



Gaze from and Toward the Silent Third Participant in a Triadic Conversation

Ichiro Umata^{1(✉)}, Koki Ijuin², Tsuneo Kato², and Seiichi Yamamoto²

¹ KDDI Research, Inc., Garden Air Tower, 3-10-10, Iidabashi, Chiyoda-ku, Tokyo 102-8460, Japan

`ic-umata@kddi-research.jp`

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

`euqll0l@mail4.doshisha.ac.jp`,

`{tsukato,seyamamo}@mail.doshisha.ac.jp`

Abstract. This study reports an analysis of the relationship between gaze and floor apportionment in a triadic conversation, focusing on the gazes from and toward the silent third participant, in other words, the participant who was not involved in the speaker change. Based on the previous observations that mutual gazes between the current and the next speaker do play an important role in coordinating floor apportionment, we analyzed the relationship between gaze and floor apportionment focusing on the silent third participant. The result suggests that the mutual gazes between the next speaker and the silent third participants had little relation to speech turn organization.

Keywords: Multimodal interaction · Multiparty conversation · Gaze · Turn organization

1 Introduction

Multimodal interactions such as meetings, negotiations, and discussions are important social activities in workplaces, classrooms, and community management, and supporting such interactions has been an important research topic in the field of human-computer interaction (HCI) studies [1, 2]. In multimodal interactions, not only verbal but also nonverbal information play an important role. The non-verbal elements have been considered particularly important not only in the affectional and attitudinal aspect of communication [3, 4], but also in coordination of communication [5, 6]. Among nonverbal cues, gaze has attracted the strong interest of researchers. They have reported on the important functions of gaze in communication, such as expressing emotional states, exercising social control, highlighting the informational structure of speech, and organizing speech turn [7–10], and it is expected to be an important cue in evaluating communication characteristics and establishing the roles of the participants in human-computer interactions.

One of the main topics communication study researchers have focused on is the relation between gaze and speech-turn organization. Several earlier psychological studies reported that gaze has a speech-turn organization function in dyadic conversations

involving participants who speak the same language [7, 8, 10], although some were skeptical about such findings [11, 12]. Some recent multiparty conversation studies in psychology, cognitive science, and information science fields have confirmed the speech-turn organization function of gaze [9, 13–20], and it is likely that the conditions under which a conversation occurs affect the relative importance of the various functions of gaze in communication [21].

The analyses of gaze and speech-turn organization mentioned above have mainly focused on the interaction between the current speaker and the next speaker. The participants constitute a ratified structure in a multi-party conversation [22], and behaviors of the side participants, in other words, the silent participants who were not involved in the speaker change, are also important cues for capturing the characteristics of such conversations. Holler and Kendrick conducted temporal analyses of the unaddressed participants' gaze shift from the current to the next speaker, and showed they can anticipate next speech turns [9]. However, the general tendency of the gazing activities among the current speaker, the next speaker, and the silent third participants has not been analyzed quantitatively. Analyzing the behavior of silent participants is also important for capturing the characteristics of multimodal multiparty interactions, and for developing systems that support smooth and active communication and designing HCI interfaces.

This study reports a preliminary analysis of the gazing activities involving the silent third participants in triadic conversations, focusing on mutual gaze and shared gaze phenomena. As for mutual gaze, the results of the correlation analysis between the current speaker and the silent third participant suggest that their mutual gaze plays a negligible role as a speech-turn organization signal. As for shared gaze toward the silent third participant, the results of the correlation analysis suggest that the silent third participants might have been attracting less shared attention in utterances without a speaker change when the conversational flow was more predictable. These results are expected to contribute to the development of a future conversation support system and interactive interface design.

2 Corpus

We analyzed a multimodal multi-party interaction corpus with eye-gaze data collected during previous studies [15, 19]. The corpus consists of conversations in the mother tongue of the participants and conversations in a second language involving the same interlocutors (for details, refer to [15, 19]). The mother tongue conversations were the focus of analysis in this 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. All participants were native speaker of Japanese.

