. The linguistic proficiency of conversational participants is expected to range widely from low to high in L2 conversations, and some studies reported that this difference in proficiency may affect nonverbal behavior in such conversation. In those conditions, interlocutors have difficulty in uttering and understanding their utterances with verbal information, so there is a possibility that the functions of eye-gaze activities for coordinating conversations may be more important in L2 conversations in order to compensate for their lack of linguistic proficiency. Hosoda [24] suggested that language expertise may affect the functions of gazes in conversation where the participants need to monitor their partner's degrees of understanding. Veinott et al. [25] found that non-native speaker pairs benefited from using video communication in route-guiding tasks compared with audio communication, whereas native-speaker pairs did not. They argued that this was because video transmitting facial information and gesture helped the non-native pairs to negotiate a common ground, whereas such video did not provide significant help to the native pairs. These observations suggest that nonverbal behavior and visual information more greatly affect efficiency in L2 conversations than in L1 conversations.

Additionally, several previous research works suggest that the gazing behaviors do not correlate highly with their verbal ratings. They reported small positive correlations between gaze and positive affective ratings in L1 conversations [58, 59]. Correlations between gaze and verbal ratings may be low because people are generally not aware of their gazing behaviors in an interaction [47]. We can interpret these results as showing that eye-gaze activities can function individually from verbal information for regulating the conversations and that eye-gaze activities function more efficiently in conditions of low linguistic proficiency.

To quantitatively and precisely analyze eye-gaze activities from the perspective of communicative efficiency, our research group [60] created a multimodal corpus of three-party conversations in L1 and L2, where there is a huge difference in linguistic proficiency between L1 and L2 conversations. In this way, it was possible to compare the features of utterance, eye gaze, and body posture in L1 and L2 conversations conducted by the same interlocutors [60]. To compare the features of eye gaze in L1 and L2 conversations, they used two metrics: (1) how long the speaker was gazed at by other participants during her or his utterance (listener's gazing ratio) and (2) how long the speaker gazed at other participants during her or his utterance (speaker's gazing ratio). The experimental results show that the averages of speaker's gazing ratios are almost the same in four kinds of conversations (two different conversation topics and two different conversation languages), whereas the averages of the listener's gazing ratios are larger in L2 conversations than in L1 conversations for both conversation topics.

Chapter 3

Multimodal Conversational Corpus

3.1 Purpose of Creating a Corpus

To analyze the eye-gaze activities in conversations, it is necessary to consider individual differences and the effect of the conversational partner. Our research group created a three-party multimodal conversational corpus in L1 and L2 in order to make comparative analyses of eye-gaze activities with the same participant between L1 and L2 conversations. We expect quantitative and comparative analyses of eye-gaze activities between L1 and L2 to highlight the function of eye-gaze activities in light of the massive difference in linguistic proficiency between L1 and L2. We collected three party-conversations in which eye gaze might be more important than in dyad conversations. This setup is expected to capture the function of eye-gaze activities for floor apportionment, which might be more efficient in selecting the next speaker. We collected 80 multimodal conversational data sets. Table 3.1 lists the specifications of the multimodal conversational corpus.

Table 3.1 Quantitative features of multimodal conversational corpus

Feature	Value
Total participants	20 conversational groups of 3 participants
Average duration of conversation	6 min.
Conversational types	2 types (free-flowing, goal-oriented)
Conversational languages	Japanese (L1), English (L2)
TOEIC scores of participants	450-985 points

3.2 Experimental Setup

Three subjects participated in a conversational group, sitting in a triangular formation around a table. The distance between participants was about 1.5 meter. Three sets of eye trackers and headsets with microphones were used to record the eye gazes and voices of all three participants. They talked about two types of predetermined topics in English as a second language and in Japanese as their native language (i.e., each group participated in four conversations). All of the sessions were conducted in a lecture room at Doshisha University. A scene of the experimental setup is shown in Figure 3.1.

The NAC EMR-9 (Nac Image Technology Inc.) eye trackers shown in Figure 3.2 were used in this experiment, and Figure 3.3 outlines the system configuration. The EMR-9 is a cap-mounted eye-tracking system, and the participants could move freely in carrying out their conversational activities. It has one eyesight camera and two eye cameras with near-infrared sensors. The eyesight camera recorded the scene that the subject was gazing at. The angle of view was 62 degrees. The eye camera recorded the eye movements of each participant at a sampling rate of 60 fps. The near-infrared sensors recorded the participants' pupils and a figure that is reflected from the cornea, which is called a Purkinje figure. Pupil corneal reflection tracking was adopted to obtain the Purkinje figure. Their eye gazes were not tracked when the participants blinked, when

Figure 3.1 Experimental setup

they laughed, or when their eyes had narrowed so much that the eye tracker could not detect their pupils. Figure 3.4 shows a screenshot of the eyesight camera. The eye marks shown with the square, cross and white circle indicate the eye mark for the right eye, left eye and compensated eye mark between the left and right eye mark, respectively. The EMR-9 has a display, and all of the settings are made with its own GUI-based system.

Figure 3.2 Image of EMR-9 and correct way to wear it

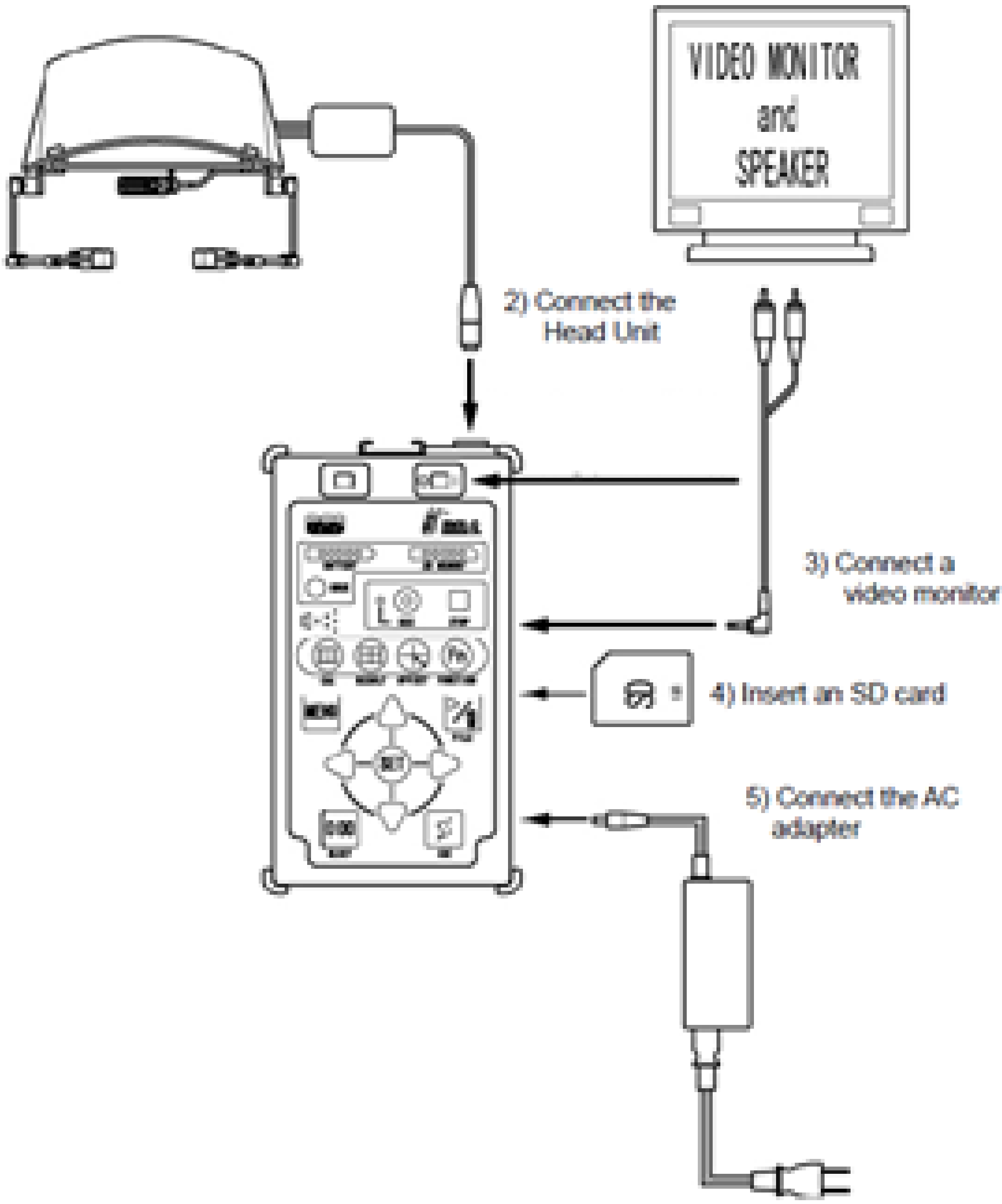

Figure 3.3 System configuration of EMR-9

Figure 3.4 Screenshot of eyesight camera of EMR-9

3.3 Participants

A total of 60 subjects (23 females and 37 males: 20 groups) between the ages of 18 and 24 years participated in this experiment. They were Japanese university students who had acquired Japanese as their L1 and had learned English as their L2. In order to normalize the effect of social position, we set the groups of participants in such a way that they were not acquainted with each other before the meeting held for data collection. Their communication levels in English were measured based on the Test of English for International Communication (TOEIC). Figure 3.5 shows the cumulative distribution of the participants' TOEIC scores. Their scores ranged from 450 to 985 (990 being the highest possible score). Note that the TOEIC scores were not related to group composition. This is because both balanced and unbalanced linguistic proficiencies of the participants in a group typically exist in real situations.

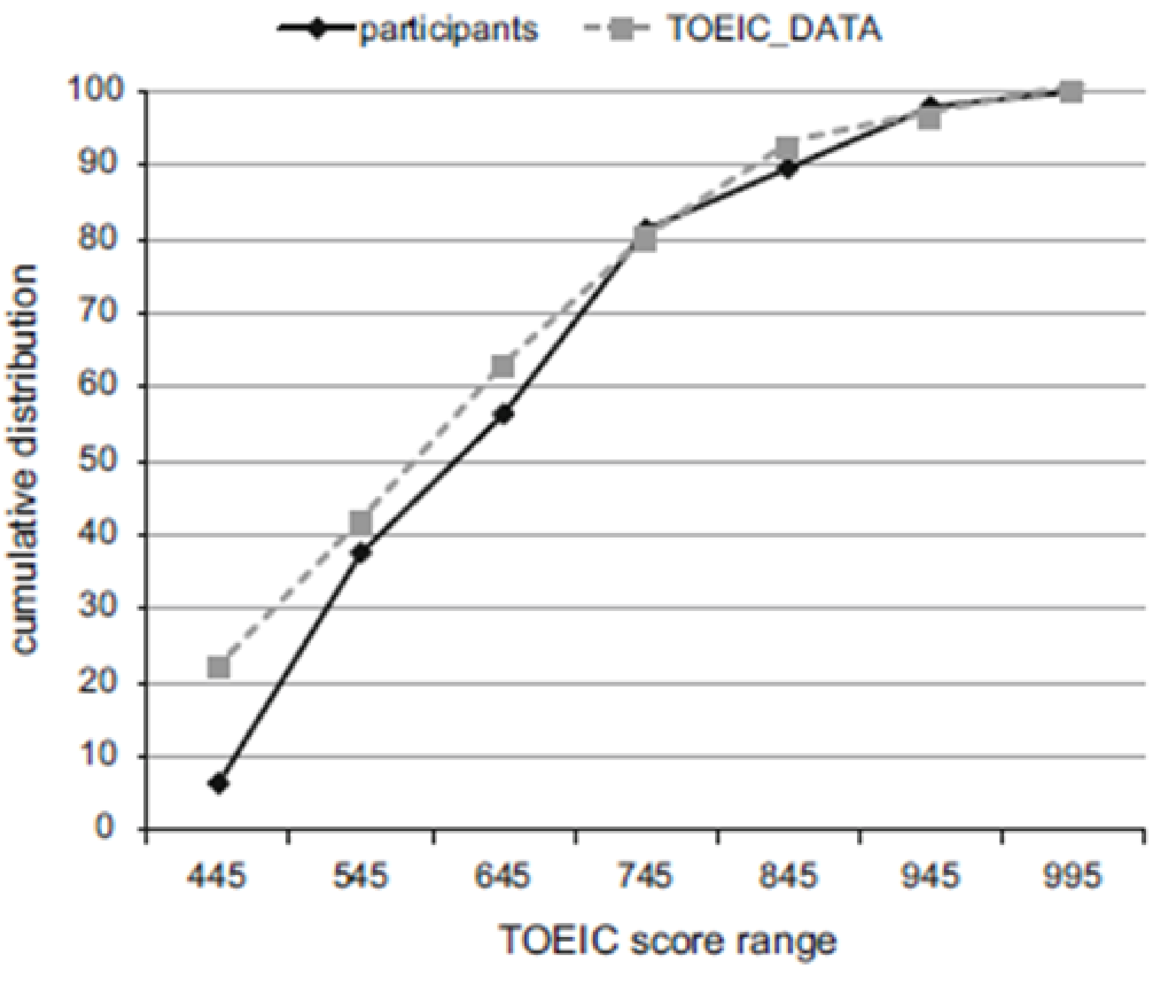

Figure 3.5 Cumulative distribution of participants' scores. Note that the gray line (TOEIC_TEST) represents cumulative distribution of the TOEIC scores of the latest TOEIC test administered nationwide

3.4 Procedure

The procedure for the experiment is as follows. First, the content of the conversational topic was explained to the participants. There were two types of conversational topics. The first one was free-flowing (Free), which was natural chatting that covered various topics such as hobbies, weekend plans, studies, and travel. The second one was goal-oriented (Goal), in which they collaboratively decided on a specific topic such as what to take with them on trips to uninhabited islands or mountains. The orders of the conversation types and languages were randomly arranged to cancel out the effect of order.

