Quantitative Analyses of Gaze Duration from the viewpoints of Grounding Acts, Conversation Topic, and Linguistic Proficiency

Ichiro Umata(ic-umata@kddi-research.jp)
KDDI Research, Inc., Garden Air Tower, 3-10-10, Iidabashi, Chiyoda-ku, Tokyo, 102-8460, Japan

Koki Ijuin(koki-ijuuin@aist.go.jp)
The National Institute of Advanced Industrial Science and Technology, 2-3-26, Aomi, Koto-ku, Tokyo 135-0064, Japan

Tsuneo Kato(tsukato@mail.doshisha.ac.jp) & Seiichi Yamamoto(seyamamo@mail.doshisha.ac.jp)
Doshisha University, Department of Information Systems Design, Kyotanabe-shi, Kyoto, 610-0321, Japan

Abstract

Although many studies have analyzed the communicative functions of gaze, it is still unclear how linguistic proficiency and communication contexts affect gazing activities. Quantitative analyses of speaker's and listener's gaze were conducted taking the factors of grounding in communication, conversation topic, and linguistic proficiency into consideration. The results showed that the duration of a listener's gaze is much longer during utterances that convey new information while the duration of a speaker's gaze did not show much difference, suggesting that the characteristics of the grounding act factor affect a listener's gazing activities but not those of the speaker. We also observed that linguistic proficiency and conversation topic have a much greater effect on the listener's gaze. The results will contribute to multi-party interaction studies that examine the effect of linguistic proficiency, and provide valuable information that could assist in the design of interaction support systems for users with different linguistic proficiency.

Keywords: gaze; communication; grounding acts; conversation topic; linguistic proficiency

Introduction

Multimodal communication is a significant part of our daily activities, and not only verbal but also non-verbal cues have been considered important in grounding, i.e. establishing a given piece of information as part of common ground (Clark & Brennan, 1991; Clark, 1996; Clark & Krych, 2004). Among nonverbal cues, gaze has attracted the attention of researchers because of its communicative functions such as expressing emotional states, exercising social control, highlighting the informational structure of speech, and organizing speech turns (Argyle, Lallijee, & Cook, 1968; Duncan 1972; Holler & Kendrick 2015; Kendon 1967).

One of the main research topics in the examination of gaze in communication is the relation between gaze and speech modalities. Several earlier psychological studies reported that gaze has a speech floor apportionment function in native two-party conversations (Kendon 1967; Argyle, Lallijee, & Cook, 1968; Duncan 1972), and several recent studies have also confirmed the function of gaze in multiparty conversations (Kalma, 1992; Learner. 2003; Rosano, 2013; Jokinen et al., 2013; Holler & Kendrick, 2015; Yamamoto et al., 2015; Ishii et al., 2016; Auer, 2018; Ijuin et al., 2018). Another study indicated that gaze can be a collaborative signal to coordinate the insertion of responses (Bavelas et al. 2002). Furthermore, another study reported that even uninvolved observers of dyadic interactions followed the interactants' speaking turns with their gaze (Hirvenkari et al., 2013).

However, most of the studies referred to above were conducted on native language (L1) speakers. Second language (L2) conversations are commonly observed in daily life, and often the conversational proficiency of the participants varies widely. Such differences in the proficiency of participants can cause serious miscommunication and may make collaboration among both native and non-native speakers difficult in face-to-face communication (Beyene et al., 2009).

Some studies observed that gazing activity during speech interaction is affected by the linguistic proficiency of the participants. A remote work study in the human-computer interaction (HCI) field argued that video transmission of facial information and gestures helped non-native pairs to negotiate a common ground, whereas this did not provide significant help to native pairs (Veinott et. al., 1999). An analysis of second language conversation reported that eye gazes and facial expressions play an important role in monitoring both participants' understanding in the repair process (i.e. a modification to the content or presentation of the current proposition under consideration (Schegloff et al., 1977; Traum, 1994)) involving participants with differing levels of linguistic proficiency (Hosoda, 2006).

