GLU 2017 International Workshop on Grounding Language Understanding
25 August 2017, Stockholm, Sweden

Comparison of Effect of Speaker's Eye Gaze on Selection of Next Speaker between Native- and Second-Language Conversations

Koki Ijuin[1], *Takato Yamashita*[1], *Tsuneo Kato*[1], *Seiichi Yamamoto*[1]

[1]Doshisha University, Japan

euq1101@mail4.doshisha.ac.jp, duq0172@mail4.doshisha.ac.jp, tsukato@mail.doshisha.ac.jp, seyamamo@mail.doshisha.ac.jp

Abstract

In face-to-face communication, eye gaze is known to play various roles such as managing the attention of interlocutors, expressing intimacy, exercising social control, highlighting particular speech content and coordinating floor apportionment. In second language (L2) communication, one's perception of eye gaze is expected to have more importance than in native language (L1) because eye gaze can be used to partially compensate for the deficiencies of verbal expressions. This paper examines the efficiency of eye gaze for floor apportionment through quantitative analyses of eye gaze during three-party conversations in L1 and L2. The authors analyze the average ratios at which the participant to whom the speaker gazes takes the floor according to the duration of pauses between two consecutive utterances. The analysis results show that this ratio decreases as the duration of a pause becomes longer in L1 conversations, whereas the gazed-at participant often takes the floor even after a longer duration of pause in L2 conversations. This suggests that the effect of the speaker's eye gaze decreased when the duration of pause was prolonged in L1 conversations, whereas this effect was maintained in L2 conversations.

Index Terms: multiparty conversations, eye gaze, floor apportionment

1. Introduction

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 [1], to express intimacy [2], [3], and to coordinate their conversation [4]. Gaze is one of the strongest and most extensively studied visual cues in face-to-face interaction, and it has been associated with a variety of functions, such as managing the attention of interlocutors [5], expressing intimacy and exercising social control, highlighting the information structure of the propositional content of speech, and coordinating turn-taking [6], [1].

The fundamental gaze patterns related to turn negotiation were discussed in Kendon [1], who demonstrated that speakers look away at a turn's beginning and then look back to their partners near the turn's ending. Kendon [1] 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. He stated there were at least four such functions: (1) to provide visual feedback; (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. Argyle and Cook [7] also found that participants gazed nearly twice as much while listening as they do while speaking. In contrast, Beattie [8] reported that there was no relation between eye gaze and floor apportionment under the experimental condition in which two participants played different social positions. Those results suggest that eye gaze activity combines many functions and that the condition of conversational setup might change the relative importance of these functions [9].

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 the coordination and interaction cannot be managed in a similar way to how it is done in dialogues between two speakers who share responsibility for coordination. Lerner reported that the speaker anticipates the next speaker explicitly in many ways, including eye gaze [10]. For turn management with eye gaze, the speaker signals the assumed next speaker using his or her gaze, thus requiring the gazed-at participant to notice this gaze while the other participant also grasps the expectation that someone else will speak next. These studies showed that the eye gaze of a conversation's speaker has a relation to floor apportionment in multiparty conversations.

Related studies have been presented within several research communities, including human-computer interaction, machine learning, speech processing, and computer vision, with the aim of furthering our understanding of human-human communication and multimodal signaling of social interactions [11], [12] [13]. In these research areas, Vertegaal discussed the importance of gaze in multiparty conversations for signaling conversational attention [5], 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 [14].

These findings on human-human interactions were mainly obtained from conversations held in the native language (L1), and little is known of the effect of linguistic proficiency on multimodal conversations. The proficiency of conversational participants typically ranges widely from low to high in second-language (L2) conversations.

As for eye gaze in L2 conversations, which was expected to have almost the same functionality as it has in L1 conversations, Hosoda [15] suggested that language proficiency may affect the functions that eye gaze performs, and Veinott et al. [16] found that non-native speaker pairs benefited from using video communication in route-guiding tasks, whereas native speaker pairs did not. They argued that this was because video transmitting facial information and gestures helped the non-native pairs to negotiate a common ground, whereas this did not provide significant help for the native pairs. These observations suggest that eye gaze and visual information play more important roles in establishing mutual understanding in L2 conversations than

10.21437/GLU.2017-7

in L1 conversations.

To quantitatively and precisely analyze the difference in eye gaze between L1 and L2 conversations, Yamamoto et al. [17] created a multimodal corpus of three-party conversations for two different conversation topics in L1 and L2. 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 [17]. 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 listener's gazing ratios are larger in L2 conversations than in L1 conversations for both conversation topics. Ijuin et al. [18],[19] classified three interlocutors into current speaker, next speaker, and other participant (not next speaker) by observing the transition of speaker in the conversation and, furthermore, compared the speaker's gaze activities in L1 and L2 from the perspective of conversational interaction. The analysis revealed two key points: (1) the speaker gazes at the interlocutor who is to be the next speaker more in L2 than in L1 conversations, whereas the averages of speakers' gazing ratios are almost the same in both L1 and L2 conversations; (2) not only the next speaker but also the other participant, who is not gazed at so much by the current speaker, gazes at the current speaker more in L2 conversations.

