

# Shared Gaze toward the Speaker and Grounding Acts in Native and Second Language Conversation

**Ichiro Umata**

Interaction Design Group  
KDDI Research, Inc., Tokyo, Japan  
ic-umata@kddi-research.jp

**Koki Ijuin**

The National Institute of Advanced  
Industrial Science and Technology  
koki-ijuin@aist.go.jp

**Tsuneo Kato and Seiichi Yamamoto**

Department of Information Systems Design  
Doshisha University, Kyoto, Japan  
{tsukato, seyamamo}@mail.doshisha.ac.jp

## Abstract

The relation between shared gazes toward the current speaker and grounding acts were analyzed from the viewpoint of floor apportionment in Native (L1) and Second language (L2) conversation. Although the shared gaze phenomenon showed common characteristics between L1 and L2 conversations, there are one notable difference: in floor hold utterances, Continue (cf. (Traum, 1994)) utterances were drawing the listener’s visual attention in L1, whereas Initiate (cf. (Traum, 1994)) utterances were in L2.

## 1 Introduction

In multimodal interactions, the non-verbal cues have been considered particularly important in grounding, i.e. establishing a given piece of information as part of common ground (Clark, 1996). Among nonverbal cues, gaze has been observed to play an important role in communication, such as by expressing emotional states, exercising social control, highlighting the informational structure of speech, and speech floor apportionment (Argyle et al., 1968) (Duncan Jr., 1972) (Holler and Kendrick, 2015) (Kendon, 1967) (Umata et al., 2018) (Ijuin et al., 2018). In this study, we examine shared gaze toward the current speaker from the next speaker and the silent third participant from the viewpoints of floor apportionment and grounding acts defined by (Traum, 1994) in L1 and L2 conversations. The results of correlation analysis of gazes showed both common and different features between the two language conditions. As a common feature, there were shared gaze in floor switch utterances other than acknowledge utterances. As a different feature, there were shared gazes only in continue utterances in L1, whereas only in initiate utterances in L2.

## 2 Data

We analyzed 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)). The data contains face-to-face three-party conversation in L1 (Japanese) and in L2 (English). We analyzed data from the goal-oriented task in L1 and L2 (20 conversations for each) in this study. Three sets of NAC EMR-9 head-mounted eye trackers and headsets with microphones recorded their eye gazes and voices. A trained annotator annotated the utterances with Grounding Act tags established by (Traum, 1994) for 20 groups of goal-oriented conversations (Umata et al., 2016).

## 3 Analyses and Results

We conducted correlation analysis of the gazes toward the current speaker (CS) from the next speaker (NS) and the silent third participant (SP) for major 4 grounding acts (*Initiate (init)*, *Continue (cont)*, *Acknowledge (ack)*, and *Acknowledge and Initiate (ack init)*). We used the average of gazing ratios based on Ijuin et al. as indices for the following analyses of gaze (Ijuin et al., 2018). The participant roles were classified into three types: CS as the speaker of the utterance, NS as the participant who takes the floor after the current speaker releases the floor, and SP who is not involved in speaking at that time. The average of role-based gazing ratios is defined as:

*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,

Lang.	GA	$\rho$	$p$
L1	<i>init</i>	.805**	.000
L1	<i>cont</i>	.660**	.002
L1	<i>ack</i>	.409	.073
L1	<i>ack init</i>	.579**	.007
L2	<i>init</i>	.594**	.006
L2	<i>cont</i>	.687**	.001
L2	<i>ack</i>	.152	.523
L2	<i>ack init</i>	.632**	.004

Table 1: Correlation of gazes in floor switch

respectively. A role-based gazing ratio is calculated for each group: i.e. a single gaze ratio is computed for each session, and for each relation.

### 3.1 Shared Gazes in Floor Switch Utterances

We formulated the following hypotheses for shared gazes toward the current speaker in floor switch utterances:

**H1:** In floor switch utterances, the next speaker and the silent third participants would try to obtain the speaker’s nonverbal cues from the visual channel, resulting in frequent shared gaze.

**H2:** There would be little shared gaze toward the current speaker in *ack* utterances where the speaker only acknowledges the previous speaker’s utterances without adding any new piece of information.

