

Quantitative Analysis of Gazes and Grounding Acts in L1 and L2 Conversations

Ichiro Umata^{1,2}, Koki Ijuin², Mitsuru Ishida²,
Moe Takeuchi², and Seiichi Yamamoto²

¹ KDDI R&D Labs, Human Centered Design Project
Garden Air Tower, 3-10-10, Idabashi, Chiyoda-ku, Tokyo, 102-8460, Japan

² Doshisha University, Department of Information Systems Design
1-3 Miyakodani, Tatara, Kyotanabe-shi, Kyoto, 610-0321, Japan

E-mail: <mailto:umataichiro@gmail.com>, <mailto:duo0108@mail4.doshisha.ac.jp>,
dup0113@mail4.doshisha.ac.jp, bum1061@mail4.doshisha.ac.jp, seyamamo@mail.doshisha.ac.jp

Abstract

The listener's gazing activities during utterances were analyzed in a face-to-face three-party conversation setting. The function of each utterance was categorized according to the Grounding Acts defined by Traum (Traum, 1994) so that gazes during utterances could be analyzed from the viewpoint of grounding in communication (Clark, 1996). Quantitative analysis showed that the listeners were gazing at the speakers more in the second language (L2) conversation than in the native language (L1) conversation during the utterances that added new pieces of information, suggesting that they are using visual information to compensate for their lack of linguistic proficiency in L2 conversation.

Keywords: Second Language Conversation, Gaze, Grounding

1. Introduction

Nowadays, the progress in transportation systems and communication networks connect wide areas with various linguistic and cultural backgrounds, and innovations in communication technologies enable us to communicate with remote partners via multimodal interaction systems that transmit the partners' visual images. Opportunities for multimodal communication in non-native languages are increasing faster than ever, and so is the importance of studying such communication to provide a basis for supporting it.

Previous studies have shown that gaze plays important roles in multimodal communication. Clark and colleagues have observed that visual information of the communication partners plays important roles in grounding, i.e. establishing a given piece of information as part of common ground (Clark & Brennan, 1991; Clark, 1996; Clark & Krych, 2004). Other studies discussed the speech turn organization role of gaze in verbal interaction (Kendon, 1967; Argyle, Lalljee, & Cook, 1968; Kleinke, 1986).

Although the opportunities for embodied second-language communication are increasing rapidly, quantificational studies of eye gaze in multimodal second-language conversations have been launched only recently. Studies of gazes during utterances have shown that the duration percentage when other participants are looking at the

speaker in a second-language (L2) conversation was significantly longer than in a native-language (L1) conversation, whereas the duration percentage when the speaker is looking at the other participants was not significantly different between L1 and L2 conversation (Kabashima, Nishida, Jokinen & Yamamoto, 2012; Yamasaki, Furukawa, Nishida, Jokinen & Yamamoto, 2012; Yamamoto, Taguchi, Umata, Kabashima & Nishida, 2013; Umata, Yamamoto, Ijuin, Nishida, 2013; Yamamoto, Taguchi, Ijuin, Umata & Nishida, 2015). These results suggest that gazes compensate for the lack of linguistic proficiency in L2 conversation by either (i) helping listeners understand the speakers' utterances with visual cues, (ii) helping participants monitor their partners' understanding (cf. Hosoda, 2006), or (iii) coordinating the conversational turns. These studies, however, did not consider the communicative function of each utterance, and the role gaze plays in L2 conversation has not been made clear.

In this study, we conduct a quantitative analysis of listeners' gazes while considering the communicative functions of utterances. We categorize each utterance according to the Grounding Act tag system established by Traum (Traum, 1994) and, for each Grounding Act, compare the quantities of listeners' gazes between L1 and L2 conversation. The results show that the listeners gaze at the speakers more in L2 than in L1 conversation when new pieces of information are presented. In the following, Section 2 describes the multimodal data collected in this

study, Section 3 shows the results of analysis, and Section 4 presents a discussion.