After the instructions were given, eye trackers were mounted on each participant and then calibrated. Figure 3.6 shows screenshots of the eyesight camera before (left) and after (right) calibration. For the calibration, we asked the participants to gaze at certain points, shown in the box in Figure 3.6, without moving their heads. If the calibration was done correctly, intersection points of the lines were located in each box, as shown in the right side of Figure 3.6. We continued calibration until its completion by changing the positions of eye cameras or changing the threshold values for detecting pupils. After the calibration was done, participants began conversations from the time of a flashlight signal made by the experimenter. Each group had free-flowing and goal-oriented conversations in L1 and in L2. The participants filled in a questionnaire after each conversation. The average duration of individual data items was 6 minutes. Consequently, the subjects participated in four conversations.

Figure 3.6 Screenshots of eyesight camera before (left) and after (right) calibration

3.5 Annotations

The multimodal corpus analyzed here was manually annotated with the time spans for utterances, backchannel, laughing, and eye movements using EUDICO Linguistic Annotator (ELAN) [61] (Figure 3.7).

We performed the annotation according to the MUMIN annotation scheme [62] used in our previous research for modeling turn-taking behaviors in L1 conversations [53]. The annotation features and values adopted in our previous research are listed in Table 3.2. We confirmed the inter-coder agreement between the annotators by measuring Cohen's kappa coefficient for DialogAct, Head Movement, and Hand Movement (Handness, Trajectory Right Hand and Trajectory Left Hand) in L2 conversations. Cohen's kappa coefficients were 0.55, 0.14, and 0.34 for Dialogue Act, Head Movement, and Hand Movement, respectively. These values are not so high, especially for Head Movement and Hand Movement. Taking these preliminary tests into account, we decided to limit annotations to the features of DialogAct and GazeObject, which were reported to be important in monitoring conversations [53]. These features also maintained high agreement among the annotators, and Cohen's kappa coefficients were 0.55 and 0.83 for DialogAct and GazeObject.

For annotating DialogAct, we manually determined the start and end times of each utterance by considering only pause durations (500 ms) between consecutive speech segments, that is, without considering the contents of consecutive speech segments. This annotation can also trace the floor apportionment based on the pause duration between consecutive utterances.

The annotation of GazeObject is composed of start and end times and attributes such as "gaze at the right person," "gaze at the left person," and "gaze at the other" in order to clarify the gaze direction of each participant. We concentrate on the local attention level and thus count each gaze shift as a separate gaze event. Note that the small movements around the same gaze object are included in one continuous gaze event. There are two reasons for this kind of eye movement: the so-called saccades, which refer to the involuntary and constant movements of the eyes' fixation points, and the agent's own involuntary eye movement around the gaze object while generally focusing on the object. Gaze events were manually annotated when the participant focused his/her visual attention on a particular object for a certain period of time (more than 200 msec).

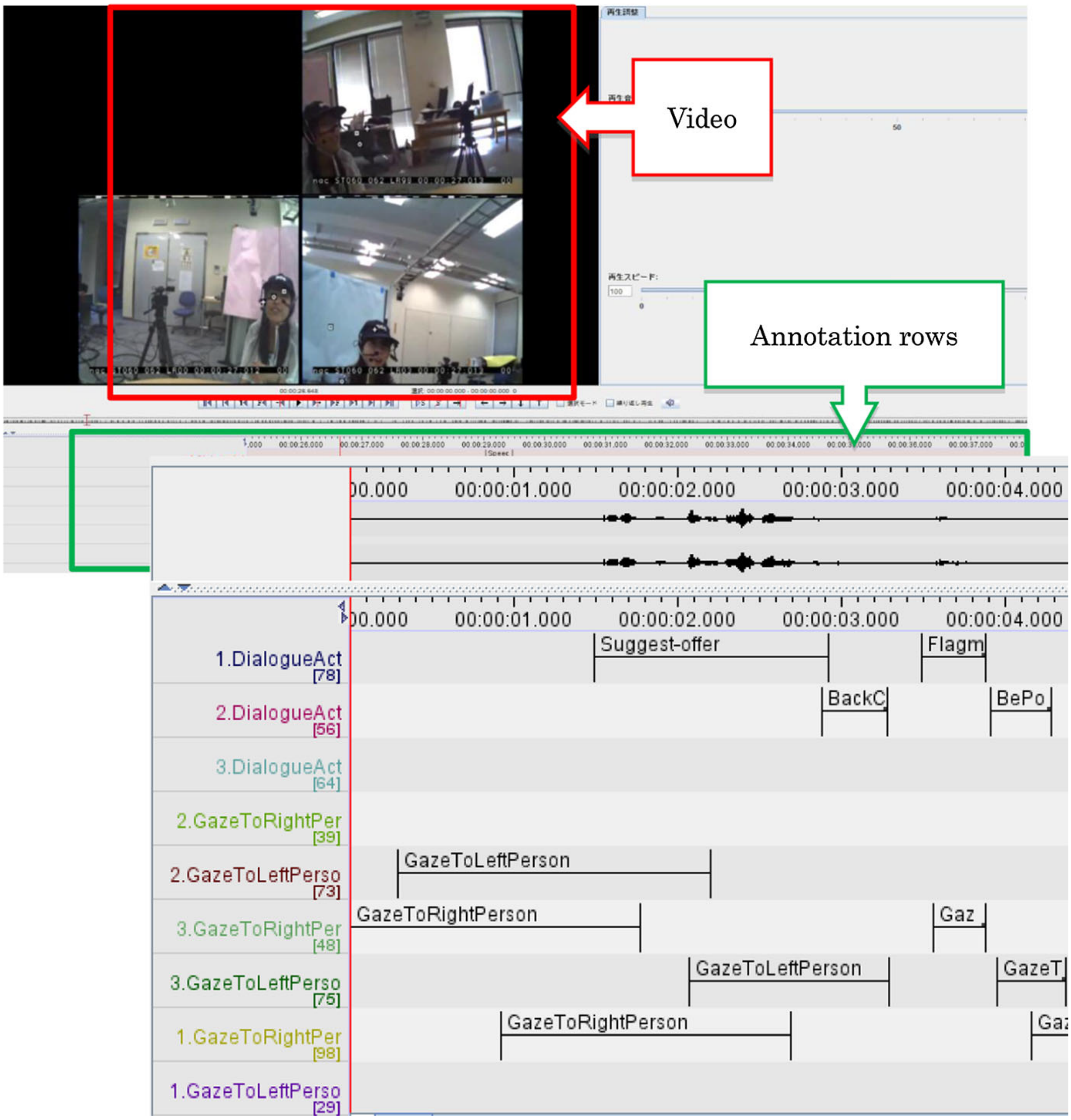

Figure 3.7 Screenshots of executing ELAN and its annotation rows

Table 3.2 Annotation features and values in previous research. Note that the features written in bold text are annotated in our multimodal conversational corpus

Annotation feature	Feature value
DialogAct	Backchannel, stall, fragment, bepositive, benegative, suggest-offer, inform, ask, other
GazeObject	Gaze at the right person, gaze at the left person, gaze at the other
Head Movement	Nod, jerk, backward, forward, tilt, TurnToPartner, TurnSide, waggle, other
Head Repetition	Single, repeated, none
Handness	Both, single
Trajectory Right Hand	Forward, backward, side, up, down, complex, other
Trajectory Left Hand	Forward, backward, side, up, down, complex, other
Hand Repetition	Single, repeated, none
Turn	Give, take, hold

Chapter 4

Analyses of Utterances and Eye-gaze Activities

To gain an overview of how the participants interacted with each other in the conversations, we calculated the basic statistics of utterance number and duration. Then we formulated "gazing ratio" in order to capture how long the participants gazed at others. Note that backchannel is excluded from utterances in the following analyses because backchannel is treated as a reaction but not as an utterance.

4.1 Basic Statistics of Utterances

Table 4.1 lists the basic statistics for the total utterance duration (TUD), the average utterance duration (AUD), and the number of utterances in four kinds of conversations, i.e., those on goal-oriented topics in L1 and L2 and free-flowing ones in L1 and L2 conversations. ANOVA tests for TUD and AUD were conducted with both the language difference and conversation topic difference as within-subject factors. The results for

Table 4.1 Total average duration of utterances, average duration of each utterance, and average number of utterances in L1 and L2 conversations

	L1 conversation	L2 conversation
Total utterance duration (s)	97.7 ± 37.6	65.0 ± 34.1
Average utterance duration (s)	1.48 ± 0.47	1.20 ± 0.44
Number of Utterances	68.7 ± 23.9	55.8 ± 27.9

TUD show a significant main effect of language difference ($F_{(1,59)} = 76.8, p < .01$). The results for AUD also show a significant main effect of language difference ($F_{(1,59)} = 30.5, p < .01$).

These results show that the participants uttered shorter utterances in L2 conversations than in L1 conversations. There is a possibility that the participants have difficulty with producing utterances in L2 conversations, and even if they did utter them, the utterances were simpler and shorter. This possibility leads us to conclude that there is a difference in communicative efficiency between L1 and L2 conversations.

4.2 Methodology of Calculating Eye-gaze Activities

We compared the eye-gaze activities between L1 and L2 conversations quantitatively, focusing on the difference in eye-gaze activities for floor apportionment between L1 and L2 conversations. We used the speaker's gazing ratio and the listeners' gazing ratios to analyze the eye-gaze activities of participants. Gazing ratios are defined as

$$Speaker's\,Gazing\,Ratio = \sum_{i=1}^{n} \frac{DSGL_{(i)}}{DU_{(i)}} \times 100\%, \tag{4.1}$$

$$Listener's\,Gazing\,Ratio = \sum_{i=1}^{n} \frac{DLGS_{(i)}}{DU_{(i)}} \times 100\%, \tag{4.2}$$

where $DSGL_{(i)}$, $DLGS_{(i)}$, and $DU_{(i)}$ represent the duration of the speaker gazing at the the listener during the i-th utterance, the duration of the listener gazing at the speaker during the i-th utterance, and the duration of the i-th utterance, respectively.

4.3 Analyses of Eye-gaze Activities in All Utterances

To understand how the participants used their eye-gaze activities throughout the conversations, we calculated the eye-gaze activities for both the speaker gazing at the listener and the listener gazing at the speaker during utterances. Table 4.2 lists the basic statistics of the speaker's and listeners' gazing ratios in L1 and L2 conversations. ANOVA tests were conducted for both speaker's gazing ratios and listener's gazing ratios with language difference and topic difference as within-subject factors. These analysis results are summarized as follows.

1. The average speaker's gazing ratio was almost the same among the four kinds of conversations.
2. The average listener's gazing ratio was larger in L2 conversations than in L1 conversations ($F_{(1,119)} = 106.7, p < .01$).

This statistical result is highly consistent with the findings of Vertegaal et al. in multiparty conversations [34], who reported that people gaze 1.6 times more while listening than speaking, even though there are differences in languages, conversation topics, and the number of participants between the two sets of multimodal data on multiparty conversations.

Table 4.2 Speaker's gazing ratios and listener's gazing ratios in L1 and L2 conversations

	L1 conversation	L2 conversation
Speaker's gazing ratio	27.9%	28.2%
Listener's gazing ratio	45.4%	57.4%

4.4 Distribution of Listener's gazing ratios

We calculated the distribution of the listener's gazing ratios for each utterance to find which factor causes the difference in the averages of Listener's gazing ratios between conversations in L1 and in L2. Figure 4.1 and Figure 4.2 show the frequency histograms, in percentage, of the listener's gazing ratios in both languages' conversations. The leftmost bar in both figures depicts the percentage of frequency in which the listener's gazing ratio was 0%. The three values written in the figures show the percentage of utterances where the listener's gazing ratio was 0%, the percentage of utterances where the listener's gazing ratio was larger than 0% but smaller than 100%, and the percentage of utterances where the listener's gazing ratio was 100% (left to right).

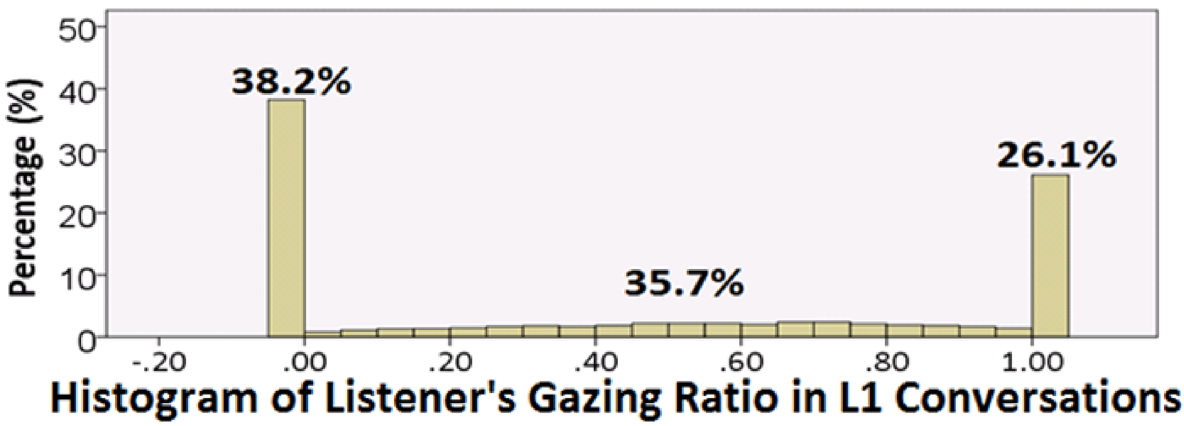

Figure 4.1 Relative frequencies of listener's gazing ratios in L1 conversations

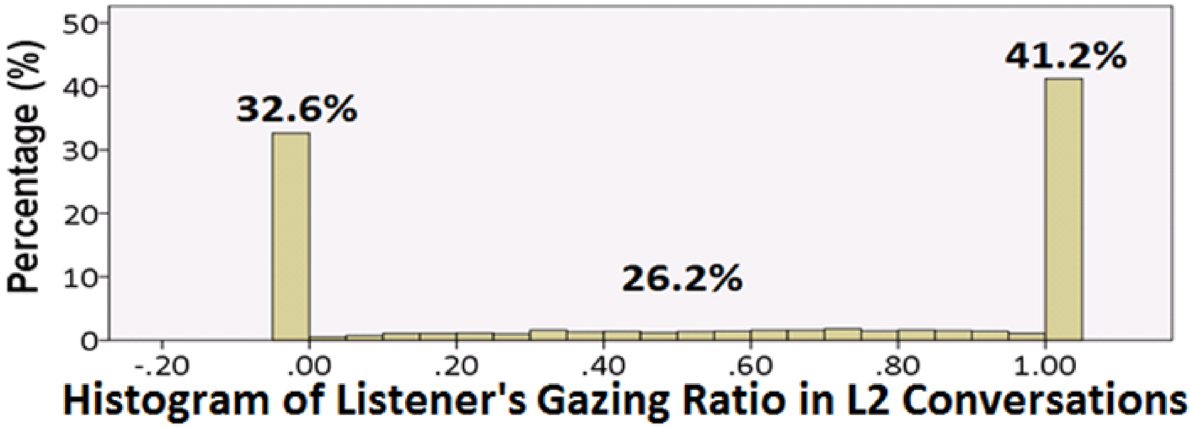

Figure 4.2 Relative frequencies of listener's gazing ratios in L2 conversations

These results show that the frequency of utterances during which the speakers were gazed at from the beginning to end was higher in L2 conversations than in L1 conversations. We then conducted a t-test on the average of listener's gazing ratios between L1 and L2 conversations by extracting the data when the listener's gazing ratio was 0% and 100%. Consequently, there was no significant difference. These results show that the difference in the frequency of utterances, where the Listener's gazing ratio was 100%, caused a significant difference in the average listener's gazing ratios between conversations in L1 and L2.