Quantitative analyses of the duration of the listener's gaze during utterances have shown that when other participants are looking at the speaker in a second-language (L2) conversation it is significantly longer than in a native-language (L1) conversation (Umata, et al., 2013; Yamamoto et al., 2015). Studies of speaker's gaze reported that the speech floor apportionment function of gaze is more prominent in second language conversations than in native language conversations (Ijuin et al., 2015; Ijuin et al., 2018). These studies, however, have not considered the effects of the communicative context. Kleinke pointed out that the conditions of a conversational setup may affect the relative importance of the multiple functions of gaze in communication (Kleinke 1986). Holler and Kendrick analyzed three-party conversations among native English speakers, focusing on question-response sequences involving

just two of the participants. They showed that unaddressed participants were able to anticipate next turns, and that the planning of these gaze shifts virtually coincided with the points at which the turns first became recognizable as possibly complete (Holler & Kendrick, 2015). There are also studies that have shown that gaze behavior is affected by the type of communication (social actions like requesting vs. complaining) and the duration of turns as projected by the social action (Kendrick & Holler, 2017; Rosano, 2013; Rossano, Brown, & Levinson, 2009; Stivers & Rossano, 2010). The role of gaze in communication is affected by the communication context, and it is important to analyze the function of gaze during utterances taking their communicative function into consideration.

Although previous studies have analyzed the effects of communication context and linguistic proficiency, it is still not very clear how these two factors interact. The current study examined the overall tendencies of speakers' and listeners' gazing activities in grounding acts while taking the effects of conversation topics and linguistic proficiency into consideration. We added the grounding tags proposed by Trauma (1994) to all the utterances in the existing triadic conversation corpus constructed by Yamamoto et al. (2015). Each utterance was categorized according to the grounding acts in the dialogue, and the gazing activities of the speakers and those of the listeners were compared between native (i.e. high linguistic proficiency) and the second language (i.e. low linguistic proficiency) conversations for two different topics. We observed that the grounding act factor affected the listener's gaze, but not that of the speaker, suggesting that the characteristics of the grounding act affects a listeners' gazing activities but not that of the speaker. We also observed that the linguistic proficiency and conversation topic have far greater effects on the listener's gaze, and that the speech direction (i.e. from someone with a high level of linguistic proficiency to someone with a low level of proficiency and vice versa) may also affect the listener's gaze but not the speaker's. The results may make a valuable contribution to multi-party interaction studies that examine the effects of linguistic proficiency, and are expected to provide information that could assist in the design of interaction support systems.

Corpus

Based on a multimodal triadic interaction corpus constructed by Yamamoto et al. (2015), we labeled all utterances in the corpus with grounding act tags by Traum (1994) for the current study (details are provided below in this section). The corpus consists of triadic conversations in a mother tongue (L1) and those in a second language (L2) made by the same interlocutors in the same group (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 participants 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 an uninhabited island or the mountains. The order of the conversation topics was randomly arranged to counterbalance any order effect. The order of the languages used in the conversations was also randomly arranged. All the conversations were subjected to the analysis in the current study.

A total of 60 subjects (23 females and 37 males: 20 groups) between the ages of 18 and 24 participated in the data collection, and each conversational group consisted of three participants unknown to each but close in age. All participants were native-Japanese speakers. Each group had six-minute conversations on both topics in both Japanese and English. The corpus contains 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) (Yamamoto et al., 2015). Twenty groups engaged in all four conversation conditions. The average duration of individual conversations was six minutes.

The English communication levels of the participants were measured based on the results of the Test of English for International Communication (TOEIC). Their scores ranged from 450 to 985 (990 being the highest score attainable), and the distribution of their TOEIC scores was nearly the same as the TOEIC test administered nationwide in Japan in the same period (cf. Yamamoto et al., 2015). The groups were assembled in various combinations of L2 proficiencies, namely, groups of participants with high TOEIC scores, those with low scores, and those with high/middle/low scores. The linguistic proficiencies of the participants were ranked also within each group according to their TOEIC scores.

Table 1. Grounding act tags and their descriptions

Grounding Act	**Description**
Initiate (*init*)	The initial presentation of a proposition
Continue (*cont*)	The 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 (*req repair*)	A request that the other participant perform a Repair
Acknowledge (*ack*)	Evidence that a previous utterance has been understood
Request-Acknowledge (*req ack*)	A request that the other participant perform an Acknowledge
Cancel	An abandonment of the proposition under consideration
Acknowledge - Initiate (*ack init*)	"*ack*" and "*init*" occurring at the same time in one utterance unit

Three sets of NAC EMR-9 head-mounted eye trackers and headsets with microphones recorded their eye gazes and voices. The EUDICO Linguistic Annotator (ELAN) developed by the Max Planck Institute was used as a tool for gaze and utterance annotation. All the utterances were annotated by two annotators using the grounding act tags

established by Traum (Traum, 1994), and each worked on 20 L1 conversations and 20 L2 conversations. Table 1 shows the grounding act tags and their descriptions, and Figure 1 shows the average occurrences of grounding acts of 20 groups in the L1 and L2 conversations. The grounding acts that only occurred a few times were not suitable for quantitative analyses, and were excluded. The major grounding acts, i.e., *init*, *cont*, *ack*, and *ack init* underwent the analyses described below. Table 2 shows the average occurrences of them.