These results suggest that the function of eye gaze in multiparty conversations in L2 is more important in coordinating a floor switch than those in L1. The eye gaze of the speaker during utterances might affect the selection of the next speaker more in L2 conversations than in L1 conversations. However, it is not clear why eye gaze has a more significant effect in L2 conversations than in L1 conversations. In this paper, we classified a conversation's listeners into two types according to the speaker's eye gaze: gazed-at participant and other participant. Then we calculated the ratios at which the gazed-at participant takes the floor, according to the duration of pauses between two consecutive utterances with floor-switch, to explore why eye gaze affects L2 conversations more significantly than it does L1 conversations.

This paper is structured as follows. We introduce the multimodal conversation corpus we used in Section 2, and in Section 3 we present the analysis results on the relation between the effect of speaker's eye gaze and the duration of pauses. Then we discuss these results in Section 4 and conclude with a summary in Section 5.

2. Multimodal Corpus

A multimodal corpus created by Yamamoto et al. [17] was used for comparing the effect of eye gaze on selection of the next speaker between native- and second-language conversations. Three subjects participated in a conversational group, sitting in a triangular formation around a table as shown in Figure 1. Three SONY video cameras were used, each capturing the face and upper body of one participant, and three NAC EMR-9 eye trackers were used to record eye gaze.

The multimodal corpus included input from a total of 60 participants (23 females and 37 males: 20 groups). They were Japanese university students who had acquired Japanese as their L1 and had learned English as their L2. The corpus contains L1 and L2 conversations held in two types of conversation. The first type was free-flowing, which was natural chatting that covered various topics such as hobbies, weekend plans, studies, and travel. The second type was goal-oriented, in which they collaboratively decided on issues related to a specific task such as deciding what to take on a trip to an uninhabited island or the mountains. Each conversation was carried out for approximately six minutes. Total number of conversations is forty for each language.

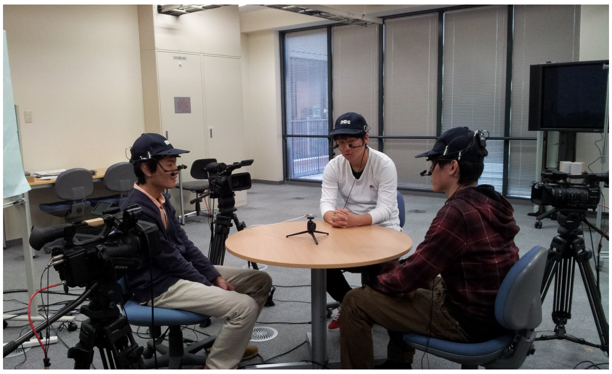


Figure 1: *Experimental setup of triad conversation*

The multimodal corpus was manually annotated in terms of the time spans for utterances, backchannel, laughing, and eye movements. Each utterance was segmented from speech at inserted pauses of more than 500 msec, and its annotation was composed of the start and end times and the attributes of utterance. The annotation of gaze events was also composed of the start and end times and attributes such as "gaze at the right-side person," "gaze at the left-side person," and "gaze at the other." Gaze events were manually annotated features defined as gazing at some object, that is, when the participant focused her/his visual attention on a particular object for a certain period of time (more than 200 msec). The annotators observed the videos and manually annotated each event.

3. Analyses

3.1. Role-based Gazing Ratio in Utterances with Floor-switch

To compare the effect of speaker's eye gaze on floor apportionment between L1 and L2 conversations, we compared the averages of Role-based Gazing Ratios in utterances after which the other takes the floor (referred to as utterances "with floor-switch") in L1 and L2 conversations.

The average of Role-based Gazing Ratios is defined as

$$Average\,of\;Role-based\,Gazing\;Ratios$$
$$= \frac{1}{n}\sum_{i=1}^{n}\frac{DG_{jk_{(i)}}}{DSU_{(i)}} * 100(\%) \qquad (1)$$

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. 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) as the participant who is not involved in floor apportionment at that time. Role-based Gazing Ratio is calculated for each group. In the following sections, Gazing Ratios

Table 1: *Gazing Ratios in utterances with floor-switch in L1 and L2 conversations*

Gazing person - gazed person	L1 conv.	L2 conv.
Current speaker - Next speaker	38.4%	46.3%
Current speaker - Other participant	19.3%	17.3%
Next speaker - Current speaker	48.9%	55.7%
Other participant - Current speaker	42.5%	47.5%

is used as the shorthand notation of the average of role-based gazing ratios.