Three participants were seated 1.5 m apart from each other in a triangular formation around a table (see Figs. 1 and 2). The corpus covers two conversation types to examine the effect of the conversation topics on their interaction behaviors. One is free-flowing, natural chatting that ranges over various topics such as hobbies, weekend plans, studies, and travels. The second type is goal-oriented, in which participants

collaboratively decided what to take with them on trips to uninhabited islands or mountains. All the participants in the goal-oriented conversations would be under pressure to contribute to the conversation in order to reach an agreement, whereas there would be far less pressure in free-flowing conversations. Conversational flow would be more predictable in the goal-oriented conversations where the vocabulary was more limited and the domain of the discourse was defined more narrowly by the task than in the free-flowing conversations.

The order of the conversation types was arranged randomly to counterbalance any order effect. The order of the languages used in the conversations was also arranged randomly. Each group had six-minute conversations of the two types in both Japanese and English. We collected multimodal data from 80 triadic conversations in L1 (Japanese) and in L2 (English) languages (20 free-flowing in Japanese, 20 free-flowing in English, 20 goal-oriented in Japanese, and 20 goal-oriented in English). Twenty groups engaged in all four conversation types. The average duration of individual conversations was 6 min. All the participants except those in the first three groups answered a questionnaire evaluating their conversation after each conversation condition to be analyzed in other studies (see [20]).

Their eye gazes and voices were recorded via three sets of NAC EMR-9 head-mounted eye trackers and headsets with microphones. The viewing angle of the EMR-9 was 62° and the sampling rate was 60 fps. We used the EUDICO Linguistic Annotator (ELAN) developed by the Max Planck Institute as a tool for gaze and utterance annotation [24] (see Fig. 3). Each utterance is segmented from speech at inserted pauses of more than 500 ms, and the corpus was manually annotated in terms of the time spans for utterances, backchannel, laughing, and eye movements.



Fig. 1. Experimental setup

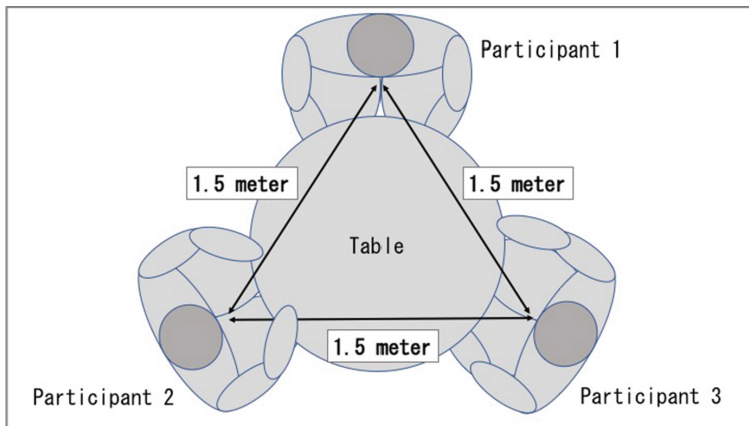


Fig. 2. Seating positions of the three participants.