The raw number of occurrences, not the number normalized for each participant, underwent analysis based on our research perspective where each utterance was regarded as not just the speaker's personal act but also a contribution to the group interaction; i.e. the differences in the number of utterances among the participants in a group reflect the interaction structure of the group.

Distributions of Major Grounding Acts

The occurrences of *init*, *cont*, *ack*, and *ack init* for each group showed different distributions in all conversation conditions, as shown in Figure 1. We conducted an ANOVA test to reveal the characteristics of this corpus, investigating the statistical differences in the averages of the grounding acts in native and second language conversations with language, topic and grounding act as within subject factors. The result showed significant main effects of topic ($F_{(1, 19)} = 7.971$, $p = .011$, $\eta_{p2} = .296$) and grounding act ($F_{(1.984, 37.694)} = 38.793$, $p = .000$, $\eta_{p2} = .671$. (note: where appropriate, the degrees of freedom have been adjusted by Greenhouse-Geisser correction in the following analyses)), and a significant second-order interaction among language, topic and grounding act ($F_{(2.008, 38.156)} = 18.367$, $p = .000$, $\eta_{p2} = .492$).

As shown in Figure 1, the occurrences of *ack init* are more frequent in L1 conversations than in L2 conversations, and more frequent in free-flowing conversations than in goal-oriented conversations, suggesting that it requires a high level of linguistic proficiency to acknowledge the previous utterance and present new information in one utterance. The occurrences of *cont* are more frequent in L2 than in L1, showing that there were more speaker changes in L1 conversations.