4.5 Discussion

The comparative analyses between L1 and L2 show that the speakers are gazed at more by their listeners in L2 conversations than those in L1, whereas the speakers gaze at listeners similarly in L1 and L2 conversations. Several possible reasons arise to explain the difference between gaze activities in the conversations in different languages, among them: (1) participants monitored their understanding of what was being said to make repairs if necessary, (2) participants used visual information to help them perceive the auditory information, (3) participants gave a polite acknowledgement of the speaker's effort in producing speech with difficulty, although the reason why the listeners gaze more at the speaker is still not clear. It is also not clear why there is no difference in the speaker's eye-gaze activities between L1 and L2 conversations.

The analyses of the distribution of listener's gazing ratios show that the listeners in L2 conversations could gaze at the speaker during his/her utterance from the beginning more often. These findings lead us to conclude that the listeners are better able to predict in L2 than in L1 conversations who the speaker will be, and thus they have more chances to gaze at the speaker from the beginning of the utterances. Detailed analyses focused on floor apportionment is needed to confirm why the listeners can predict the next speaker more often in L2 than in L1 conversations.

Chapter 5

Eye-gaze for Floor Apportionment

In our previous analyses, eye-gaze activities were analyzed regardless of the sequential information of utterances. Although these analyses provided interesting results regarding the difference in eye gaze between L1 and L2 conversations, they could not explain why the listeners gazed more at the speaker in L2 conversations. The results of the distribution of listener's gazing ratios show that the listeners in L2 conversations might be able to predict the next speaker, and thus they could gaze at the speaker from the very beginning of his/her utterance. This interpretation leads us to analyze the relationship between eye-gaze activities and floor apportionment. The previous studies suggested that eye-gaze activities support smooth floor apportionment, especially in multi-party conversations [34], [53]. There is a high possibility that eye-gaze activities can be used efficiently for floor apportionment when the participants have difficulty in generating utterances and in understanding; here, the eye-gaze activities compensate for low linguistic proficiency in order to establish smooth turn taking and giving.

We formulated two hypotheses based on the previous results:

1. The speaker's eye gaze affects floor apportionment more significantly in L2 than

in L1 conversations,

2. Longer listeners' eye gazes in L2 conversations function efficiently in executing smooth floor apportionment.

5.1 Classification of Participants and Utterances

The participant roles are classified into three types: current speaker (CS) as the speaker of the utterance, next speaker (NS) as the participant who takes the floor after the current speaker releases the floor, and other participant (OP), who is not involved in floor apportionment at that time. The participants were classified into these roles for each utterance.

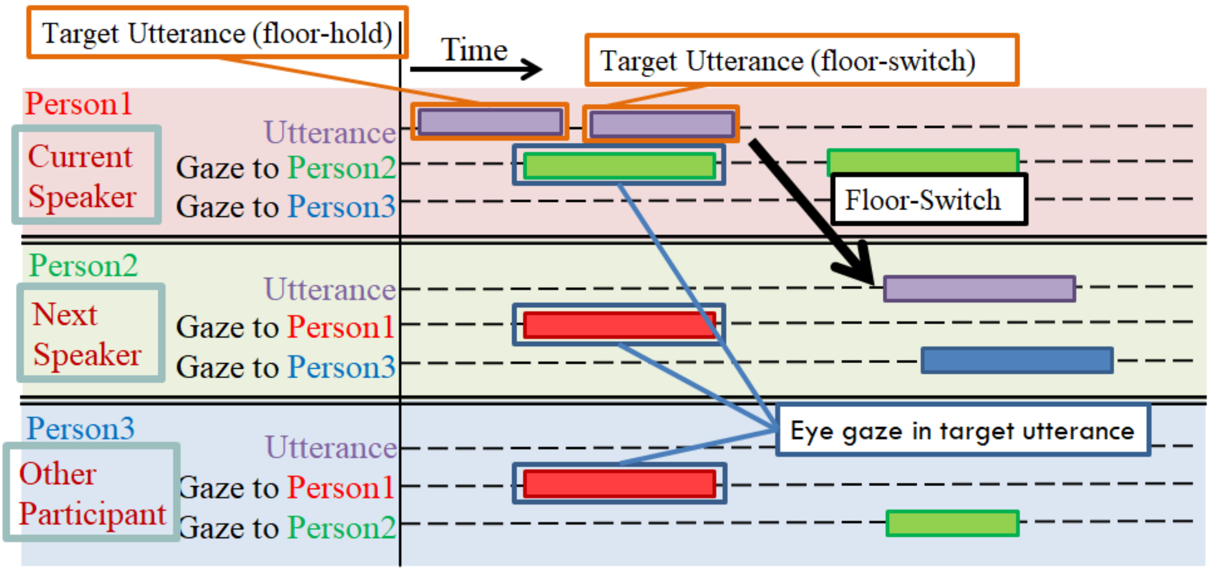

Figure 5.1 Classification of participants

To analyze quantitatively the relationship between eye gaze and floor apportionment, we classified utterances into two groups: utterances after which the speaker held the floor and utterances after which the other took the floor (referred to as utterances "with floor-hold" and utterances "with floor-switch," respectively, in the following discussion).

Table 5.1 Proportions of utterances with floor-hold and with floor-switch for each language's conversation

	L1 conv.	L2 conv.
Utterances with floor-hold	22.2%	43.1%
Utterances with floor-switch	77.8%	56.9%

Table 5.1 shows the proportion of utterances with floor-hold and floor-switch for each language's conversation. In order to compare the eye gazes of each participant role between the two conditions (utterances with floor-hold and those with floor-switch), the gazing ratio was calculated for each condition. Figure 5.1 shows how the participants and utterances were classified for the target utterances. In the following sections, the gazing ratio calculated for the utterances with floor-hold is shortened to gazing ratio with floor-hold, while similarly the gazing ratio calculated for the utterances with floor-switch is termed gazing ratio with floor-switch.

To analyze the eye gaze activity of each role, we defined the gazing ratios:

$$Average\,of\;\,gazing\,ratios = \frac{1}{n}\sum_{i=1}^{n}\frac{DG_{jk_{(i)}}}{DU_{(i)}} * 100(\%), \tag{5.1}$$

where $DSU_{(i)}$ and $DG_{jk(i)}$ represent the duration of the i-th utterance and the duration of participant j gazing at participant k during that utterance, respectively.

5.2 Speaker's Eye Gaze

To test the first hypothesis, we calculated gazing ratios for CStoNS (eye gaze of current speaker toward next speaker) in both conditions of floor apportionment and in both languages. Figure 5.2 compares these gazing ratios and also shows the results of the gazing ratios for CStoOP as reference. As the figure shows, the current speaker gazes more at the next speaker than at the other participant in both conditions of floor apportionment in L2 conversations as well as in L1 conversations. There is a greater difference between gazing ratios with floor apportionment for CStoNS in L2 than in L1 conversations.

In order to test these hypotheses on the speaker's and listeners' eye gazes in L2 conversations, we conducted an ANOVA test to investigate the statistical differences between languages, conversation topics, gaze channels (gazer-target pairs), and floor apportionment conditions (floor-hold or floor-switch). The difference in gazing ratios with floor-switch between CStoNS and CStoOP (eye gaze of current speaker toward other participant) in L2 conversations is much larger than that in L1 conversations. These observations show that the speakers gaze more at a listener who will be the next speaker in utterances with floor-switch in L2 conversations than in L1 conversations.

5.3 Listener's Eye Gaze

We calculated gazing ratios for both listeners' gaze toward the speaker. Figure 5.3 shows gazing ratios with floor-hold and with floor-switch for NStoCS (eye gaze of next speaker toward current speaker) and OPtoCS (eye gaze of other participant toward current speaker). As shown in this figure, the gazing ratios with floor-switch for both NStoCS and OPtoCS are smaller than the gazing ratios with floor-hold for those gaze pairs. This result shows that both the next speaker and the other participant gaze at the current speaker significantly less in utterances with floor-switch than in utterances with floor-hold. The results also show that those differences of floor apportionment are larger in L2 conversations than in L1 conversations.

These results demonstrate that the listeners gaze less at the speaker in utterances with floor-switch than in utterances with floor-hold in both L1 and L2 conversations. This

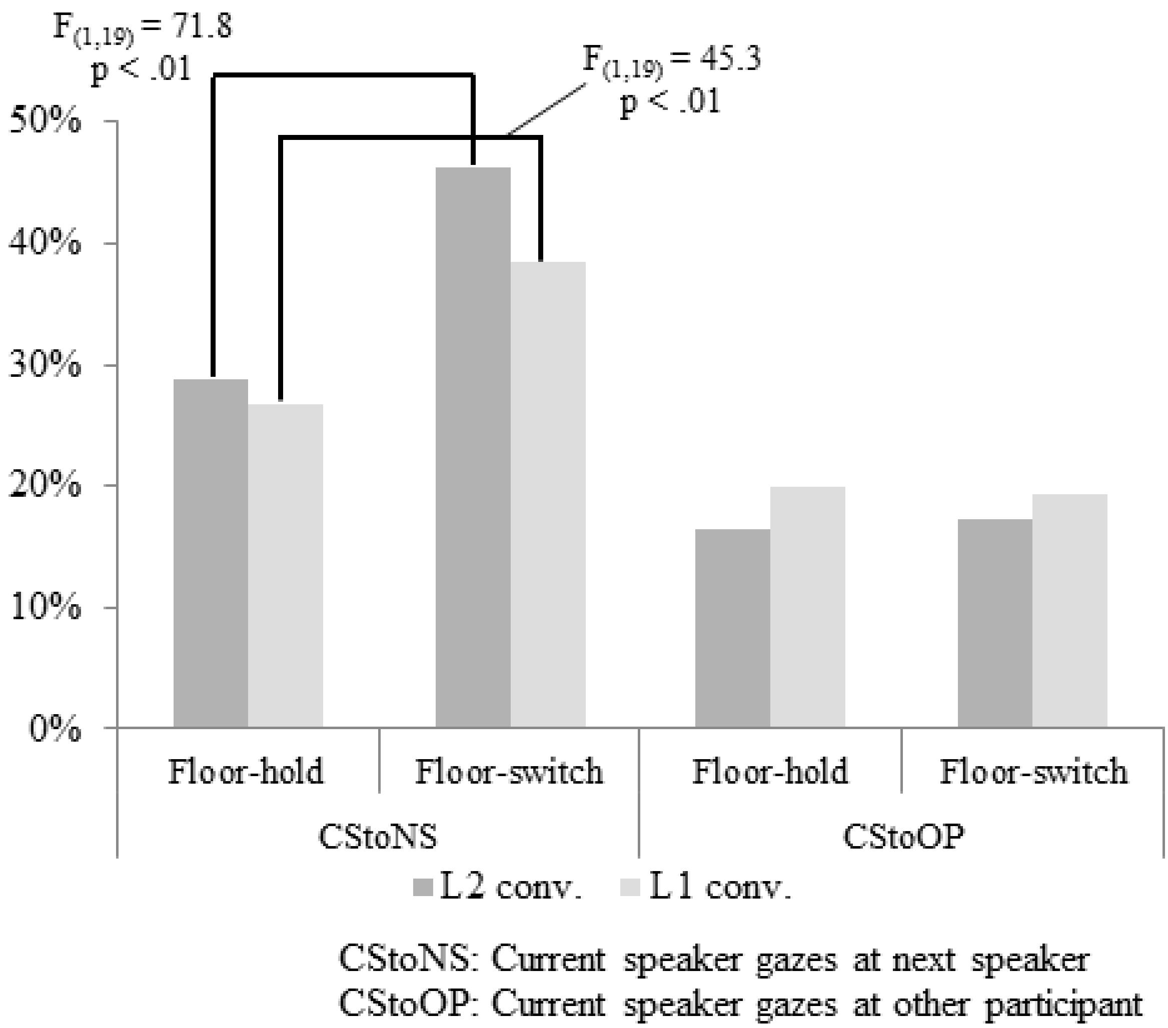

Figure 5.2 Averages of gazing ratios for the current speaker gazing to the next speaker (CStoNS) and to the other participant (CStoOP) in utterances with floor-hold and with floor-switch in L1 and L2 conversations

suggests the possibility that both listeners mainly gaze at the speaker while he/she holds the floor and they somehow notice the difference between utterances with floor-hold and with floor-switch.