The results of Spearman’s correlation analyses are as in Table 1 (The correlation coefficients with their false discovery rates (FDR)  $q < .01$  are marked with “\*\*”).

The result showed there were strong correlations other than *ack* utterances, supporting our hypotheses H1 and H2.

### 3.2 Shared Gazes in Floor Hold Utterances

We formulated the following hypotheses for shared gazes toward the current speaker in floor hold utterances:

**H3:** In floor hold utterances, the speaker’s nonverbal cues would be not as salient as floor switch utterances, resulting in less shared gaze toward the current speaker.

The results of Spearman’s correlation analyses are as in Table 2.

Lang.	GA	$\rho$	$p$
L1	<i>init</i>	.090	.705
L1	<i>cont</i>	.583**	.001
L1	<i>ack</i>	-.272	.246
L1	<i>ack init</i>	.128	.591
L2	<i>init</i>	.705**	.001
L2	<i>cont</i>	.309	.185
L2	<i>ack</i>	.323	.164
L2	<i>ack init</i>	.110	.655

Table 2: Correlation of gazes in floor switch

Our hypothesis H3 was partially supported: the results suggest less shared gaze in floor hold utterances. There were, however, high correlations in *cont* in L1, and in *init* in L2, suggesting that the speaker was drawing the listeners’ shared attention in these utterances, and the attention drawing utterance categories were different in these two language conditions.

## 4 Discussion and Future Work

The analysis of shared gazes in floor switch utterances supported our hypotheses: the speaker gathered shared attention of the listeners other than *ack* utterances. For floor hold utterances, however, the result showed differences between L1 and L2: in floor hold utterances, *cont* utterances were drawing the listener’s visual attention in L1, whereas *init* utterances were in L2. Although the cause of this difference is not clear, one factor that might have affected the listeners’ shared attention is the difference in linguistic proficiency: the speakers might have tended to start their long-lasting speech turns with cushioning utterances with less information content in L1, whereas they could not use such rhetoric in L2 where their linguistic proficiency was not high. Further analyses of utterance content would be required to elucidate the cause of this phenomenon. The current analyses were conducted for Japanese and English conversations as L1 and L2 for each, and extending the analyses to other languages will also be necessary.

## References

Michael Argyle, Mansur Lallijee, and Mark Cook. 1968. The effects of visibility on interaction in a dyad. *Human relations*, 21:3–17.

- Herbert H. Clark. 1996. *Using language*. Cambridge University Press.
- Starkey Duncan Jr. 1972. Some signals and rules for taking speaking turns in conversations. *Journal of Personality and Social Psychology*, 23:283–292.
- Judith Holler and Kobin H. Kendrick. 2015. [Unaddressed participants' gaze in multi-person interaction](#). *Frontiers in psychology*, 6.
- Koki Ijuin, Ichiro Umata, Tsuneo Kato, and Seiichi Yamamoto. 2018. Difference in eye gaze for floor apportionment in native- and second-language conversations. *Journal of Nonverbal Behavior*, 42:113–128.
- A. Kendon. 1967. Some functions of gaze-direction in social interaction. *Acta Psychologica*, 26:22–63.
- David R. Traum. 1994. *A computational theory of grounding in natural language conversation*. Ph.D. thesis, University of Rochester.
- Ichiro Umata, Koki Ijuin, Mitsuru Ishida, and Seiichi Yamamoto. 2016. Quantitative analysis of gazes and grounding acts in I1 and I2 conversations. In *Proceedings of 10th edition of the Language Resources and Evaluation Conference (LREC 2016)*, pages 4249–4252.
- Ichiro Umata, Koki Ijuin, Tsuneo Kato, and Seiichi Yamamoto. 2018. Floor apportionment and mutual gazes in native and second-language conversation. In *Proceedings of the international conference on multimodal interaction (ICMI2018)*, pages 334–341.
- Seiichi Yamamoto, Keiko Taguchi, Koki Ijuin, Ichiro Umata, and Masafumi Nishida. 2015. Multimodal corpus of multiparty conversations in I1 and I2 languages and findings obtained from it. *Language Resources and Evaluation*, 49:857–882.