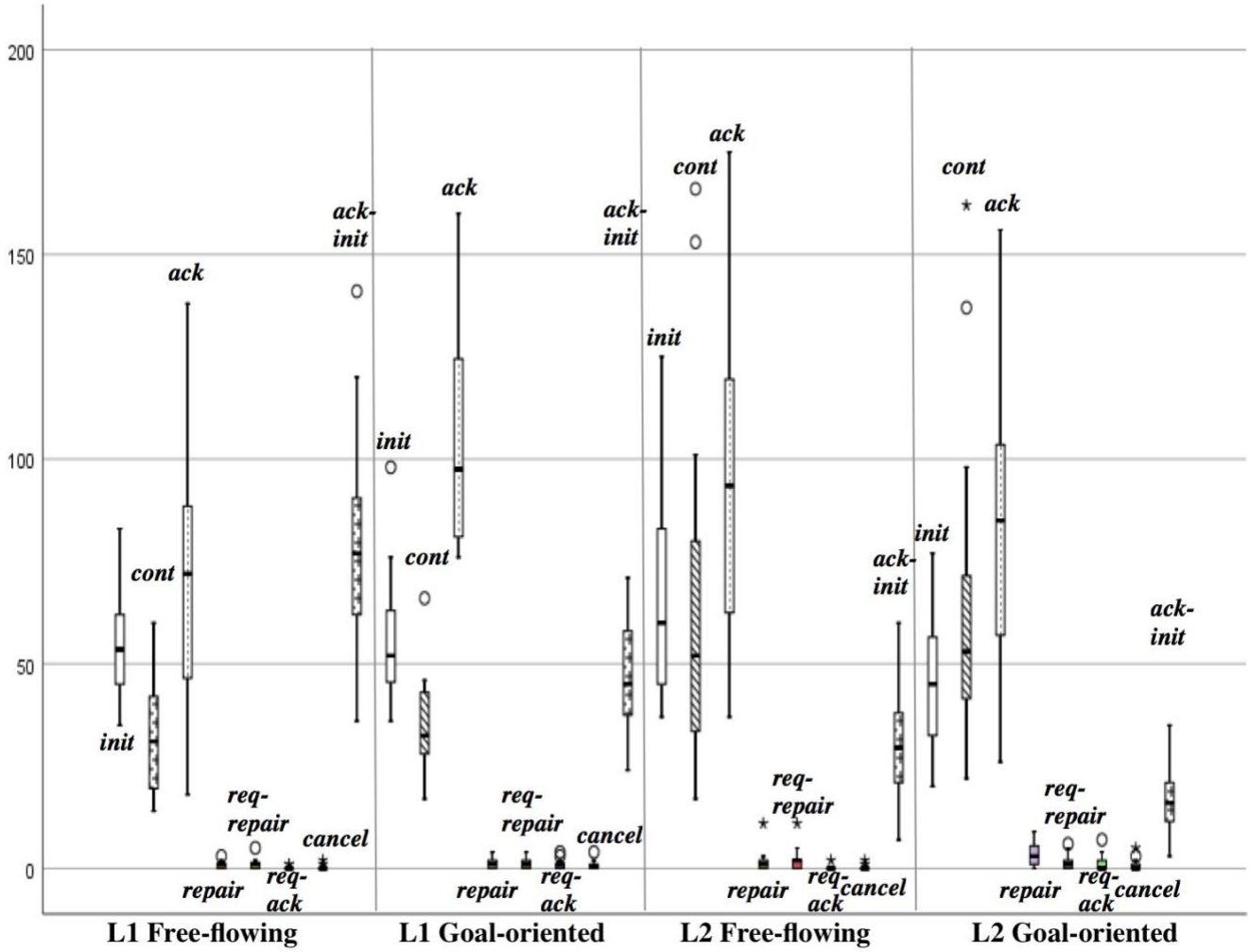

Figure 1. Average occurrences of grounding acts

Table 2. The average occurrences of *init*, *cont*, *ack*, and *ack init*

Language	Topic	Grounding Act	*Av.*	*SD*
L1	Free-flowing	*init*	54.05	12.382
		cont	32.00	13.545
		ack	72.30	27.774
		ack init	79.30	26.525
	Goal-oriented	*init*	55.20	14.424
		cont	34.55	11.381
		ack	104.75	25.961
		ack init	46.60	13.663
L2	Free-flowing	*init*	66.05	24.839
		cont	62.10	40.524
		ack	94.80	37.902
		ack init	30.75	14.134
	Goal-oriented	*init*	46.30	16.598
		cont	63.45	34.836
		ack	82.45	32.422
		ack init	17.05	7.776

Analyses: Gazes during Utterances

We compared the duration of speakers' and listeners' gazes during four major categories of grounding act (i.e., *init*, *ack init*, *cont*, *ack*) between L1 (as the high linguistic proficiency condition) and L2 (as the low linguistic proficiency condition) conversations on two different topics, respectively.

The speaker's gazing ratio indicates how long the speaker gazed at other participants during his/her utterances and is defined as the ratio of the duration of the speaker gazing at other participants to his/her speaking duration. The listener's gazing ratio indicates how long a participant gazed at the speaker during his/her utterance and is defined as the ratio of the duration of a participant gazing at the speaker to the speaking duration. The speakers' and listeners' gazes were calculated for each dyadic combination of the participants for both directions within the group, with six values for each group.

To examine the effect of linguistic proficiency in detail, we categorized each utterance direction for each dyadic combination according to their L2 linguistic proficiency: "high-to-low" represents from a participant with a higher TOEIC score to one with a lower score, and the reverse relation from low-to-high.

The average of the speaker's gazing ratios is defined as:

$$\textbf{Average of speaker's gazing ratio} = \frac{1}{n}\sum_{i=1}^{n}\frac{DSG_j(i)}{D(i)}$$

Here, D(i) is the duration of the ith utterance and DSGj(i) is the duration of the speaker gazing at the jth participants (j = 1, 2, 3) in the ith utterance.

The average of the listener's gazing ratios is defined as:

$$\textbf{Average of listener's gazing ratio} = \frac{1}{n}\sum_{i=1}^{n}\frac{DLG_j(i)}{D(i)}$$

Here, $DLG_j(i)$ is the total duration of the *j*th participant (*j* = 1, 2, 3) in each group gazing at the speaker in the *i*th utterance.

Speakers' Gaze

First, we analyzed the speakers' gaze, taking the factors of linguistic proficiency, conversation topic, and grounding act into consideration. As mentioned above, previous studies showed that the duration of the speaker's gaze is not affected by the linguistic proficiency (Umata, et al., 2013; Yamamoto et al., 2015). Another study showed that the speech floor apportionment function of gaze is observed in both L1 and L2 conversations, and it is more prominent in second language conversations than in native language conversations (Ijuin et. al., 2015; Ijuin et al., 2018). Because speakers make use of gaze to coordinate the speech floor in both L1 and L2 conversations, they would gaze at the listeners anyway although the prominence of its function may be affected by linguistic proficiency of the participants. Also, the effect of the topic would not be so salient because the resource in gaze modality is largely occupied by the floor apportionment function.