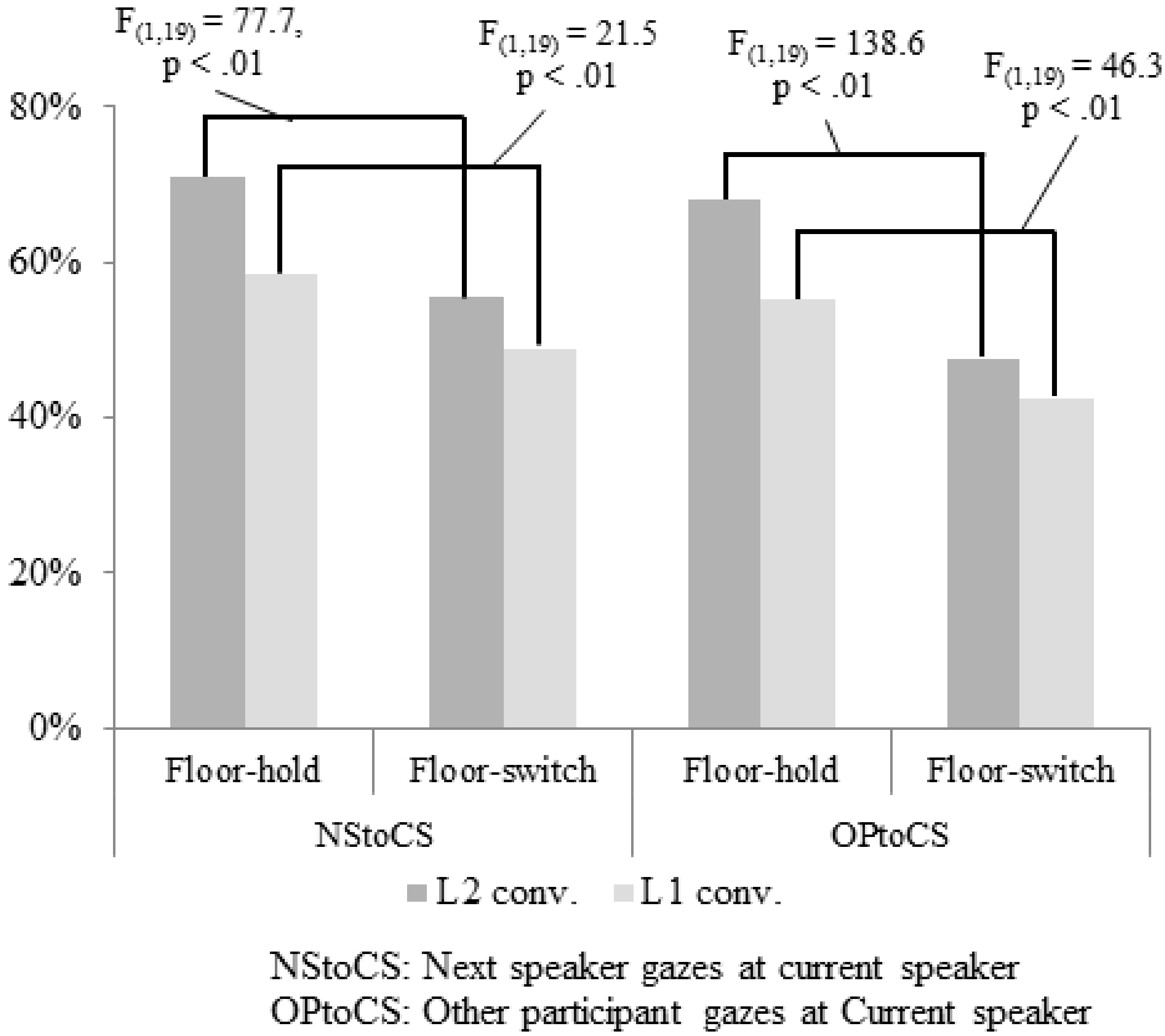

Figure 5.3 Averages of gazing ratios for the next speaker gazing to the current speaker (NStoCS) and for the other participant gazing to the current speaker (OPtoCS) in utterances with floor-hold and with floor-switch in L1 and L2 conversations

5.4 Discussion

Our analyses of the speaker's eye gaze for floor apportionment revealed the following observations:

(1) Concerning the speaker's eye gaze target, the speaker gazes more at the next speaker than at the other participant in both utterances with floor-switch and those with floor-hold in L2 conversations as well as in L1 conversations.

(2) Concerning the difference between the two conditions of floor apportionment, the speaker gazes more at the next speaker in utterances with floor-switch than in utterances with floor-hold in both L1 and L2 conversations.

(3) Concerning the difference between languages, the speaker gazes more in L2 than in L1 conversations at the next speaker in utterances with floor-switch, in contrast to those with floor-hold.

Observations (1) and (2) imply that the speaker directed his/her gaze to the person who will be the next speaker in both L1 and L2 conversations and that the speaker gazed more at the recipient during the utterance with floor-switch in almost the same manner as that used in L1 conversations. Observation (3) implies the probability that the gazed-at participant will be the next speaker is higher in L2 conversations than in L1 conversations. This shows that the speaker has a greater tendency to focus his/her overt attention to the next speaker in L2 than in L1 conversations.

Observation (1) suggests that the functionality of the speaker's eye gaze in floor apportionment is likely to be common to L1 and L2 conversations; however, the probability that the gazed-at participant will be the next speaker is higher in L2 than in L1 conversations. This further suggests that the speaker's eye gaze affects floor apportionment more significantly in L2 conversations than in L1 conversations, which supports our first hypothesis. Lower conversational competence in L2 might reinforce this interpretation because the lack of vocabulary, shortage of grammatical knowledge, and poor prosodic control may hinder L2 interlocutors in sending verbal signals for floor apportionment. The speaker's directing gaze does not automatically ensure who will actually speak next because additional interactive contingencies can intervene; however, the probability that the gazed-at participant, who perceives the speaker's eye gaze, will be the next speaker is higher in L2 than in L1 conversations.

As Lerner (2003) pointed out, to understand that he/she is the intended recipient, the gazed-at participant must notice the gaze, and others may also need to see this to comprehend that someone else has been indicated as the intended recipient. In L2 conversations in which the speaker proactively directs his/her gaze to the intended recipient, it is expected that the listeners need to direct their gaze to the speaker in order not to miss the visual signals of floor apportionment given by the speaker. There is a possibility that the signal of floor apportionment from the speaker might be shown by his/her eye gaze more explicitly in L2 conversations than in L1 conversations; at least the statistical results show that the gazing ratio of the speaker toward the next speaker is larger in L2 conversations than in L1 conversations.

The analysis results of listeners' eye gazes in the previous section demonstrated the following observations:

(4) Concerning the difference between the two conditions of floor apportionment, both the next speaker and the other participant gaze less at the speaker in utterances with floor-switch than in those with floor-hold in both L1 and L2 conversations.

(5) Concerning the difference between languages, the differences of gazing ratios for NStoCS between utterances with floor-hold and with floor-switch are larger in L2 than in L1 conversations.

The observation of the listeners' eye gaze (4) suggests the possibility that listeners notice the utterances with floor-switch and decrease their gaze attention to the speaker after noticing the point of floor switch. This result is consistent with the finding of Kendon [16] that the next speaker reduces the amount of gaze toward the current speaker in order to prepare an utterance in dyadic conversations. This suggests that the next speakers need more concentration to prepare for the next utterance and to reduce the visual attention to the speaker due to the lack of L2 proficiency.

The other observation (5) shows that the decrease in the next speaker's eye gaze is larger in L2 than in L1 conversations. This larger decrease may lead to a higher chance that the speaker notices this change more easily in L2 conversations than in L1 conversations.

5.5 Summary

These findings confirm our first hypothesis that the speaker's eye gaze affects floor apportionment more significantly in L2 than in L1 conversations, since the gazed-at participant in utterances with floor-switch will be the next speaker at a higher probability in L2 conversations than in L1 conversations. The findings are also consistent with our second hypothesis that longer eye gazes by listeners in L2 conversations work efficiently for smooth floor apportionment, although we may not yet be able to verify our second hypothesis to a sufficient degree. The larger decrease in the next speaker's eye gaze in utterances with floor-switch compared to those with floor-hold may lead to a higher possibility that the speaker notices this change more easily in L2 than in L1 conversations.

Chapter 6

Effect of Speaker's Eye Gaze for Floor Apportionment

Our previous analyses revealed that the speaker's eye-gaze activities perform more effectively for assigning the next speaker in L2 than in L1 conversations.

As a result, it is not yet clarified why the speaker's eye-gaze activities function more effectively as a signal in L2 and how this finding relates to previous findings that listeners gaze more at the speaker in L2.

In order to explore this question, we compared the speaker's gaze activities in utterances with floor-hold and floor-switch to two types of listeners, mainly-gazed-at listeners and not-mainly-gazed-at listeners, as well as their eye gaze activities to the speaker; this distinction was used to evaluate how differently the two listeners' gazing ratios related to the ratios of taking the floor. We analyzed the correlation between the two listeners' gazing ratios toward the speaker to investigate how the gaze activities of the two listener types interact with each other in triad conversations.

6.1 Classification of the Listener with Speaker's Eye Gaze

We classified the listeners into two groups according to their being targets of the speaker's eye gaze: "mainly-gazed-at listener," who is gazed at more by the speaker during the utterance than the other listener, and "not-mainly-gazed-at listener." The speaker's gazing ratio was used to classify the listeners. The listener who was the gaze target with the higher gazing ratios was classified as the mainly-gazed-at listener, and the other listener was classified as the not-mainly-gazed-at listener. We used the average gazing ratios of these participants to analyze the effect of gaze activities. The gazing ratio is used as a shorthand notation of the average of the gazing ratios among the three participants in the following sections. Here, SPtoGL is used as the speaker's (SP) eye gaze toward the mainly-gazed-at listener (GL), while GLtoSP and NGLtoSP are used as the mainly-gazed-at listener's and the not-mainly-gazed-at listener's (NGL) eye gazes toward the speaker. To grasp the big picture of the participants' eye-gaze activities and their effect on floor apportionment, we calculated the average gazing ratios of SPtoGL, GLtoSP, and NGLtoSP in L1 and L2 conversations (Figure 6.1).

6.2 Speaker's Gazing Ratios toward Mainly-Gazed-At Listeners and Ratios of their Taking the Floor

We calculated the gazing ratios for SPtoGL and SPtoNGL in order to compare how the speaker gazes at the listeners between L1 and L2 conversations. As shown in Table 6.1, the results show that the gazing ratios for SPtoGL are almost the same for utterances with floor-hold in L1 and L2 conversations, but they are larger for utterances with floor-switch in L2 than in L1 conversations. To analyze the difference in gazing ratios for SPtoGL

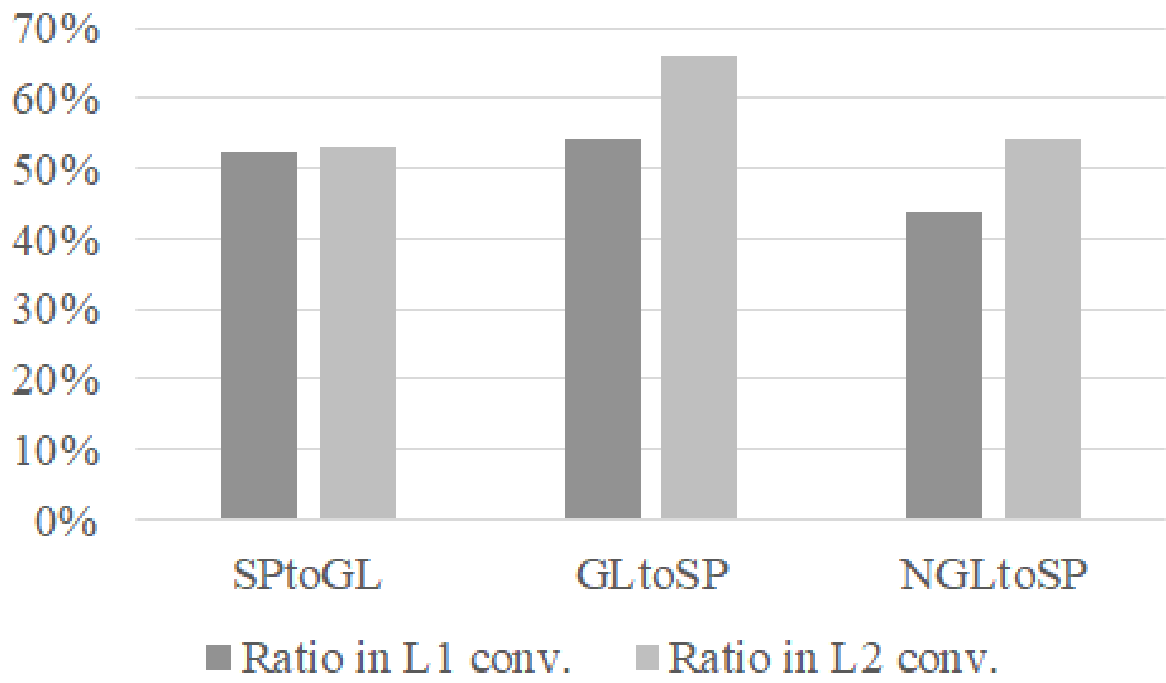

Figure 6.1 Basic statistics of gazing ratios of SPtoGL, GLtoSP, and NGLtoSP in L1 and L2 conversations. Note that SPtoGL is used as the speaker's (SP) eye gaze toward the mainly-gazed-at listener (GL), while GLtoSP and NGLtoSP are used as the mainly-gazed-at listener's and the not-mainly-gazed-at listener's (NGL) eye gazes toward the speaker.

between L1 and L2 conversations, we conducted an ANOVA test with language difference, topic difference, and floor apportionment (floor hold and floor switch) difference as within-subject factors. The ANOVA test shows significant main effects of floor apportionment difference ($F_{(1,19)} = 79.4, p < .01$) and significant interaction of language difference and floor apportionment difference ($F_{(1,19)} = 9.41, p < .01$). To verify the effects of speaker's eye gaze on floor apportionment, we calculated the ratios of how often the mainly-gazed-at participant takes the floor. Figure 6.2 represents the ratios at which the listener takes the floor in L1 and L2 conversations.