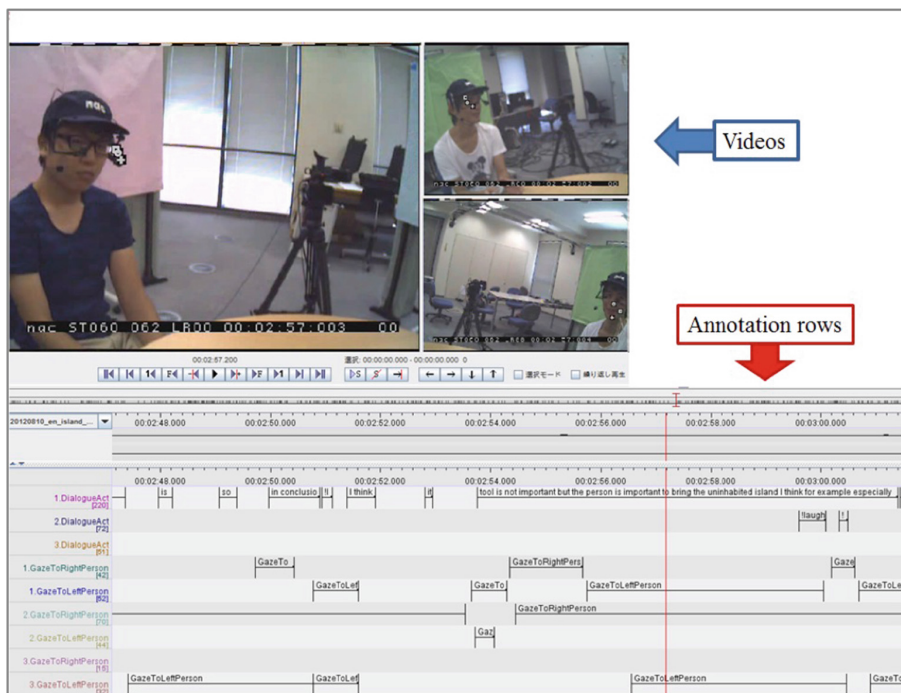


Fig. 3. Example of annotation result by using ELAN.

Studies have been conducted that observed cultural differences in gazing activities, as introduced in [13]. Rosano et al. showed that gazing activities may vary across cultures and may also be strongly related to the social actions the participants are initiating [25]. The participants gaze 1.6-fold more while listening than while speaking

in L1 conversations in the corpus analyzed here [19]. This statistical result is highly consistent with that of Vertegeal et al. [26] in multiparty conversations, regardless of the differences in languages, cultural background, and conversation topics.

3 Analyses

We focused on the mutual gaze and the shared gaze phenomena that involve the silent third participant. For the mutual gaze study, we conducted Spearman rank-order correlation analyses of the gaze from the next speaker toward the silent third participant and that from the silent third participant toward the next speaker. A previous study showed that in native language conversations there were significant positive correlations between gazes from the current to the next speaker and those from the next to the current speaker only during utterances preceding the speaker change but not utterances without a speaker change, suggesting that mutual gaze acted as a turn transition signal [20]. We assumed that the same tendency would be observed between the current speaker and the silent third participant, who is not involved in speaking at that time if mutual gaze also acts as a turn transition signal between them.

We used the average of gazing ratios for the correlation analyses based on Ijuin et al. [19]. The participant roles were 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 the silent third participant (Silent 3rd). The average of role-based gazing ratios is defined as

$$\text{average role-based gazing ratio (gazing ratio)} = \frac{1}{n} \sum_{i=1}^n \frac{DG_{jk(i)}}{DSU(i)} \times 100(\%)$$

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. A role-based gazing ratio is calculated for each group. In the following sections, “gazing ratio” is used as the shorthand notation for the average of role-based gazing ratios.

Spearman rank-order correlation analyses showed significant correlations both with and without speaker changes and in both the free-flowing and goal-oriented conversations (see Table 1), suggesting that there were mutual gazes between the current speaker and the silent third participant regardless of speaker changes.

Table 1. Spearman rank-order correlation coefficients of the gaze from the current speaker toward the silent third participant and that from the silent third toward the current speaker. (*: $p < .05$; **: $p < .01$)

| | From CS to Silent-3 rd ⇔ From Silent 3 rd to CS |
|--|--|
| Free-flowing: utterances before speaker change | .522* ($p = .018$) |
| Free-flowing: utterances without speaker change | .534* ($p = .015$) |
| Goal-oriented: utterances before speaker change | .848** ($p = .000$) |
| Goal-oriented: utterances without speaker change | .696** ($p = .001$) |

For the shared gaze study, we conducted correlation analyses of the gaze from the current speaker toward the silent third participant and that from the next speaker toward the silent third participant. We expected that the current speaker's gazing activity toward the silent third participant would invite the next speaker's gazing activity resulting in their shared gaze toward the silent third participant. The silent third participants were expected to be less prominent in conversation, and another expectation was that they might do something noteworthy that attracted the current and the next speakers' attention when they were gazed at. We assumed that a correlation would be observed both with and without speaker changes and in both the free-flowing and goal-oriented conversations.

The analyses showed a significant correlation both with and without speaker change in free-flowing conversations, whereas they showed a significant correlation only with a speaker change in goal-oriented conversations (see Table 2), suggesting that there were shared gazes toward the silent third participants other than utterances without speaker change in goal-oriented conversations.