Table 1 lists Gazing Ratios in utterances with floor-switch in both L1 and L2 conversations. The Gazing Ratio showing how long the speaker gazed at the next speaker during her or his utterance is 38.4% in L1 conversations, whereas this value for gazing at the other participant is 19.3%. On the other hand, there is a relatively small difference between the averages of the listeners' gazing ratios at the speaker by the next speaker and by the other participant, which are 48.9% and 42.5%, respectively. In other words, the speaker gazes more at the interlocutor who is to be the next speaker than at the other participant, whereas the two listeners gaze at the speaker to nearly the same degree.

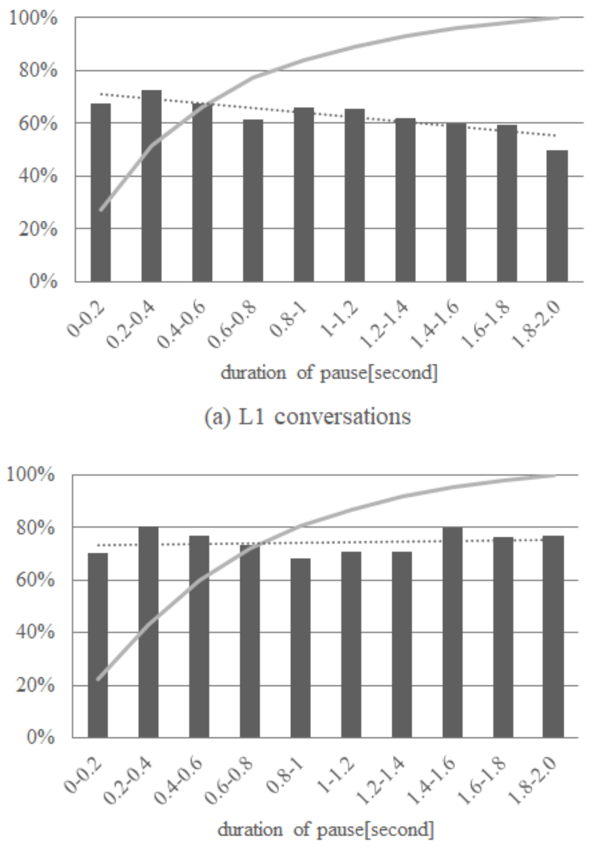
These effects of eye gaze in L1 conversations have almost the same tendency in L2 conversations. However, comparing Gazing Ratios in utterances with floor-switch in L1 and L2 conversations, the analysis revealed two key points:

1. The speaker gazes more at the next speaker, especially in utterances with floor-switch, in L2 than in L1 conversations.
2. Not only the next speaker but also the other participant, who is not gazed at so much by the current speaker, gazes at the current speaker more in L2 than in L1 conversations, even in utterances with floor-switch.

3.2. Ratios at which Gazed-at Participant Takes the Floor

To compare how long the effect of speaker's eye gaze on floor apportionment holds between native- and second-language conversations, we calculated the change of the ratios at which the gazed-at participant takes the floor according to pause duration between two consecutive utterances with floor-switch. Figure 2 compares the ratios at which the gazed-at participant takes the floor in each 0.2-second pause duration in both conversations. Each black bar represents the ratio at which the gazed-at participant takes the floor during each pause of a 0.2-second interval between two consecutive utterances with floor-switch. The dotted line indicates the linear approximation with minimum mean square error. The solid line indicates the cumulative distribution of pause duration between two consecutive utterances with floor-switch. We calculated Spearman's rank correlation coefficient between the length of pause and ratio at which the gazed-at participant takes the floor in both L1 and L2 conversations. The results showed the negative corelation in L1 conversations($\rho = -.92, p < .01$), whereas no corelation was not shown in L2 conversations($\rho = .13, p = .72$). The experimental results shown in the figure revealed the following findings:

1. The average pause duration between two consecutive utterances with floor-switch is slightly longer in L2 conversations than in L1 conversations.



(a) L1 conversations

(b) L2 conversations

Figure 2: *Change in ratios at which the gazed-at participant takes the floor in L1 and L2 conversations. The solid lines indicate cumulative distribution of pause duration between two consecutive utterances with floor-switch, and the dotted lines indicate the linear approximation with minimum mean square error.*

2. The ratios at which the gazed-at participant takes the floor during a short pause (0.2 seconds) are nearly equal in L1 and L2 conversations.
3. The ratios at which the gazed-at participant takes the floor decreased when the duration of pauses was prolonged in L1 conversations.
4. The ratios at which the gazed-at participant takes the floor reach the same degree with all durations of pauses in L2 conversations.