Thus, our assumption for the speakers' gaze is expressed as H1 below:

H1: The factors of linguistic proficiency, conversation topic, and the grounding act do not affect the duration of the speaker's gaze toward listeners.

The distributions of the speakers' gazes are shown in Figure 2. We conducted an analysis of variance (ANOVA) with language difference, topic difference, and grounding act as within-subject factors, and with the utterance direction ("high-to-low" indicates direction from a participant with a higher L2 linguistic proficiency to one with a lower L2 proficiency, and "low-to-high" is the reverse relation) being a between-subject factor. The results revealed a significant main effect of grounding act ($F_{(2.704,\ 281.195)} = 3.592$, $p = .014$, $\eta_{p2} = .033$) , and a significant interaction between topic and grounding act: $F_{(3,\ 312)} = 2.948$, $p = .033$, $\eta_{p2} = .028$). Sub-effect tests of the topic factor for each grounding act showed marginally significant differences in *cont* ($F_{(1,\ 104)} = 2.836$, $p = .095$, $\eta_{p2} = .027$), and *ack* ($F_{(1,\ 104)} = 3.525$, $p = .063$, $\eta_{p2} = .033$), but no significant differences in *init* ($F_{(1,\ 104)} = .517$,

p = .474, η_{p2} = .005) and *ack init* ($F_{(1, 104)}$ = .001, p = .976, η_{p2} = .000).

As shown in Figure 2, while the differences in the values were not particularly large they were significant, and the distributions were similar for all categories. Our hypothesis H1 is substantially if not fully supported: the main effects of linguistic proficiency and topic factors were not found to be significant, and the significant main effect of grounding act and the significant interaction of topic and grounding act were not so large in terms of the magnitude of differences in the gazing ratio values.

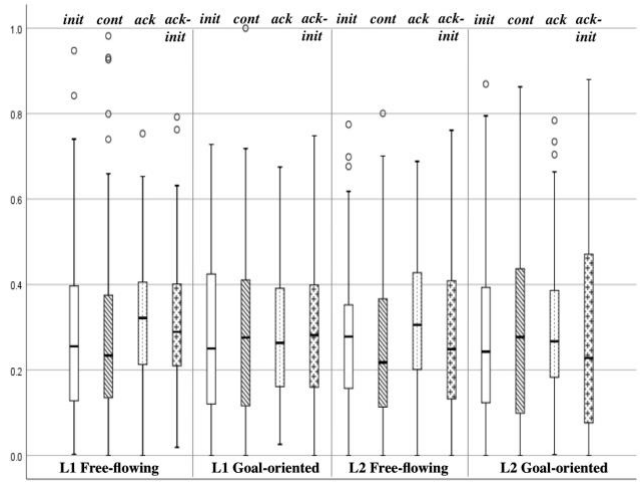

Figure 2. Distributions of the speakers' gazes

Listeners' Gaze

Next, we analyzed listener's gazes, taking the factors of linguistic proficiency, conversation topic, and grounding act into consideration. As mentioned above, previous studies showed that the duration of listener's gaze is longer in L2 conversations (Umata, et al., 2013; Yamamoto et al., 2015), possibly because the listeners were relying on the speaker's gazing cues of floor apportionment to compensate for their lack of linguistic proficiency in L2 conversations. Thus, the linguistic proficiency factor was expected to affect the duration of the listener's gaze. It is also likely that the differences in linguistic proficiency between the speaker and the listener is also a factor: a listener would make use of visual cues more often during the utterance of a speaker with higher linguistic proficiency than during the utterance of a speaker with lower proficiency in L2 conversations. Because the listeners were just monitoring the speaker's gazing cues of floor apportionment, their resource in gaze modality would not be exhausted to the same degree as that of the speaker. We expect that the listeners make use of multiple functions of gaze, and consequently the duration of their gaze would be affected by other factors such as the conversation topic and grounding act. Thus, our assumptions for the listener's gaze are expressed as H2 to H5 below:

H2: The language (L1/L2) factor strongly affects the duration of a listener's gaze toward the speaker. The duration of a listener's gaze would be longer in low linguistic proficiency (L2) conversations than in high linguistic proficiency (L1) ones.

H3: The difference in L2 linguistic proficiency between the speaker and the listener affect the duration of a listener's gaze toward the speaker in L2 conversations. The duration of a listener's gaze would be longer in utterances of the higher proficiency speaker to the lower proficiency listener in L2 conversations.