2. Data Collection

We collected data from conversations in a mother tongue and those in a second language made by the same interlocutors (for details, refer to Yamamoto et al., 2015). One of two conversational topics was assigned before each trial. One was a free-flowing conversation in which they chatted about their favorite foods. The other was a goal-oriented task in which they collaboratively decided what to take with them on a trip to a deserted island or the mountains. We randomly arranged the order of the conversation topics to counterbalance any order effect. We also randomly arranged the order of the languages used in the conversations. Each group had six-minute conversations on the free-flowing and goal-oriented topics in both Japanese and English. We collected multimodal data from 80 (20 free-flowing in Japanese, 20 free-flowing in English, 20 goal-oriented in Japanese, and 20 goal-oriented in English) three-party conversations in L1 (Japanese) and in L2 (English) languages [4], and we analyzed 20 goal-oriented in L1 and 20 goal-oriented in L2 in this paper. Twenty groups engaged in all four conversation conditions. The average duration of individual data was 6 min. All participants were native-Japanese speakers whose second language was English.

We measured their English communication levels based on the Test of English for International Communication (TOEIC). Participants were ranked within the group into three degrees of linguistic proficiency (Rank 1, Rank 2, Rank 3, from higher to lower proficiency) according to their TOEIC scores.

Three sets of NAC EMR-9 head-mounted eye trackers and headsets with microphones recorded their eye gazes and voices. The participants talked about two predetermined topics in English (second language) and in Japanese (mother tongue). Each group participated in two conversations in each language. We used the EUDICO Linguistic Annotator (ELAN) developed by the Max Planck Institute as a tool for gaze and utterance annotation.

Currently, we are annotating utterances with Grounding Act tags established by Traum (Traum, 1994), and we have finished tagging 20 groups of goal-oriented conversations analyzed in this paper. Table 1 shows the Grounding Act tags and their descriptions, and Figure 1 shows the frequency of Grounding Acts in L1 and L2 conversation.

Grounding Act	Description
Initiate (<i>init</i>)	The initial presentation of a proposition
Continue (<i>cont</i>)	A continuation of a previous act performed by the same speaker
Repair	A modification to the content or presentation of the current proposition under consideration
Request-Repair (<i>reqRepair</i>)	A request that the other participant perform a Repair
Acknowledge (<i>ack</i>)	Evidence that a previous utterance has been understood
Request-Acknowledge (<i>reqAck</i>)	A request that the other participant perform an Acknowledge
Cancel	An abandonment of the proposition under consideration
Acknowledge - Initiate (<i>ack init</i>)	"ack" and "init" occurring at the same time in one utterance unit

Table 1: Traum’s Grounding Acts

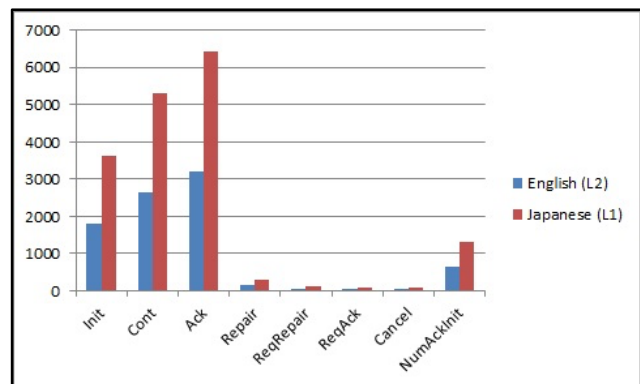


Figure 1: Frequency of Grounding Acts in L1 and L2 conversation

3. Analyses of Gazes in Utterances

We compared the amount of listeners' gaze during four major categories of Grounding Acts (i.e., init, ack init, cont, ack) between L1 and L2 conversations. We used the average of listener's gazing ratio to analyze how long the speaker was gazed at by other participants [4]. The average of listener's gazing ratios was defined as

Average of listener's gazing ratios

$$= \sum_{i=1}^n DPOS(i) / \sum_{i=1}^n D(i)$$

Here, DPOS(i) is the total duration when each participant is gazing at the speaker in the i-th utterance.