The ANOVA on the ratios of the mainly-gazed-at listener taking the floor was conducted with language difference, topic difference, and floor apportionment difference as within-

Table 6.1 Basic statistics of gazing ratios of SPtoGL, GLtoSP, and NGLtoSP in L1 and L2 conversations for utterances with both floor-hold (FH) and floor-switch (FS)

	Ratio in L1 conv. All utterances (FH, FS)	Ratio in L2 conv. All utterances (FH, FS)
SPtoGL	52.5% (44.5%, 55.0%)	53.3% (44.1%, 62.3%)
SPtoNGL	2.7% (2.2%, 2.8%)	1.3% (1.2%, 1.4%)
GLtoSP	54.1% (60.6%, 51.9%)	66.0% (74.3%, 58.8%)
NGLtoSP	43.9% (53.0%, 39.4%)	54.2% (64.7%, 44.4%)

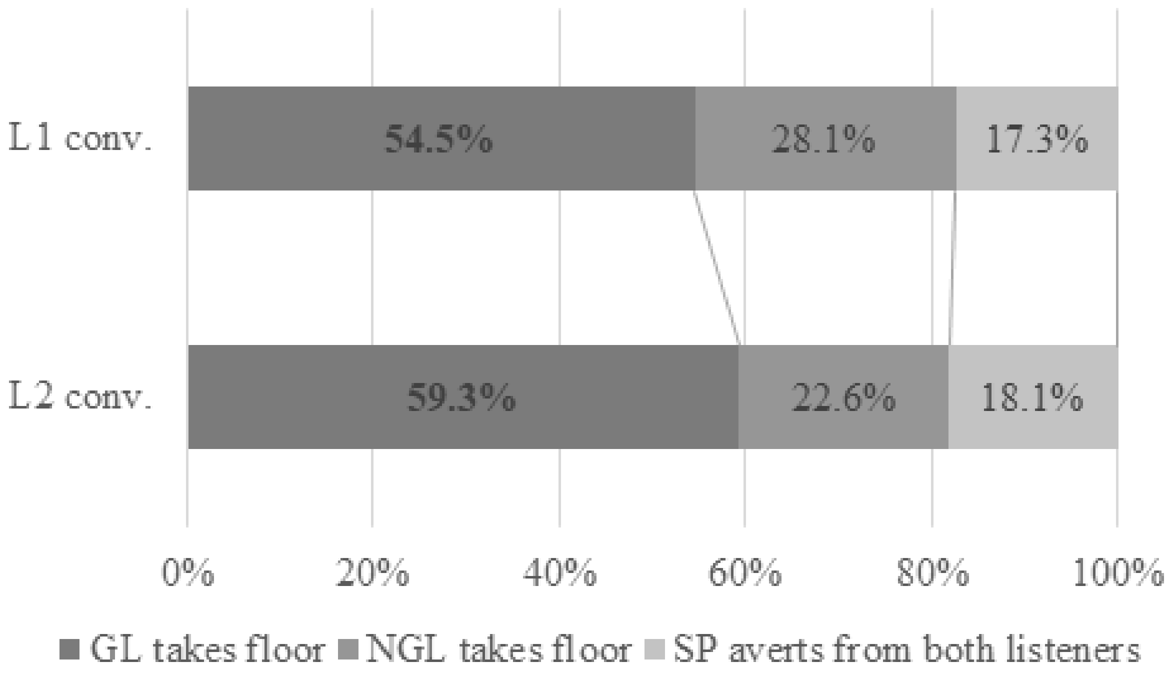

Figure 6.2 Ratios at which listeners take the floor in L1 and L2 conversations. Note that the utterances in which the speaker does not gaze at any listener are classified as "SP averts gaze from both listeners."

subject factors. The results show a significant main effect of language difference ($F_{(1,19)} = 7.2, p < .05$). The statistical results revealed that the mainly-gazed-at listener takes the floor more often in L2 conversations than in L1 conversations, perhaps because of the larger speaker's gazing ratio in floor-switch utterances in L2 conversations, whereas the duration of the speaker's gaze at the mainly-gazed-at listener is almost the same in floor-hold utterances in both L1 and L2 conversations.

Then, we calculated the relationship between gazing ratio for SPtoGL with floor-switch and ratios of GL taking the floor (Figure 6.3). Correlation analyses were done for each conversation, and the results show that there is a significant positive correlation in L2 conversations ($r = .55, p < .01$), whereas this was not shown in L1 conversations ($r = .26, p = .11$). These results suggest that the speaker's eye-gaze activities strongly affect floor apportionment in L2 conversations as a way to compensate for their low linguistic proficiency, which creates difficulties in producing and understanding their utterances.

6.3 Two Listeners' gazing ratios toward Speaker in response to Speaker's Gaze Activities

To analyze the difference in gazing ratios for GLtoSP and NGLtoSP between the floor-hold and floor-switch utterances shown in Figure 6.4, we conducted an ANOVA test with language difference, topic difference, and floor apportionment difference as within-subject factors. The ANOVA test shows significant main effects of language difference ($F_{(1,19)} = 42.0, p < .01$) and floor apportionment difference ($F_{(1,19)} = 83.0, p < .01$) for GLtoSP as well as significant main effects of language difference ($F_{(1,19)} = 30., p < .01$) and floor apportionment difference ($F_{(1,19)} = 142.8, p < .01$) for NGLtoSP. These results suggest that the gazing activities of both listeners change depending on the change of the speaker's gazing activities between floor-hold and floor-switch utterances.

We conducted correlation analysis between the gazing ratios for GLtoSP and the ratios of the mainly-gazed-at listener taking the floor. Figure 6.5 is a scatter plot of those ratios for each conversation. The results show that there is a significant positive correlation

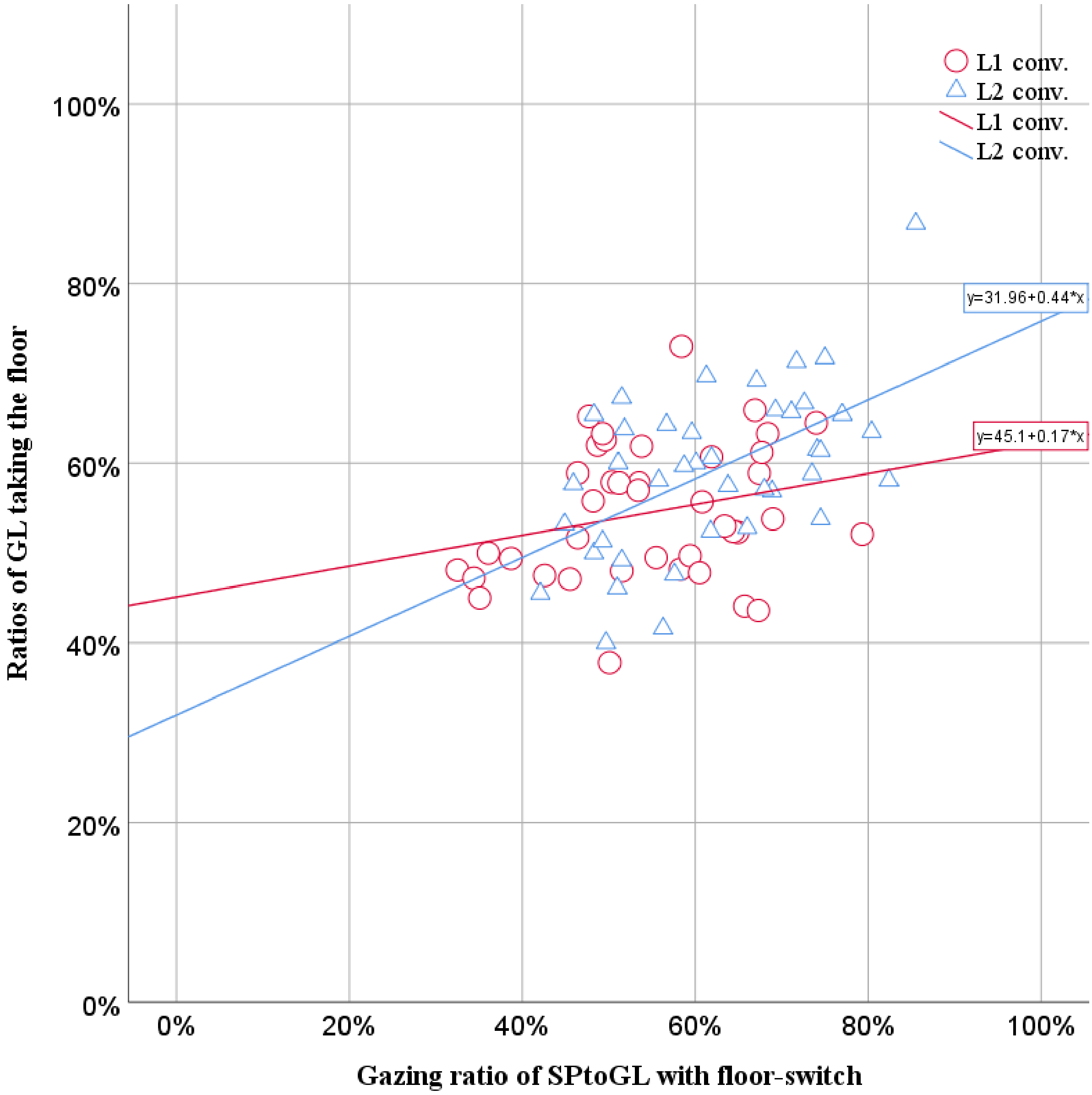

Figure 6.3 Scatter plots of gazing ratio of SPtoGL with floor-switch and ratio of GL taking the floor in L1 and L2 conversations. Solid lines indicate the regression line for each language.

($r = .62, p < .01$). Consequently, there is a possibility that the more the mainly-gazed-at listener gazes at the speaker, the more often she/he takes the floor, and the gazing ratio for GLtoSP may indicate her/his intention of taking the floor. On the other hand, there is almost no correlation between the gazing ratios for NGLtoSP and the ratios of the

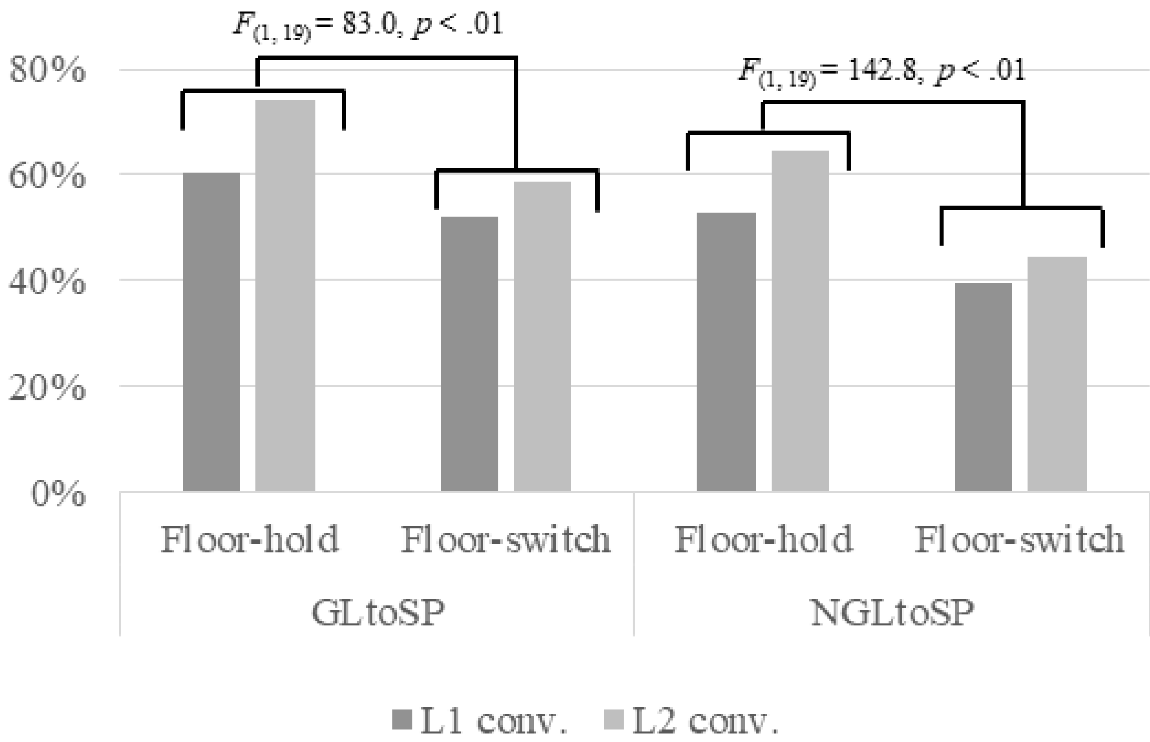

Figure 6.4 Gazing ratios of GLtoSP and NGLtoSP in utterances with floor-hold and floor-switch in L1 and L2 conversations

not-mainly-gazed-at listener taking the floor, as shown in Figure 6.6. This result suggests that the gazing ratios for NGLtoSP are independent of the intention to take the floor and that other factors may affect the gazing ratios.

The statistical results show that the proportion between GLtoSP and NGLtoSP ratios is similar in L1 and L2 conversations. To verify the detailed proportion between the listeners' eye-gaze activities toward a speaker, we analyzed the correlation of eye-gaze activities between the listeners in both L1 and L2 conversations. Figure 6.7 shows a scatter plot of the GLtoSP and NGLtoSP ratios.