4. Discussion

Finding 1 in Section 3.2 shows that the next speaker might have difficulty with uttering due to a low proficiency in L2 conversations. The next speaker in general needed a longer duration to generate her/his utterance to respond to the previous utterance in L2 conversations.

Table 1 indicates that the gazed-at participant tends to take the floor more than the other participant does in both L1 and L2 conversations, and the probability that the gazed-at participant takes the floor is higher in L2 than in L1 conversations. Finding 2 in Section 3.2 shows that the ratios at which the gazed-at participant takes the floor after a short pause from the end of an utterance are almost equal in both L1 and L2 conversations. The utterances after a short pause are expected to be responses to easily understandable utterances. This suggests that the ratios at which the gazed-at participant takes the floor are correlated with the ease of generating utterances and are indirectly related to the proficiency of the participants.

Finding 3 in Section 3.2 shows that the effect of the speaker's eye gaze is reduced when the duration of pauses was prolonged, and thus the ratios come closer to 50%. In L1 conversations, the speaker's eye gaze for floor apportionment has an effect only with short pauses because the participants' proficiency in using L1 is high enough to significantly lessen the importance of the nonverbal information gained from eye gaze. The long pauses in L1 conversations might be regarded as the gazed-at participant refusing to take the floor or failing to give the floor, and thus both listeners have a chance to take the floor.

On the other hand, finding 4 in Section 3.2 suggests that the effect of speaker's eye gaze in L2 conversations was not reduced by the duration of pauses, although it was reduced in L1 conversations. In L2 conversations, all participants generally have difficulty with understanding and generating the utterances, so they needed longer pauses for response than in L1 conversations. The use of eye gaze for signaling and capturing visual information is more important in L2 conversations than in L1 conversations in order to complement the verbal information. This might be the reason why the effect of speaker's eye gaze was maintained after long durations of pauses in L2 conversations. The long pauses in L2 conversations might be regarded as the interval for understanding the previous utterances and for generating the next utterance. There is a possibility that much longer pauses might be regarded by participants as abandonment of taking the floor or failure in smooth floor apportionment, although we could not confirm this due to the small occurrence of pauses longer than 2000 msec. These results imply that the effect of eye gaze for floor apportionment might change with the level of proficiency in a language.

5. Conclusions

We compared the ratios at which a listener takes the floor from the speaker in L1 and L2 conversations. The analysis results revealed that the ratios of the gazed-at participant (person who was gazed at by the current speaker) becoming the next speaker were reduced when the duration of pauses was prolonged in L1 conversations, whereas these ratios were maintained with long durations of pauses in L2 conversations. These results suggest that the effect of the speaker's eye gaze on selection of the next speaker was maintained longer in L2 conversations than in L1 conversations. The long pauses might be interpreted as the abandonment or failure of the effort to achieve smooth floor apportionment in L1 conversations, whereas they might be regarded as intervals for understanding and generating utterances in L2 conversations. These results show that the effect of eye gaze for floor apportionment might vary with the participants' proficiency levels in a language.

The participants' groups were constructed randomly with regard to the TOEIC (Test Of English for International Communication) scores in the multimodal corpus we used. Therefore, there are various groups ranging from one in which all three participants are relatively fluent in L2 to one in which all three are seriously struggling with L2. The relation between proficiency levels of participants and non-verbal information has not been fully explored. We are now collecting a new multimodal conversational data in a better controlled way with groups categorized by L2 proficiency levels and with a balanced number of participants to analyze the effects of linguistic proficiency on eye gaze.

6. Acknowledgements

The authors would like to thank Dr. Ichiro Umata of KDDI Research, Inc. and Professor Kristiina Jokinen of the University of Helsinki (currently at AIST) for their suggestions and various discussions. This research was supported by a contract with the Ministry of Education, Culture, Sports, Science and Technology, grant number 15K02738.

7. References

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

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

[3] 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.

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

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

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

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

[8] 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.

[9] C. L. Kleinke, "Gaze and eye contact: a research review," *Psychological Bulletin*, vol. 100, p. 78100, 1986.

[10] 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.

[11] 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.

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

[13] 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.

[14] 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.

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

[16] 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," in *Proceedings of the Conference on Computer Human Interaction. CHI'99, ACM Press, PA, USA*, 1999, pp. 302–309.

[17] 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.

[18] K. Ijuin, Y. Horiuchi, I. Umata, and S. Yamamoto, “Eye gaze analyses in l1 and l2 conversations: Difference in interaction structures,” in *Text, Speech, and Dialogue - 18th International Conference, Proceedings*, 2015, pp. 114–121.

[19] K. Ijuin, I. Umata, T. Kato, and S. Yamamoto, “Difference in eye gaze for floor apportionment in native- and second-language conversations,” *Journal of Nonverbal Behavior*, 2017, in press.