H4: The grounding act factor strongly affects the duration of a listener's gaze toward the speaker. The duration of a listener's gaze toward speakers would be longer in grounding acts *init*, *cont*, and *ack init*, where speakers present new information.

H5: The topic factor strongly affect the duration of a listener's gaze toward the speaker. The duration of a listener's gaze toward the speaker would be longer in goal-oriented conversations where the pressure to coordinate communication is stronger because of the requirement to reach an agreement.

The distributions of the listeners' gazes are shown in Figure 3. We conducted an analysis of variance (ANOVA) with language difference, topic difference, and grounding act as within-subject factors, and with the utterance direction being a between-subject factor. The results revealed significant main effects of language ($F_{(1, 112)}$ = 47.075, p = .000, η_{p2} = .296), topic ($F_{(1, 112)}$ = 16.326, p = .000), grounding act ($F_{(2.587, 289.698)}$ = 360.591, p = .000, η_{p2} = .763). The multiple comparison analysis showed that the differences among four major grounding acts were all significant (p = .000; *av*. of *init* = 0.578, *av*. of *cont* = 0.680, *av*. of *ack* = 0.280, *av*. of *ack init* = 0.520). We also observed significant interactions between language and grounding act: $F_{(2.707, 303.186)}$ = 24.828, p = .000, η_{p2} = .181), and topic and grounding act: $F_{(2.809, 314.622)}$ = 4.606, p = .004, η_{p2} = .040). Sub-effect tests of the topic factor for each grounding act showed significant differences in *cont* ($F_{(2.809, 314.622)}$ = 6.232, p = .014, η_{p2} = .053) and *ack init* ($F_{(2.809, 314.622)}$ = 25.409, p = .000, η_{p2} = .185), and a marginally significant difference in *ack init* ($F_{(2.809, 314.622)}$ = 3.137, p = .079, η_{p2} = .027), but no significant difference in *ack* ($F_{(2.809, 314.622)}$ = 1.154, p = .285, η_{p2} = .010). The results showed the duration of a listener's gaze was longer in L2 and goal-oriented conversations, and in grounding acts presenting new information.

We also observed a marginally significant main effect of utterance direction: $F_{(1, 112)}$ = 3.226, p = .075 η_{p2} = .028), and a marginally significant interaction between language and speech direction ($F_{(1, 112)}$ = 2.972, p = .087, η_{p2} = .026). Sub-effect tests of the utterance direction factor for each language condition (i.e. L1 and L2) showed significant differences for

the L2 condition ($F_{(1, 112)}$ = 5.144, p = .025 η_{p2} = .044; *av*. of *high-to-low* = 0.583, *av*. of *low-to-high* = 0.527) but not for the L2 condition ($F_{(1, 112)}$ = .518, p = .473 η_{p2} = .005; *av*. of *high-to-low* = 0.482, *av*. of *low-to-high* = 0.467), suggesting that the listeners gazed at the speaker with higher linguistic proficiency for longer.

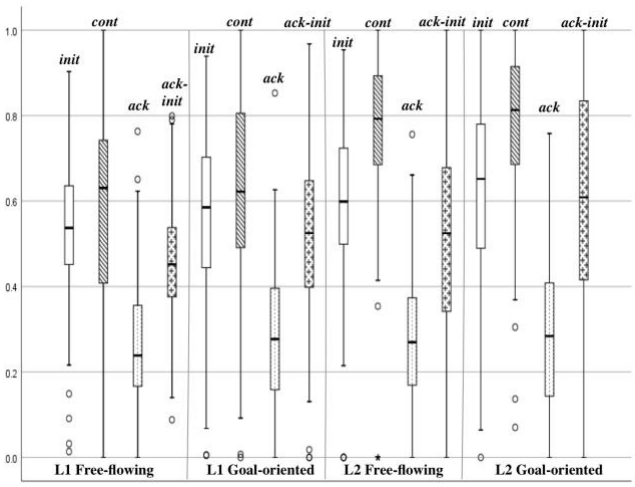

Figure 3. Distributions of the listeners' gazes

Discussion

The analyses of the speaker's gaze duration did not reveal a significant main effect of language, suggesting that the linguistic proficiency factor did not affect the duration of a speaker's gaze. Although we observed a significant main effect of grounding act and a significant interaction between topic and grounding act, the differences were not very large as shown in Figure 2. Thus, our hypothesis H1 was mostly supported.