The data points represent the gazing ratios obtained in 20 conversations, in both L1 and L2, and the solid line shows the regression line. The results show that there is a significant positive correlation between the ratios for GLtoSP and NGLtoSP ($r = .80, p < .01$). This

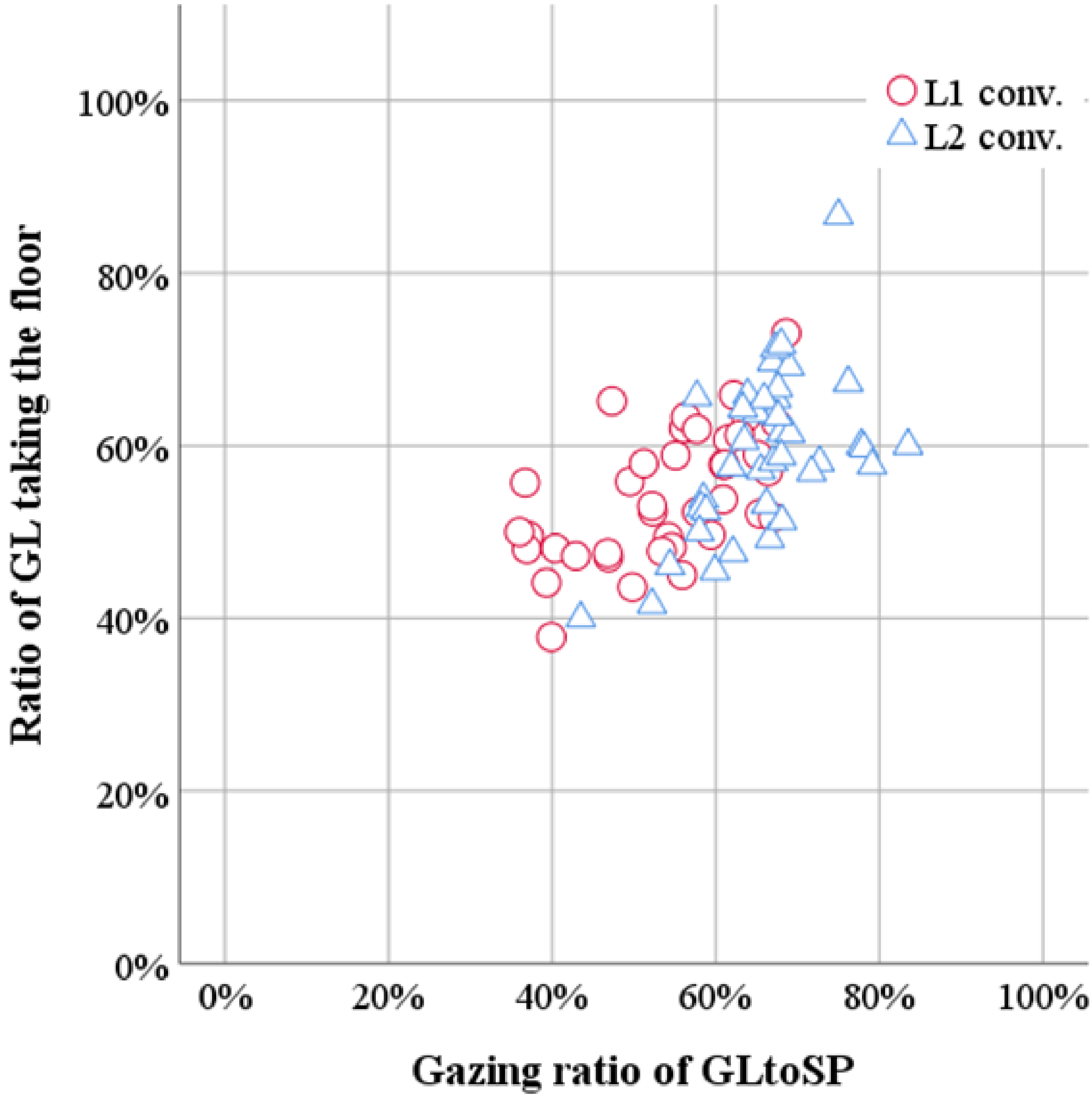

Figure 6.5 Scatter plot of gazing ratio of GLtoSP and ratios of GL taking the floor in human-human L1 and L2 conversations

shows that the more one of the listeners gazes at the speaker, the more the other listener also gazes at the speaker, in both L1 and L2 conversations.

To analyze the difference in the correlation between the ratios of GLtoSP and NGLtoSP owing to language difference, we compared the correlation coefficients of GLtoSP and NGLtoSP between L1 and L2 conversations. We conducted Fisher's z-transformation to

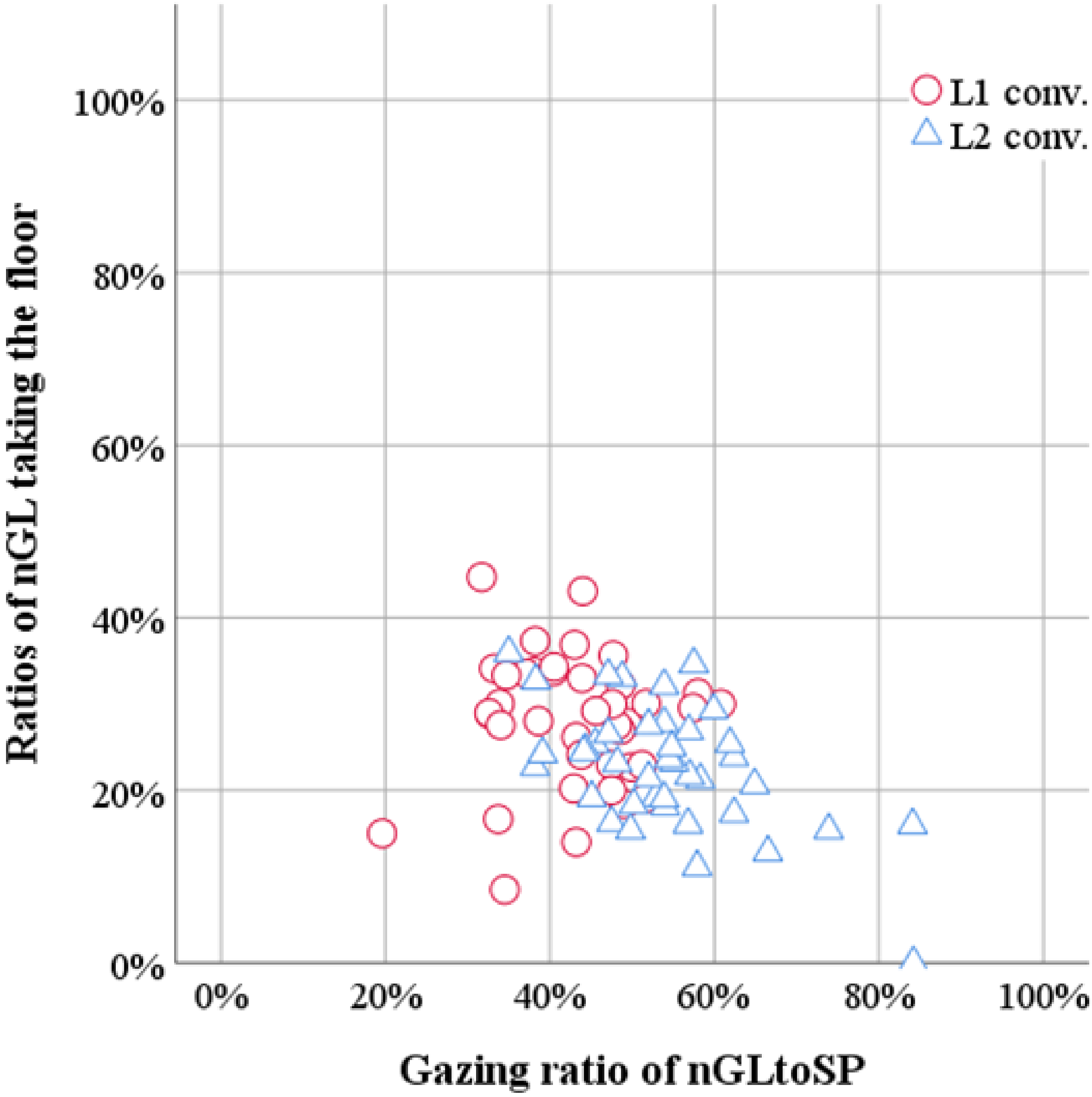

Figure 6.6 Scatter plot of gazing ratio of NGLtoSP and ratios of NGL taking the floor in human-human L1 and L2 conversations

those correlation coefficients to confirm whether there is a difference. The results show that there is no significant difference between the correlation of the listener's eye gaze between L1 and L2 conversations (p = .40).

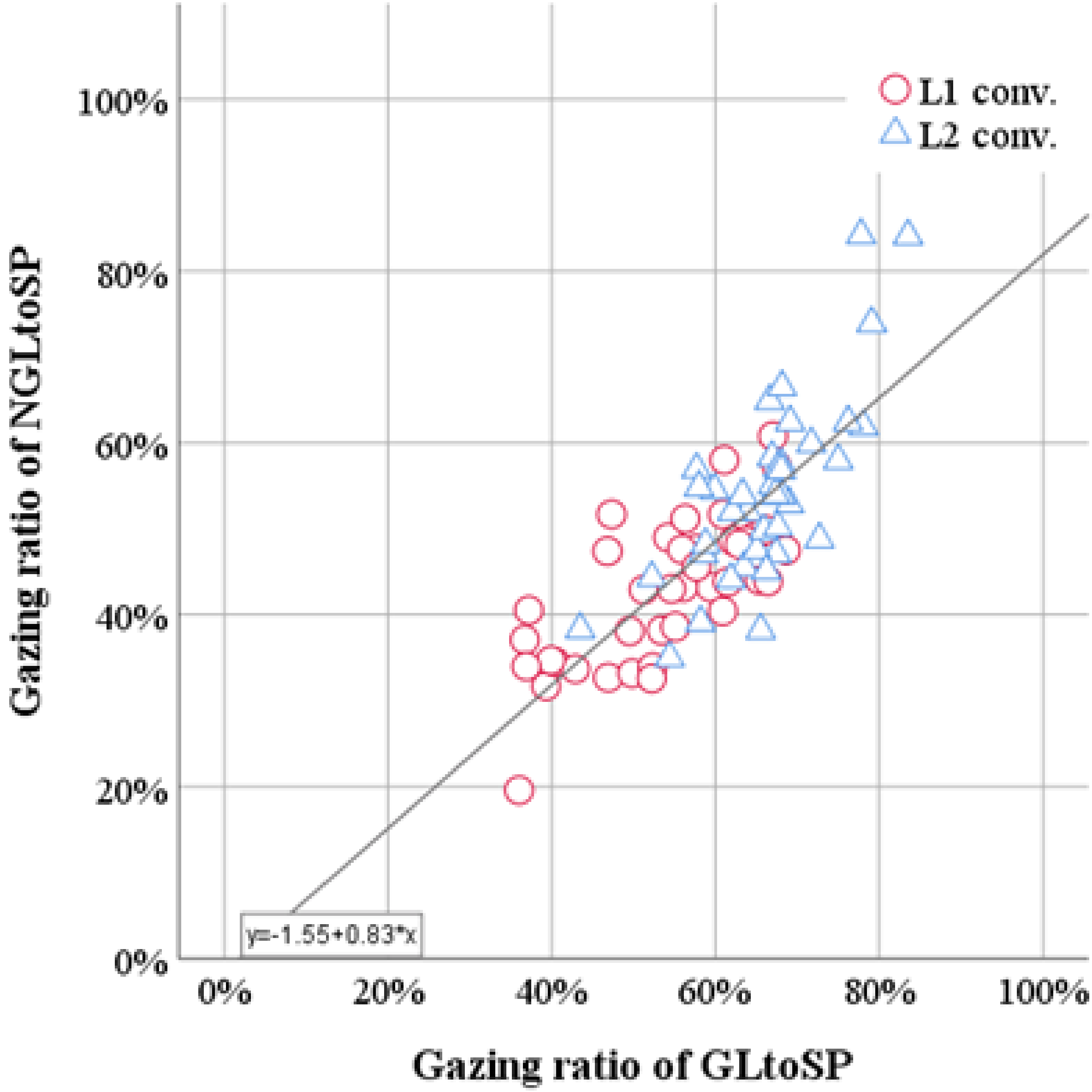

Figure 6.7 Scatter plot for gazing ratios of GLtoSP and NGLtoSP in L1 and L2 conversations. Note that the solid line shows a regression line.

6.4 Discussion

The ANOVA results demonstrated the following observations:

(1) The speaker's gazing ratio to the mainly-gazed-at listener increases in utterances with floor-switch compared to those with floor-hold. The increase in the speaker's gazing ratio in floor-switch, which is assumed to be a signal for floor apportionment, is larger in L2 than in L1.

(2) The mainly-gazed-at listener gazes more at the speaker corresponding to the larger speaker's gazing ratio in utterances with floor-switch in L2 conversations than in L1 conversations and takes the floor more often in L2 conversations than in L1 conversations.

(3) The more the mainly-gazed-at listener gazes at the speaker, the more often she/he takes the floor; in contrast, no positive correlation is shown with the not-mainly-gazed-at listener.

(4) There is a significant positive correlation between the GLtoSP and NGLtoSP ratios in both L1 and L2 conversations, and the correlation coefficient of the ratios are almost the same in L1 and L2 conversations.

Result (1) suggests that the speaker's eye-gaze activity has more effect in setting the mainly-gazed-at listener as the next speaker in L2 conversations, owing to a greater increase of gazing activities to the mainly-gazed-at listener. The results also show that there is a positive correlation between the gazing ratio of SPtoGL in floor switch and the ratios of GL taking the floor in L2 conversations, whereas there is no significant correlation in L1 conversations. This observation also supports the view that the effect of the speaker's eye-gaze activities in L2 conversations is strong for floor apportionment in order to determine the next speaker.

Result (2) suggests that this stronger effect of the speaker's eye-gaze activities for determining the next speaker in L2 conversations may lead to a higher gazing ratio of the mainly-gazed-at listener toward the speaker in L2 conversations than in L1 conversations. Furthermore, this might be caused by the low proficiency of L2 listeners, where they tried to grasp the information for floor apportionment given by the speaker, especially for the mainly-gazed-at listener who received the eye-gaze signal from the speaker.