Table 2. Spearman rank-order correlation coefficients of the gaze from the current speaker toward the silent third participant and that from the next speaker toward the silent third.

| | From CS to Silent-3 rd ⇔ From NS to Silent 3 rd |
|--|---|
| Free-flowing: utterances before speaker change | .752** ($p = .000$) |
| Free-flowing: utterances without speaker change | .445* ($p = .049$) |
| Goal-oriented: utterances before speaker change | .866** ($p = .000$) |
| Goal-oriented: utterances without speaker change | .141 ($p = .552$) |

4 Discussion

For the mutual gaze between the next speaker and the silent third participant, our analysis revealed an interesting result: the correlation analysis of the gaze from the next speaker toward the silent third participant and that from the silent third participant toward the next speaker showed that their mutual gaze was not related to speech-turn transition to the same extent as the mutual gaze between the current speaker and the next speaker was. Together with the results of previous studies that showed the duration of the current speaker's gaze toward the next speaker was significantly longer than that toward the silent participants in multiparty conversation [9, 13–17, 19], the results of this study also suggest that their mutual gaze had little relation to speech-turn organization and did not act as a speech-turn organization signal. There might have been other reasons for their mutual gaze, and it would be an interesting future research direction to examine the context of the interaction where their mutual gaze was observed.

In terms of the shared gaze toward the silent third participant, correlation analysis of gaze from the current speaker toward the silent third participant and that from the next speaker toward the silent third participant also revealed an interesting result.

Contrary to our expectation, there was an exception: their shared gaze toward the silent third participant was not observed for utterances without speaker change in goal-oriented conversations.

The cause of this phenomenon is not clear, although the predictability of the goal-oriented conversation might have been an important factor. It may be the case that the silent third participants might have been attracting less shared attention during utterances without a speaker change in goal-oriented conversation where the conversational flow was less dynamic and the current and the next speaker might have felt less need to observe the behavior of the silent third participant who was not actively involved in their speech interaction. Detailed analyses of the differences among these interaction conditions would also be an interesting future extension of this study.

5 Conclusion

We analyzed gazing activities of the current speaker, the next speaker, and the silent third participant during utterances with/without speaker change, from the viewpoints of mutual gaze and shared gaze phenomena that involve the silent third participant.

For mutual gaze between the current speaker and the silent third participant, the analysis of gaze from the current speaker toward the silent third participant and that from the silent third participant toward the current speaker showed significant correlations under all utterance conditions, suggesting that their mutual gaze had little relation to speech-turn organization, contrary to our initial expectation.

For shared gaze toward the silent third participant, the analysis of gaze from the current speaker toward the silent third participant and that from the next speaker toward the silent third participant showed a significant correlation both with and without speaker change in free-flowing conversations, whereas they showed a significant correlation only with a speaker change in goal-oriented conversations. These results suggest that the silent third participants might have been attracting less shared attention for utterances without a speaker change when the conversational flow was more predictable.

Although the causes of these phenomena are still not clear and require more detailed studies in the future, these results show that the functions of gaze are affected by the role of the participants in multimodal multiparty interaction, and are expected to contribute to forming the basis of the development of a conversation support system and interactive interface design.

References

1. Chiu, P., Kapuskar, A., Wilcox, L., Reitmeier, S.: Meeting capture in a media enriched conference room. In: Streitz, N.A., Siegel, J., Hartkopf, V., Konomi, S. (eds.) CoBuild 1999. LNCS, vol. 1670, pp. 79–88. Springer, Heidelberg (1999). https://doi.org/10.1007/10705432_8
2. Cutler, R., et al.: Distributed meetings: a meeting capture and broadcasting system. In: Proceedings of the Tenth ACM International Conference on Multimedia (MULTIMEDIA 2002), pp. 503–512. ACM Press (2002)