We expected that the amount of the listener's gaze would be affected by the language, the function of the utterances, and L2 linguistic proficiency, and conducted an Analysis of Variance (ANOVA) with language difference and Grounding Act being within-subject factors and L2 proficiency rank of the speaker being between-subject factors. The results revealed significant main effects of language differences ($F(1, 99) = 27.9, p < .01$) and Grounding Acts ($F(1, 297) = 204.8, p < .01$), and significant interaction between language difference and Grounding Acts ($F(3, 297) = 11.4, p < .01$). A sub-effect test showed significant simple main effect of language in Grounding act "init" ($F(1, 99) = 4.53, p < .05$), "ack-init" ($F(1, 99) = 8.91, p < .01$) "cont" ($F(1, 99) = 50.84, p < .001$). The average of listener's gaze for each Grounding Act is shown in Figure 2.

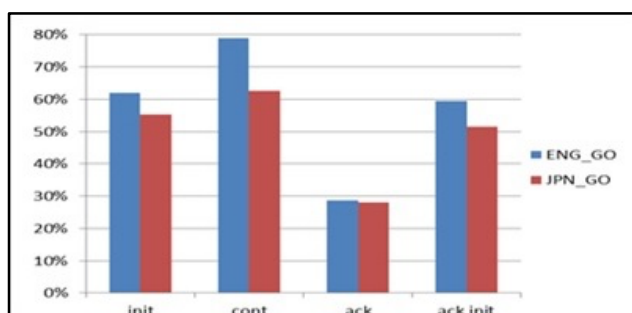


Figure 2: Average of listener's gaze for each Grounding Act

4. Discussion

The results show that the difference in the duration of listeners' gaze between L1 and L2 is notable during "int," "ack-int" and "cont" utterances, while "ack" utterances do not show any significant difference. This suggests that the "monitoring the partner's understanding" function mentioned in section 1 may not be so salient in the L2 goal-oriented conversation analyzed here: If this were the case, there would be more listener's gaze during "ack" utterances in L2 conversation. On the contrary, listeners gazed at speakers more in L2 conversation than in L1 conversation during "init," "ack-init" and "cont" utterances. This suggests that

listeners need more visual cues to compensate for their lack of linguistic proficiency in L2 during utterances where new pieces of information are presented.

The difference in the quantities of listener's gaze between L1 and L2 conversations was much bigger in "cont" than in "init" and "ack-init" utterances. A previous study suggested that the function of speaker's gaze to assign the next speaker is more salient in L2 conversation than in L1 conversation (Ijuin, Horiuchi, Umata & Yamamoto, 2015). Speakers keep another turn with a "cont" utterance, and it may be the case that they hold stronger control over the turn organization at such a moment. If so, then perhaps listeners are gazing at the speaker not only to obtain visual information that supplements the utterance contents but also to obtain visual cues for speech turn organization during "cont" utterances.

These results suggest that the listeners compensate for their lack of linguistic proficiency with visual information either to understand the utterance contents or to organize speech turns in L2 conversation, or both. This implies that multimodal communication support systems must take such characteristics of L2 conversation into consideration. For example, meeting capture systems that identify interaction events or infer the objects of interest (cf. Chiu, Kapuskar & Wilcox, 1999; Culter, Rui, Gupta, Cadiz, Tashev, He, Colburn, Zhang & Liu, 2002) must process gaze information of L2 participants with a different strategy than that used for L1 participants.

5. Summary

We conducted a quantitative analysis of listeners' gazes considering communicative functions of utterances. The results show that the difference in the duration of the listener's gaze between L1 and L2 is notable during "int" "ack-int" and "cont" utterances, and the difference is biggest in "cont" utterances. This suggests listeners are making use of gazes mostly for help in understanding their partners' utterances or for turn organization cues. This observation implies that multimodal communication support systems must take into consideration such characteristics of gaze in L2 conversation.