Result (3) shows that there is a positive correlation between the gazing ratios of GLtoSP and the ratios of the mainly-gazed-at listeners taking the floor in both languages. This result suggests that the listeners show their intention to take the floor with their eye-gaze activities toward the speaker. There is a possibility that the listeners gaze more at the speaker to show their intention in order to compensate for their lower proficiencies in L2 conversations than in L1 conversations. Result (3) also shows that the gazing activities of the not-mainly-gazed-at listener are different from those of the mainly-gazed-at listener whose ratio of taking the floor is correlated with the gazing ratio toward the speaker, and another factor is assumed to affect the gazing activities of the not-mainly-gazed-at listener.

Result (4) shows a significant positive correlation between the ratios of GLtoSP and NGLtoSP in triad conversations. Furthermore, there is no significant difference between the positive correlation values in L1 and L2 conversations. This suggests that the listeners' eye-gaze activities toward the speaker affect each other independently of the proficiency of the participants. A possible cause of this phenomenon is the joint attention of listeners that occurs when two people share an interest in an object or event and there is an understanding between these two people that they are both interested in the same object or event.

The experimental results show that when the mainly-gazed-at listener gazes at the speaker more in the utterances with floor-hold, the not-mainly-gazed-at listener also gazes at the speaker more. On the other hand, when the conversation moves from utterances with floor-hold to those with floor-switch, the target of focus of the conversations shifts from the speaker to the mainly-gazed-at listener, and both the mainly-gazed-at participant and the not-mainly-gazed-at participant gaze at the speaker to a lesser extent. The speaker and the mainly-gazed-at listener are expected to be the main players in triad conversations, and the attention of participants shifts from the speaker in utterances with floor-hold to the mainly-gazed-at listener in utterances with floor-switch. These observations suggest that the not-mainly-gazed-at listener gazes at the speaker to maintain joint attention with the mainly-gazed-at listener. A previous study arguing the importance of joint attention showed that it played the dominant role in making coherent discourse for sharing the location to which the participants are paying attention [63]. Previous research on human-robot communications suggested that joint attention occurred when the participants shared interest in the gazed-at object [64].

The statistical results show that the strong effects of the speaker's eye gaze on floor apportionment in L2 conversations lead the mainly-gazed-at listeners to gaze at the speaker in order to grasp the information from the visual input due to their low proficiency in L2. The statistical results also suggest that the effect of joint attention might increase the durations of not-mainly-gazed-at listeners' eye gazes toward the speaker when the mainly-gazed-at listeners gaze at the speaker in both L1 and L2 conversations. Those two effects, the strong effect of the speaker's eye-gaze activities on floor apportionment and the joint attention between listeners, are the reasons why the listener's gazing ratios are higher in L2 conversations than in L1 conversations; at least the statistical results support this possibility.

6.5 Summary

These results suggest that the mainly-gazed-at listener gazes more at the speaker corresponding to the larger speaker's gazing ratio in utterances with floor-switch in L2 conversations, and the higher ratio of the mainly-gazed-at listener toward a speaker in L2 conversations could lead to a similarly higher ratio of the not-mainly-gazed-at listener toward a speaker in L2 conversations. Various factors may increase the listener's gazing ratio in L2 conversations, such as showing polite acknowledgement of the effect made by the speakers of lower proficiency through eye gaze. However, from these findings, the stronger effect of the speaker's eye gaze for floor apportionment in L2 conversations and joint attention between listeners may be the main reasons why the listener's gazing ratio is higher in L2 than in L1 conversations; at least the statistical results support this possibility.

Chapter 7

Overall Discussion

The analysis results revealed a difference in the importance of eye gaze activities of both speaker and listener in L1 and L2 conversations. In the following, we present an overall discussion on the speakers' and listeners' eye-gaze activities one by one.

The speakers' gaze durations toward listeners are almost the same in L1 and L2 conversations. However, when it comes to the utterances just before floor switch, speakers gaze more at the listener who is to be the next speaker, and the durations of these gazes are larger in L2 conversations than in L1 conversations. The results also show that the longer a speaker gazes at a listener, the more the listener takes the floor in L2 conversations, whereas that tendency was not observed in L1 conversations. Those results reveal that speakers use their eye gaze activities more efficiently for floor apportionment in L2 conversations than in L1 conversations.

Listeners gaze more at the speaker in L2 conversations than in L1 conversations in any conversational condition. The listener who is to be the next speaker shifts his/her gaze away during utterances in which the speaker gives the floor to him/her, compared with other utterances in both L1 and L2 conversations. The listener begins to think of the next utterance while shutting out visual information in order to concentrate, which is consistent with the results obtained in the previous studies. The statistical results also show that the longer the listener who is gazed at by the speaker in return gazes at the speaker, the more often that listener takes the floor. The results also show that when the listener gazes away from the speaker, even if he/she is gazed at by the speaker, that listener less often takes the floor. This suggests that even if the speaker tries to compel the gaze-target listener to take the floor, that listener avoids taking the floor by gazing away from the speaker. These observations suggest that the listeners show their intention to take the floor with their eye-gaze activities toward the speaker, and there is a higher chance of mutual gaze being constructed between speaker and listener so that they might confirm with each other their agreement on floor apportionment, and since this is shown in both L1 and L2 conversations, this phenomenon is independent of linguistic proficiency.

From these findings, eye-gaze activities function efficiently to establish floor apportionment in conversations under the influence of communicative insufficiency. These statistical findings can be used to predict the participants' intentions despite whether the speaker wants to give or hold the floor. Furthermore, they can be applied to the eye-gaze model for understanding the interlocutor's intention to speak or not speak. They can also be applied to create behavior models to control a robot's eye gaze in order to establish the flow of smooth conversations.

Chapter 8

Conclusion and Future Work

This dissertation presented eye gaze analyses under the influence of communicative insufficiency through comparing the eye gaze activities in L1 and L2 conversations. There are massive differences of linguistic proficiency between native-language and second-language speakers, and thus there is a highly possibility that the interlocutor's eye gaze activities compensate for their low linguistic proficiency in L2 conversations. A multimodal conversational corpus was created to conduct quantitative and comparative analyses of eye-gaze activities in L1 and L2 conversations. This thesis shows the results of three analyses: eye-gaze activities for all utterances to capture the general difference between L1 and L2 conversations, eye gaze activities for floor apportionment, and the effect of the speaker's eye gaze activities on floor apportionment.

The results revealed that the listeners gaze more at the speaker in L2 conversations than in L1 conversations, whereas the speaker gazes at the listeners with almost the same duration in L1 and L2 conversations. Detailed analyses from the perspective of floor apportionment found that the speaker gazed more at the listener who is to be the next speaker in L2 conversations than in L1 conversations. The results also show that the more a speaker gazes at the listener, the more that listener takes the next turn in

L2 conversations. Those results reveal that the strong effects of the speaker's eye gaze on floor apportionment in L2 conversations lead the listener, who was gazed at by the speaker, to gaze at the speaker in order to grasp information from the visual input due to their low proficiency in L2. Such listener's eye gaze toward the speaker leads the other listener, who was not gazed at by the speaker, to gaze at the speaker in both L1 and L2 conversations. Those two effects, i.e., the strong effect of the speaker's eye-gaze activities on floor apportionment and the joint attention between listeners, are the main reasons why the listener's gazing ratios are higher in L2 conversations than in L1 conversations; at least the statistical results support this possibility.

This dissertation presents the finding that eye-gaze activity plays a more important role when the conversational participants have low levels of communicative efficiency. These observations can be adapted to a model for natural gaze patterns to improve a robot's engagement and interaction capabilities by estimating the participant's intention from his/her eye gaze activities and coordinating conversations with the robot's eye-gaze activities. Our research group started to collect human-robot conversations in order to monitor the participants' eye gaze activities in those situations. It seemed that the participants used their eye-gaze activities in a way quite similar to that in human-human conversations, although this work is in progress and thus we need more data and analyses. We might also need a deep analysis of the timing of eye-gaze activities, since this dissertation focused on the duration of gaze, to create a better eye-gazing model for robots.

Acknowledgment

First and foremost, I would like to express my deepest gratitude to my excellent advisor Professor Seiichi Yamamoto, for his encouragement, selfless support and continual guidance throughout my Ph.D. His insights, meticulous reading, editing of this dissertation and every other publication resulting from this research, have definitely improved the quality of my work. He always provides me freedom to investigate and explore this field, guides my future directions. Without my dear advisor, I might not be inspired with this study, not enjoy it and be excited for this research.

I must thank Dr. Ichiro Umata and Prof. Tsuneo Kato for their encouragement and help on this research. Their devoted help and advise improved the quality of my work.

I also thank Prof. Kristiina Jokien and Prof. Graham Wilcock for their advises for this research. Their kindness comments improve this research. Prof. Jokinen also gave me an opportunity to attend the internship to do the research in AIST. This wonderful time broaden my viewpoint.

Then, I would like to thank all of my friends and research colleagues who helped me with collecting conversational corpus.

Last, I would like to express my deep gratitude to my family for enormous supports and encouragement. Without their support, I will not have been free to chase my dream until now.

Bibliography

[1] E.C. Hanson, “A history of international communication studies,” 11 2017.

[2] Ministry of Economy, Trade and Industry, Japan, White Paper on International Economy and Trade, Japanese Government, 2018.

[3] Ministry of Education, Culture, Sports, Science and Technology, Japan, Research on Interaction of researchers, Japanese Government, 2018.

[4] R.E. Grinter, J.D. Herbsleb, and D.E. Perry, “The geography of coordination: Dealing with distance in r&d work,” Proceedings of the international ACM SIGGROUP conference on Supporting group work, pp.306–315, ACM, 1999.

[5] C.D. Cramton, “The mutual knowledge problem and its consequences for dispersed collaboration,” Organization science, vol.12, no.3, pp.346–371, 2001.

[6] T. Beyene, J. Pamela, P.J. Hinds, and C.D. Cramton, Walking through jelly: language proficiency, emotions, and disrupted collaboration in global work, SSRN eLibrary, 2009.

[7] I. Aaltonen, A. Arvola, P. Heikkilä, and H. Lammi, “Hello pepper, may i tickle you?: Children’s and adults’ responses to an entertainment robot at a shopping mall,” Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction, HRI ’17, New York, NY, USA, pp.53–54, ACM, 2017.

[8] E. Gilmartin, M. Collery, K. Su, Y. Huang, C. Elias, B.R. Cowan, and N. Campbell, “Social talk: Making conversation with people and machine,” Proceedings of the 1st ACM SIGCHI International Workshop on Investigating Social Interactions with

Artificial Agents, ISIAA 2017, New York, NY, USA, pp.31–32, ACM, 2017.

[9] Y. Yoshikawa, K. Shinozawa, H. Ishiguro, N. Hagita, and T. Miyamoto, "Responsive robot gaze to interaction partner.," Robotics: Science and systems, 2006.

[10] J.D. Boucher, U. Pattacini, A. Lelong, G. Bailly, F. Elisei, S. Fagel, P.F. Dominey, and J. Ventre-Dominey, "I reach faster when i see you look: gaze effects in human-human and human-robot face-to-face cooperation," Frontiers in neurorobotics, vol.6, p.3, 2012.

[11] B. Mutlu, T. Shiwa, T. Kanda, H. Ishiguro, and N. Hagita, "Footing in human-robot conversations: how robots might shape participant roles using gaze cues," Proceedings of the 4th ACM/IEEE international conference on Human robot interaction, pp.61–68, ACM, 2009.

[12] H. Sacks, E.A. Schegloff, and G. Jefferson, "A simplest systematics for the organization of turn taking for conversation," in Studies in the organization of conversational interaction, pp.7–55, Elsevier, 1978.

[13] A. Pentland, "Socially aware, computation and communication," Computer, vol.38, no.3, pp.33–40, 2005.

[14] D. Gatica-Perez, "Automatic nonverbal analysis of social interaction in small groups: A review," Image and Vision Computing, vol.27, no.12, pp.1775–1787, 2009.

[15] A. Vinciarelli, M. Pantic, and H. Bourlard, "Social signal processing: Survey of an emerging domain," Image and vision computing, vol.27, no.12, pp.1743–1759, 2009.

[16] A. Kendon, "Some functions of gaze-direction in social interaction," Acta Psychologica, vol.26, pp.22–63, 1967.

[17] S. Duncan, "Some signals and rules for taking speaking turns in conversations," Journal of Personality and Social Psychology, vol.23, pp.283–292, 1972.

[18] S. Duncan, "Towards a dyadic grammar for conversation," Semiotica, vol.1, pp.29–47, 1973.

[19] C. Goodwin, Conversational Organization: Interaction between Speakers and Hearers, New York: Academic Press, 1981.

[20] G.W. Beattie, "The regulation of speaker turns in ace-to-face conversation: Some

implications for conversation in sound-only channels," Semiotica, vol.34, pp.55–70, 1981.

[21] M. Argyle and J. Dean, "Eye-contact, distance and affiliation," Sociometry, pp.289–304, 1965.

[22] E. Boyle, A. Anderson, and A. Newlands, "The effect of visibility on dialogue and performance in a cooperative problem solving task," Language and Speech Journal, pp.1–20, 1994.

[23] H.H. Clark and M.A. Krych, "Speaking while monitoring addressees for understanding," Journal of Memory and Language, pp.62–81, 2004.

[24] Y. Hosoda, "Repair and relevance of differential language expertise in second language conversations," Applied Linguistics, pp.25–50, 2006.

[25] E. Veinott, J. Olson, G. Olson, and X. Fu, "Video helps remote work: Speakers who need to negotiate common ground benefit from seeing each other," Proceedings of the Conference on Computer Human Interaction. CHI'99, ACM Press, PA, USA, pp.302–309, 1999.

[26] E. Goffman, "Replies and responses," Language in Society, vol.5, pp.257–313, 1976.

[27] A. Mehrabian and M. Wiener, "Decoding of inconsistent communications," Journal of Personality and Social Psychology, vol.6, no.1, pp.109–114, 1967.

[28] A. Mehrabian and S.R. Ferris, "Inference of attitudes from nonverbal communication in two channels," Journal of Consulting Psychology, vol.31, no.3, pp.248–252, 1967.

[29] H.H. Clark, Using Language, Cambridge: Cambridge University Press, 1996.

[30] H. Kobayashi and S. Kohshima, "Unique morphology of the human eye," Nature, vol.387, no.6635, p.767, 1997.