In terms of the duration of the listener's gaze, we found that linguistic proficiency and topic had significant main effects, demonstrating that these factors affect the duration of the listener's gaze. The listeners gazed at the speaker for longer in L2 conversations in the case where their linguistic proficiency was low, and in goal-oriented conversations where the pressure to coordinate communication is stronger. The analysis also showed a significant main effect of the grounding act, and the multiple comparison analysis also demonstrated that the differences among four major grounding acts (*init*, *cont*, *ack*, and *ack init*) were all significant. Among them, the average of *ack* was the lowest by a considerable margin, showing that the duration of a listener's gaze was longer in utterances where the speaker present new information. Thus, the results were consistent with our hypotheses H2, H4, H5.

The analyses also revealed a marginally significant main effect of utterance direction and a marginally significant interaction between the language factor and speech direction, and sub-effect tests of the utterance direction factor for each linguistic proficiency condition showed significant differences only for the L2 condition. The results suggest that the difference in linguistic proficiency between the speaker and the listener affect the gazing activities of the listeners in L2 conversations: the listeners gazed at the speaker with higher linguistic proficiency for longer. Although the results were marginally significant, they can be regarded as positive arguments supporting our hypothesis H3.

Thus, our analyses suggest that the effects of linguistic proficiency and conversation topic are much greater for listeners' gaze, and that the difference in linguistic proficiency between a speaker and a listener may also affect the listener's gaze but not that of the speaker. There are, however, some factors that were not fully examined in this study. As Healey et al. observed, there might have been differences between primary and secondary recipients. Further analyses will be required to examine how the differences in listeners' roles affect their gazing behavior. Another possible factor is the difference in the length of the utterance, as this might have some impact on gazing behavior: for example, *ack* utterances were generally shorter than the other grounding acts. Not only the content of the information but also the length of an utterance might have affected gazing patterns.

Our findings may make a valuable contribution to the future multimodal communication studies that include participants with varying levels of linguistic proficiency and where there are several conversation topics. Also, these findings demonstrated factors to be considered in communication support system (ex. meeting capture, attention estimation, information highlighting, etc.) for users with varying levels of linguistic proficiency.

Conclusion

We conducted quantitative analyses of gaze during utterances, considering the factors of grounding in communication, conversation topic, and linguistic proficiency. The results showed that the gazing activities of the listeners were affected by all the factors examined in this study, although the activities of speakers were affected to a much lesser extent. The results also suggest that the difference in linguistic proficiency between the speaker and the listener affect the gazing activities of the listeners. These results should provide future multimodal communication studies with useful information on how linguistic proficiency and conversation topic affect speaker-listener interactions.

References

Auer, P. Gaze, addressee selection and turn-taking in three-party interaction. (2017). In: *InLiSt - Interaction and Linguistic Structures*, *No. 60*, August 2017, URL: http://www.inlist.uni-bayreuth.de/issues/60/index.htm

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

Argyle, M., Lallijee, M. & Cook, M. (1968). The effects of visibility on interaction in a dyad, *Human relation*s, 21, pp. 3-17.

Bavelas, J. B., Coates, L., & Johnson, T. (2002). Listener responses as a collaborative process: The role of gaze. *Journal of Communication 52, 3*, pp.566–580.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc. ser. B, 57*, pp. 298-300.

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

Clark, H. H. (1996). *Using language*, Cambridge University Press.

Clark, H. H. & Brennan, S. E. (1991). Grounding in communication, In Resnik, L. B., Levine, J. M., & Teasley, S. D. (Eds.) *Perspectives on socially shared cognition*, pp.127-149, Washington, DC: APA Books.

Clark H. H., & Krych, M. A. (2004). Speaking while monitoring addressees for understanding, Journal of memory and language, 50, 62-81.

Duncan, S. (1972). Some Signals and Rules for Taking Speaking Turns in Conversations. *Journal of Personality and Social Psychology 23*, pp. 283–292.

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

Healey, P., Lavelle, M., Howes, C., Battersby, S., & McCabe, R. (2013). How listeners respond to speaker's troubles. In *Proc. of the 35th Annual Meeting of the Cognitive Science Society*, pp. 2506-2511.

Hirvenkari, L., Ruusuvuori, J., Saarinen, V. M., Kivioja, M., Peräkylä, A., & Hari, R. (2013). Influence of turn-taking in a two-person conversation on the gaze of a viewer. *PloSOne*,*8*,*8*,e71569. https://doi.org/10.1371/journal.pone.0071569

Holler, J., & Kendrick, K. H. (2015). Unaddressed participants' gaze in multi-person interaction: optimizing recipiency. *Frontiers in psychology 6*, Article 98, 14 pages. https://doi.org/10.3389/fpsyg.2015.00098

Hosoda, Y. (2006). Repair and relevance of differential language expertise, *Applied Linguistics, 27*, pp.25-50.