3. Mehrabian, A., Ferris, S.R.: Inference of attitudes from nonverbal communication in two channels. *J. Consult. Psychol.* **31**(3), 248–252 (1967)
4. Mehrabian, A., Wiener, M.: Decoding of inconsistent communications. *J. Pers. Soc. Psychol.* **6**(1), 109–114 (1967)
5. Clark, H.H.: *Using Language*. Cambridge University Press, Cambridge (1996)
6. Clark, H.H., Brennan, S.E.: Grounding in communication. In: Resnik, L.B., Levine, J.M., Teasley, S.D. (eds.) *Perspectives on Socially Shared Cognition*, pp. 503–512. APA Books (1991)
7. Argyle, M., Lalljee, M., Cook, M.: The effects of visibility on interaction in a dyad. *Hum. Relat.* **21**(1968), 3–17 (1968)
8. Duncan, S.: Some signals and rules for taking speaking turns in conversations. *J. Pers. Soc. Psychol.* **23**, 283–292 (1972)
9. Holler, J., Kendrick, K.H.: Unaddressed participants' gaze in multi-person interaction: optimizing reciprocity. *Front. Psychol.* **6**, 14 (2015). <https://doi.org/10.3389/fpsyg.2015.00098>. Article 98
10. Kendon, A.: Some functions of gaze-direction in social interaction. *Acta Psychol.* **26**, 22–63 (1967)
11. Beattie, G.W.: Floor apportionment and gaze in conversational dyads. *Br. J. Soc. Clin. Psychol.* **17**(1), 7–15 (1978)
12. Rutter, D.R., Stevenson, G.M., Ayling, K., White, P.A.: The timing of looks in dyadic conversation. *Br. J. Soc. Clin. Psychol.* **17**, 17–21 (1978)
13. Rossano, F.: Gaze in conversation. In: Stivers, T., Sidnell, J. (ed.) *The Handbook of Conversation Analysis*, pp. 308–329. Wiley-Blackwell, Malden (2013). <https://doi.org/10.1002/9781118325001.ch15>
14. Jokinen, K., Furukawa, H., Nishida, M., Yamamoto, S.: Gaze and turn-taking behavior in casual conversational interactions. *ACM Trans. Interact. Intell. Syst.* **3**(2), 12 (2013)
15. Yamamoto, S., Taguchi, K., Ijuin, K., Umata, I., Nishida, M.: Multimodal corpus of multiparty conversations in L1 and L2 languages and findings obtained from it. *Lang. Res. Eval.* **49**, 857–882 (2015)
16. Ishii, R., Otsuka, K., Kumano, S., Matsuda, M., Yamato, J.: Predicting next speaker and timing from gaze transition patterns in multi-party meetings. In: *Proceedings of the 15th ACM on International Conference on Multimodal Interaction (ICMI 2013)*, pp. 79–86 (2013). ACM, New York. <https://doi.org/10.1145/2522848.2522856>
17. Ishii, R., Otsuka, K., Kumano, S., Yamato, J.: Prediction of who will be the next speaker and when using gaze behavior in multiparty meetings. *ACM Trans. Interact. Intell. Syst.* **6**(1), 31 (2016). <https://doi.org/10.1145/2757284>. Article 4
18. Auer, P.: Gaze, addressee selection and turn-taking in triadic interaction. In: *InLiSt - Interaction and Linguistic Structures*, no. 60, August 2017 (2018). <http://www.inlist.uni-bayreuth.de/issues/60/index.htm>
19. Ijuin, K., Umata, I., Kato, T., Yamamoto, S.: Difference in eye gaze for floor apportionment in native- and second-language conversations. *J. Nonverbal Behav.* **42**, 113–128 (2018)
20. Umata, I., Ijuin, K., Kato, T., Yamamoto, S.: Floor apportionment and mutual gazes in native and second-language conversation. In: *Proceedings of the 20th ACM International Conference on Multimodal Interaction (ICMI 2018)*, pp. 334–341. ACM, New York (2018). <https://doi.org/10.1145/3242969.3242991>
21. Kleinke, C.L.: Gaze and eye contact: a research review. *Psychol. Bull.* **100**, 78–100 (1986)
22. Goffman, E.: Replies and Responses. *Lang. Soc.* **5**, 257–313 (1976)

23. Umata, I., Yamamoto, S., Ijuin, K., Nishida, M.: Effects of language proficiency on eye-gaze in second language conversations: toward supporting second language collaboration. In: Proceedings of the International Conference on Multimodal Interaction (ICMI 2013), pp. 413–419 (2013)
24. ELAN. <http://www.lat-mpi.eu/tools/elan>
25. Rossano, F., Brown, P., Levinson, S.C.: Gaze, Questioning, and Culture, pp. 187–249. Cambridge University Press, Cambridge (2009). <https://doi.org/10.1017/CBO9780511635670.008>
26. Vertegaal, R., Slagter, R., Verr, G., Nijholt, A.: Eye gaze patterns in conversations: there is more to conversational agents than meets the eyes. In: CHI 2001 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 301–308 (2001)