6. Acknowledgements

This research was supported in part by a grant from the Japan Society for the Promotion of Science (JSPS) (No. 15K00293). The authors also thank Professor Kristiina Jokinen of the University of Helsinki for her suggestions and discussions. We also deeply appreciate the reviewers for their valuable suggestions and comments.

7. Bibliographical References

- Argyle, M. Lallijee, M. and Cook, M. (1968). The effects of visibility on interaction in a dyad, *Human relations*, vol. 21, pp. 3-17.
- Chiu, P., Kapuskar, A. and Wilcox, L. (1999). Meeting capture in a media enriched conference room, in *Proceedings of CoBuild '99*, LNCS 1670, pp. 79-88.
- Clark, H. H. (1996). *Using language*, Cambridge University Press.
- Clark, H. H. and Brennan, S. E. (1991). Grounding in communication, In Resnik, L. B., Levine, J. M. and Teasley, S. D. (Eds.) *Perspectives on socially shared cognition*, pp.127-149, Washington, DC: APA Books.
- Clark H. H. and Krych, M. A. (2004). Speaking while monitoring addressees for understanding, *Journal of memory and language*, vol. 50, pp. 62-81.
- Cutler, R., Rui, Y., Gupta, A., Cadiz, J. J., Tashev, I., He, L., Colburn, A., Zhang, Z., Liu, Z. and Silverberg, S. (2002). Distributed meetings: a meeting capture and broadcasting system, in *Proceedings of the Tenth ACM International Conference on Multimedia (MULTIMEDIA2002)*, ACM Press, pp. 503-512.
- ELAN. <http://www.lat-mpi.eu/tools/elan>
- Hosoda, Y. (2006). Repair and relevance of differential language expertise, *Applied Linguistics*, Vol. 27, pp. 25-20.
- Ijuin, K., Horiuchi, Y., Umata, I. and Yamamoto, S. (2015)
- Eye gaze analyses in L1 and L2 conversations: difference in interaction structures, P. Král and V. Matoušec (Eds.) *LNAI 9302*, pp. 11-121, 2015.
- Kabashima, K., Nishida, M., Jokinen, K. and Yamamoto, S. (2012). Multimodal corpus of conversations in mother tongue and second language by same interlocutors, *Proceedings of 4th Workshop on Eye Gaze in Intelligent Human Machine Interaction*.
- Kendon, A. (1967). Some functions of gaze-direction in social interaction, *Acta Psychologica*, vol. 26, pp. 22-63.
- Kleinke, C. L. (1986). Gaze and eye contact: a research review, *Psychological Bulletin*, vol. 100, pp. 78-100.
- TOEIC. <http://www.ets.org/toEIC>
- Traum, D. (1994). *A computational theory of grounding in natural language conversation*. PhD thesis, University of Rochester.
- Umata, I., Yamamoto, S., Ijuin, K. and Nishida, M. (2013). Effects of language proficiency on eye-gaze in second language conversations: toward supporting second language collaboration, *Proceedings of the international conference on multimodal interaction ICMi2013*, pp. 413-419.
- Yamamoto, S., Taguchi, K., Ijuin, K., Umata, I. and Nishida, M. (2015). Multimodal corpus of multiparty conversations in L1 and L2 languages and findings obtained from it, *Language Resources and Evaluation*, DOI: 10.1007/s10579-015-9299-2.
- Yamamoto, S., Taguchi, K., Umata, I., Kabashima, K. and Nishida, M. (2013). Differences in interactional attitudes in native and second language conversations: quantitative analyses of multimodal three-party corpus, *Proceedings of the 35th Annual Meeting of the Cognitive Science Society*, pp. 3823-3828.
- Yamasaki, S., Furukawa, H., Nishida, M., Jokinen, K., and Yamamoto, S. (2012). Multimodal corpus of multi-party conversations in second language, *Proceedings of 8th International Conference on Language Resources and Evaluation, LREC2012*.