Ijuin, K., Horiuchi, Y., Umata, I., & Yamamoto, S. (2015) Eye gaze analyses in L1 and L2 conversations: difference in interaction structures, Král, P., & Matoušec, V. (Eds.) *Proc. of 18th International Conference on Text, Speech and Dialogue (TSD 2015)*, pp. 114-121, LNAI 9302.

Ijuin, K., Umata, I., Kato, T., & Yamamoto, S. (2018). Difference in eye gaze for floor apportionment in native- and second-language conversations. *Journal of Nonverbal Behavior 42*, pp. 113–128.

Ishii, R., Otsuka, K., Kumano, S., & Yamato, J. (2016). Prediction of Who Will Be the Next Speaker and When Using Gaze Behavior in Multiparty Meetings. *ACM Trans. Interact. Intell. Syst. 6, 1, Article 4*, 31 pages. DOI:

Jokinen, K., Furukawa, H., Nishida, M., & Yamamoto, S. (2013). Gaze and turn-taking behavior in casual conversational interactions. *ACM Transactions on Interactive Intelligent Systems 3, 2, 12*.

Kalma, A. (1992). Gazing in triads: A powerful signal in floor apportionment. *British Journal of Social Psychology*, 31(1), 21–39.

Lerner, G. H. (2003). Selecting next speaker: The context-sensitive operation of a context-free organization. *Language in Society, 32*, 177–201.

Kendon, A. (1967). Some functions of gaze-direction in social interaction, *Acta Psychologica, 26*, pp.22- 63.

Kendrick, K. H. & Holler, J. (2017). Gaze direction signals response preference in conversation. *Research on Language and Social Interaction 50*, pp.12–32.

Kleinke, C. L. (1986). Gaze and eye contact: a research review, *Psychological Bulletin, 100*, pp.78-100.

Rutter, D.R., Stevenson, G.M., Ayling, K., & White, P.A. (1978). The timing of looks in dyadic conversation. *British Journal of Social and Clinical Psychology, 17*, pp.17–21.

Rossano, F. (2013). Gaze in conversation. *The handbook of conversation analysis*. Stivers, T. Sidnell, J. (Ed.). Wiley-Blackwell, Malden, MA, pp. 308–329. https://doi.org/10.1002/9781118325001.ch15

Rossano, F., Brown, P., & Levinson., S. C. (2009). *Gaze, questioning, and culture*. Cambridge University Press, pp. 187–249. https://doi.org/10.1017/CBO9780511635670.008

Stivers, T. & Rossano, F. (2010). Mobilizing response. *Research on Language & Social Interaction 43*, 3–31.

Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The preference for self- correction in the organization of repair in conversation. *Language 53*, 361–382. 34.

Traum, D., R. 1994. *A computational theory of grounding in natural language conversation*. Ph.D. Dissertation. University of Rochester.

Veinott, E., Olson, J., Olson, G., and Fu, X. 1999. Video helps remote work: speakers who need to negotiate common ground benefit from seeing each other. In *Proc. of the Conference on Computer Human Interaction (CHI'99)*, pp. 302-309.

Vertegaal, R., Slagter, R., Verr, G., & Nijholt, A. (2001). Eye gaze patterns in conversations: there is more to conversational agents than meets the eyes. *CHI '01 proc. of the SIGCHI conference on human factors in computing systems*, pp. 301–308.

TOEIC. http://www.ets.org/toeic

Umata, I., Yamamoto, S., Ijuin, K. & Nishida, M. (2013). Effects of language proficiency on eye-gaze in second language conversations: toward supporting second language collaboration. *Proc. of the international conference on multimodal interaction (ICMI2013)*, pp. 413-419.

Veinott, E., Olson, J., Olson, G., & Fu, X. (1999). Video helps Remote Work: Speakers Who Need to Negotiate Common Ground Benefit from Seeing Each Other. *Proc. of the Conference on Computer Human Interaction. CHI'99 (pp. 302–309)*, ACM Press, PA, USA.

Yamamoto, S., Taguchi, K., Ijuin, K., Umata, I., & Nishida, M. (2015). Multimodal corpus of multiparty conversations in L1 and L2 languages and findings obtained from it, *Language Resources and Evaluation, 49*, pp. 857-882, DOI: 10.1007/s10579-015-9299-2.