[31] H. Kobayashi and S. Kohshima, "Unique morphology of the human eye and its adaptive meaning: comparative studies on external morphology of the primate eye," Journal of human evolution, vol.40, no.5, pp.419–435, 2001.

[32] M.S. Gobel, H.S. Kim, and D.C. Richardson, "The dual function of social gaze," Cognition, vol.136, pp.359–364, 2015.

[33] J.B. Bavelas, L. Coates, and T. Johnson, “Listener responses as a collaborative process: The role of gaze,” Journal of Communication, vol.52, no.3, pp.566–580, 2002.

[34] R. Vertegaal, R. Slagter, G. Veer, and A. Nijholt, “Eye gaze patterns in conversations: there is more to conversational agents than meets the eyes,” CHI ’01 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp.301–308, 2001.

[35] S. Baron-Cohen, Mindblindness: An essay on autism and theory of mind, MIT press, 1997.

[36] A. Mehrabian, “Inference of attitudes from the posture, orientation, and distance of a communicator.,” Journal of consulting and clinical psychology, vol.32, no.3, pp.296–308, 1968.

[37] C.L. Kleinke, R.A. Staneski, and D.E. Berger, “Evaluation of an interviewer as a function of interviewer gaze, reinforcement of subject gaze, and interviewer attractiveness.,” Journal of Personality and Social Psychology, vol.31, no.1, pp.115–122, 1975.

[38] R.W. Wheeler, J.C. Baron, S. Michell, and H.J. Ginsburg, “Eye contact and the perception of intelligence,” Bulletin of the Psychonomic Society, vol.13, no.2, pp.101–102, 1979.

[39] S. Duncan Jr and G. Niederehe, “On signalling that it’s your turn to speak,” Journal of experimental social psychology, vol.10, no.3, pp.234–247, 1974.

[40] M. Tice and T. Henetz, “Turn-boundary projection: looking ahead,” Proceedings of the Annual Meeting of the Cognitive Science Society, pp.838–843, 2011.

[41] M. Casillas and M.C. Frank, “Cues to turn boundary prediction in adults and preschoolers,” SemDial 2012 (SeineDial), pp.61–69, 2012.

[42] A. Keitel, W. Prinz, A.D. Friederici, C. von Hofsten, and M.M. Daum, “Perception of conversations: the importance of semantics and intonation in children’s development,” Journal of Experimental Child Psychology, vol.116, no.2, pp.264–277, 2013.

[43] L. Hirvenkari, J. Ruusuvuori, V.M. Saarinen, M. Kivioja, A. Peräkylä, and R. Hari,

"Influence of turn-taking in a two-person conversation on the gaze of a viewer," PLoS One, vol.8, no.8, p.e71569, 2013.

[44] M. Argyle and J. Dean, "Gaze and mutual gaze," Cambridge University Press, 1976.

[45] G.W. Beattie, "Floor apportionment and gaze in conversational dyads," British Journal of Social and Clinical Psychology, vol.17, no.1, pp.7–15, 1978.

[46] D.R. Rutter, G.M. Stevenson, K. Ayling, and P.A. White, "The timing of looks in dyadic conversation," British Journal of Social and Clinical Psychology, vol.17, pp.17–21, 1978.

[47] C.L. Kleinke, "Gaze and eye contact: a research review," Psychological Bulletin, vol.100, pp.78–100, 1986.

[48] C. Goodwin, "The interactive construction of a sentence in natural conversation," Everyday language: Studies in ethnomethodology, pp.97–121, 1979.

[49] A. Kalma, "Gazing in triads: A powerful signal in floor apportionment," British Journal of Social Psychology, vol.31, no.1, pp.21–39, 1992.

[50] G.H. Lerner, "Selecting next speaker: The context-sensitive operation of a context-free organization," Language in Society, vol.32, no.02, pp.177–201, 2003.

[51] J. Edlund, S. Alexandersson, J. Beskow, L. Gustavsson, M. Heldner, A. Hjalmarsson, P. Kallioinen, and E. Marklund, "3rd party observer gaze as a continuous measure of dialogue flow," LREC-The eighth international conference on Language Resources and Evaluation, LREC, 2012.

[52] J. Holler and K.H. Kendrick, "Unaddressed participants' gaze in multi-person interaction: optimizing recipiency," Frontiers in Psychology, vol.6, no.98, 2015.

[53] K. Jokinen, H. Furukawa, M. Nishida, and S. Yamamoto, "Gaze and turn-taking behavior in casual conversational interactions," ACM Transactions on Interactive Intelligence Systems, vol.3, no.2, pp.12:1–30, 2013.

[54] F. Rossano, P. Brown, and S.C. Levinson, "Gaze, questioning and culture," Conversation analysis: Comparative perspectives, pp.187–249, 2009.

[55] F. Rossano, "Gaze in conversation," The handbook of conversation analysis, pp.308–329, 2012.

[56] T. Stivers and F. Rossano, "Mobilizing response," Research on Language and social

interaction, vol.43, no.1, pp.3–31, 2010.

[57] K.H. Kendrick and J. Holler, "Gaze direction signals response preference in conversation," Research on Language and Social Interaction, vol.50, no.1, pp.12–32, 2017.

[58] G. Breed and M. Porter, "Eye contact, attitudes, and attitude change among males," The Journal of Genetic Psychology, vol.120, no.2, pp.211–217, 1972.

[59] G.N. Goldberg, C.A. Kiesler, and B.E. Collins, "Visual behavior and face-to-face distance during interaction," Sociometry, pp.43–53, 1969.

[60] S. Yamamoto, K. Taguchi, K. Ijuin, I. Umata, and M. Nishida, "Multimodal corpus of multiparty conversations in l1 and l2 languages and findings obtained from it," Language Resources and Evaluation, 2015.

[61] ELAN. http://www.lat-mpi.eu/tools/elan.

[62] J. Allwood, L. Cerrato, K. Jokinen, C. Navarretta, and P. Paggio, "The mumin coding scheme for the annotation of feedback, turn management and sequencing phenomena," Language Resources and Evaluation, vol.41, no.3-4, pp.273–287, 2007.

[63] H. Kozima and A. Ito, "The role of shared-attention in human-computer conversation," In International Conference of Research on Computational Linguistics (ROCLING-97), pp.224–228, 1997.

[64] R. Ishii and Y.I. Nakano, Estimating User's Conversational Engagement Based on Gaze Behaviors, pp.200–207, Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

Publication

Journal Article

1. Seiichi Yamamoto, Keiko Taguchi, **Koki Ijuin**, Ichiro Umata, Masafumi Nishida, "Multimodal corpus of multiparty conversations in L1 and L2 languages and findings obtained from it," Language Resources and Evaluation 49(4): pp. 857-882, March 2015
2. **Koki Ijuin**, Ichiro Umata, Tsuneo Kato, Seiichi Yamamoto, "Difference in eye gaze for floor apportionment in native-and second-language conversations,"Journal of Nonverbal Behavior 42(1): pp. 113-128, February 2018

Proceedings of International Conferences

1. Ichiro Umata, Seiichi Yamamoto, **Koki Ijuin**, Masafumi Nishida, "Effects of language proficiency on eye-gaze in second language conversations: toward supporting second language collaboration," Proceedings of the 15th ACM on International conference on multimodal interaction, pp. 413-420, November 2013
2. Keiko Taguchi, **Koki Ijuin**, Ichiro Umata, Seiichi Yamamoto, "Multimodal Japanese corpus of multi-party conversation on two different topic types," The Oriental Chapter of International Committee for the Coordination and Standardization of Speech Databases and Assessment Techniques (COCOSDA),

pp.39-42, September 2014

3. **Koki Ijuin**, Keiko Taguchi, Ichiro Umata, Seiichi Yamamoto, "Eye Gaze Analyses in L1 and L2 Conversations: From the Perspective of Listeners' Eye Gaze Activity,"Proceedings of the Workshop on Understanding and Modeling Multiparty, Multimodal Interactions at International conference on Multimodal Interaction pp.33-37, November 2014
4. **Koki Ijuin**, Yasuhiro Horiuchi, Ichiro Umata, Seiichi Yamamoto, "Eye Gaze Analyses in L1 and L2 Conversations: Difference in Interaction Structures," The 18th International Conference of Text, Speech and Dialogue, pp. 114-121, September 2015
5. Ichiro Umata, Tomoya Tanizoe, **Koki Ijuin**, Seiichi Yamamoto, "Quantitative analyses of Gaze Activity during Silence: Comparison between Native-Language and Second-Language Conversation," Proceedings of the EuroAsianPacific Joint Conference on Cognitive Science , September 2015
6. Ichiro Umata, **Koki Ijuin**, Mitsuru Ishida, Moe Takeuchi, Seiichi Yamamoto, "Quantitative Analysis of Gazes and Grounding Acts in L1 and L2 Conversation," 10th Edition of its Language Resources and Evaluation Conference, May 2016
7. Ichiro Umata, **Koki Ijuin**, Tsuneo Kato, Seiichi Yamamoto, "Gaze during utterances and silence in L1 and L2 Conversations," CogSci 2017, , pp 178-179, July 2017
8. Ichiro Umata, **Koki Ijuin**, Seiichi Yamamoto, "Factor Analysis of Gazing Activities in Native and Second Language Conversations," SEMDIAL 2017 SaarDial, , pp 178-179, August 2017
9. **Koki Ijuin**, Takato Yamashita, Tsuneo Kato, Seiichi Yamamoto, "Comparison of Effect of Speaker's Eye Gaze on Selection of Next Speaker between Native- and Second-Language Conversations," Proc. GLU 2017 International Workshop on Grounding Language Understanding, pp.32-36, August 2017
10. **Koki Ijuin**, Shohei Fujio, Albara Khalifa, Tsuneo Kato and Seiichi Yamamoto, "Comparison on Effect of Eye Gaze Activities between Human-human and Human-robot Conversations in Second-Language", The AI for Multimodal

Human Robot Interaction Workshop at the Federated AI Meeting (FAIM), July 2018

11. Ichiro Umata, **Koki Ijuin**, Tsuneo Kato, Seiichi Yamamoto, "Floor Apportionment and Mutual Gazes in Native and Second-Language Conversation", 20th ACM International Conference on Multimodal Interaction, November 2018

Proceedings of Domestic Conferences

1. **Koki Ijuin**, Keiko Taguchi, Seiichi Yamamoto, Ichiro Umata, "Multimodal Corpus of Multi-party Conversations in Second Languages,” Proceedings of the 2014 IEICE General Conference, p. 179, March 2014 (In Japanese)
2. **Koki Ijuin**, Keiko Taguchi, Ichiro Umata, Seiichi Yamamoto, "Eye Gaze Activities in Multi-party Conversations in Second Languages,” Proceedings of the 2014 IEICE General Conference, p. 180, March 2014 (In Japanese)
3. **Koki Ijuin**, Keiko Taguchi, Seiichi Yamamoto, Ichiro Umata, "Comparative analysis of Eye Gaze Activities in L1 and L2 Conversations,” Proceedings of Forum on Information Technology 13(3): pp. 323-324, September 2014 (In Japanese)
4. Ichiro Umata, Keiko Taguchi, **Koki Ijuin**, Seiichi Yamamoto, "during Silence between the Native Language and the Second Language Conversations,” Proceedings of Forum on Information Technology 13(3): pp. 325-330, September 2014 (In Japanese)
5. **Koki Ijuin**, Keiko Taguchi, Ichiro Umata, Seiichi Yamamoto, "Listener’s Eye Gaze Activities in Multi-party Conversations in Second Language,” Proceedings of the 2015 IEICE General Conference, p. 196, March 2015 (In Japanese)
6. Yasuhiro Horiuchi, **Koki Ijuin**, Keiko Taguchi, Seiichi Yamamoto, "Difference of Eye Gaze Activities between First and Second Languages,” Proceedings of the 2015 IEICE General Conference, p. 197, March 2015 (In Japanese)
7. **Koki Ijuin**, Mitsuru Ishida, Yasuhiro Horiuchi, Seiichi Yamamoto, Ichiro Umata, "Eye Gaze Analyses of Different Interaction Structure between L1 and L2 Conver-

sation," Proceedings of Forum on Information Technology 14(3): pp. 411-414, September 2015 (In Japanese)

8. Ichiro Umata, Tomoya Tanizoe, **Koki Ijuin**, Seiichi Yamamoto, "Analysis of Gaze in Second Language Conversation: from a Viewpoint of Linguistic Proficiency," Proceedings of Forum on Information Technology 14(3): pp. 315-316, September 2015 (In Japanese)
9. Moe Takeuchi, Mitsuru Ishida, **Koki Ijuin**, Ichiro Umata, Tsuneo Kato, Seiichi Yamamoto, "Difference of Eye Gaze Activities in Discourse Function between L1 and L2," Proceedings of the 2016 IEICE General Conference, p. 273, March 2016 (In Japanese)
10. Yasuhiro Horiuchi, **Koki Ijuin**, Ichiro Umata, Tsuneo Kato, Seiichi Yamamoto, "Difference of Eye Gaze Activities between Dialogue Structure of First and Second Languages," Proceedings of the 2016 IEICE General Conference, p. 280, March 2016 (In Japanese)
11. Takato Yamashita, **Koki Ijuin**, Tsuneo Kato, Seiichi Yamamoto, "Comparison of Effect of Speaker's Gaze on Selection of Next Speaker between Native- and Second-Language Conversations," Proceedings of Forum on Information Technology 16(3), September 2017 (In Japanese)

    Shohei Fujio, **Koki Ijuin**, Tsuneo Kato, Seiichi Yamamoto, "Measurement of Gaze Activities of Learners with Joining-in-type RALL System," Proceedings of the 2018 IEICE General Conference, March 2018 (In Japanese)

    Takato Yamashita, **Koki Ijuin**, Tsuneo Kato, Seiichi Yamamoto, "Comparison of Continuing Effect of Speaker's Gaze on Selection of Next Speaker between Native- and Second- Language Conversations," Proceedings of the 2018 IEICE General Conference, March 2018 (In Japanese)