

**Investigating the relation between oral and visual feedback in  
dyadic conversations: a multimodal approach**

Malin Spaniol, *Department of Psychiatry, University Hospital Cologne, Germany*

Alicia Janz, *IfL Phonetik, University of Cologne, Germany*

Simon Wehrle, *IfL Phonetik, University of Cologne, Germany*

Kai Vogeley, *Department of Psychiatry, University Hospital Cologne, Germany*

Martine Grice, *IfL Phonetik, University of Cologne, Germany*

Very short utterances produced by listeners can function as feedback signals demonstrating understanding and acknowledgement. Such feedback can indicate passive reciprocity (PR) or incipient speakership (IS). PR tokens support the ongoing turn of the interlocutor, while IS tokens signal the listener's intention to take the floor. The properties of oral feedback tokens are complex, but show consistent evidence for the relationship between a token's lexical form, its intonation contour, and its function (PR vs. IS; Sbranna et al., 2022). In face-to-face conversation, it can be assumed that other communicative channels, such as gaze, help to discriminate between PR and IS tokens. Only a small number of studies have investigated the relation between oral and visual feedback in the context of turn-taking to date.

Direct gaze by the speaker (entailing mutual gaze), creating a so-called "gaze window" (Bavelas et al., 2002), plays an important role in turn-taking (Auer, 2021) and has been proposed to function as backchannel-inviting (Skantze et al., 2014) and turn-yielding signal (Degutyte & Astell, 2021; Kendon, 1967). However, the precise interplay between turn-taking function, oral feedback, and the listener's gaze is yet to be elucidated.

During both oral and visual feedback, the listener is generally expected to use more directed gaze than the speaker. However, some studies have reported averted gaze at the beginning of turns (Degutyte & Astell, 2021). As IS tokens exclusively occur at the beginning of turns, we can expect that direct gaze during IS tokens will be reduced compared to PR tokens.

We have developed a novel multimodal approach for studying dyadic face-to-face conversation, recording both eye-gaze (using mobile eye-tracking glasses) and speech. We measured oral feedback and gaze, in three different conversational contexts, in dialogues between 8 native speakers of German (four dyads). Speakers first engaged in an introductory conversation, followed by a task-based conversation (Tangram task) and a subsequent discussion thereof. We investigated if and how oral feedback and gaze complement each other during the production of PR and IS tokens. Directed vs. averted gaze was automatically coded using fixation detection and face detection. Speech data were annotated in *Praat*.

Our analysis revealed relatively low amounts of speaker-directed gaze during feedback production, contrary to expectations (IS: 34%, PR: 39%). Still, PR tokens involved slightly more speaker-directed gaze than IS tokens, as predicted. We also observed less speaker-directed gaze in the task-based dialogues (likely due to task demands). Further, we also found clear differences between dyads in the time spent producing oral feedback and the amount of directed gaze, independent of conversational context.

The setup introduced offers opportunities for enriching the study of multimodal communication, and in a second step, can make a contribution to related fields, such as the modelling of human-agent interaction.

**Keywords:** dyadic multimodal interaction, backchannels, social gaze, feedback, turn-taking

## Figures

Figure 1. *Experimental setup.*

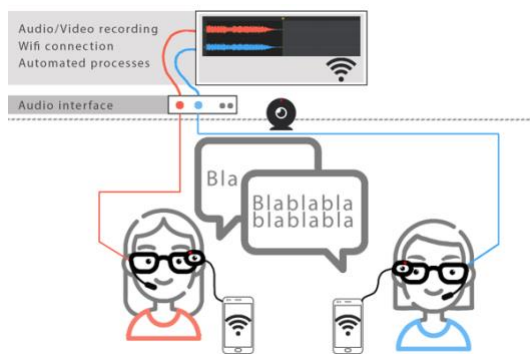
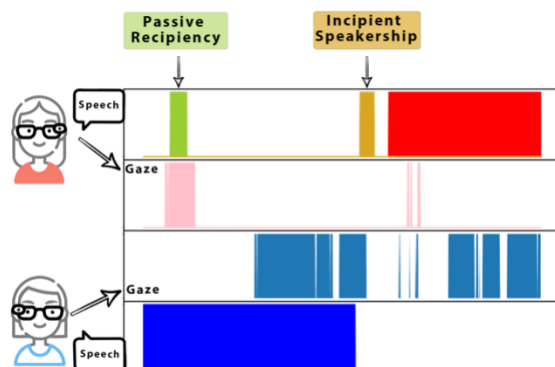


Figure 2. *Example activity plot showing speech and gaze of both interlocutors, featuring an IS token (in yellow) during which gaze is briefly averted, and a PR token (in green) with directed gaze.*



## References

- Auer, P. (2021). Turn-allocation and gaze: A multimodal revision of the “current-speaker-selects-next” rule of the turn-taking system of conversation analysis. *Discourse Studies*, 23(2), 117–140.
- Bavelas, J. B., Coates, L., & Johnson, T. (2002). Listener Responses as a Collaborative Process: The Role of Gaze. *Journal of Communication*, 52(3), 566–580.
- Degutyte, Z., & Astell, A. (2021). The Role of Eye Gaze in Regulating Turn Taking in Conversations: A Systematized Review of Methods and Findings. *Frontiers in Psychology*, 12(April), 1–22.
- Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta Psychologica*, 26, 22–63.
- Sbranna, S., Möking, E., Wehrle, S., & Grice, M. (2022). Backchannelling across Languages: Rate, Lexical Choice and Intonation in L1 Italian, L1 German and L2 German. *Proc. Speech Prosody 2022*, 734–738.
- Skantze, G., Hjalmarsson, A., & Oertel, C. (2014). *Turn-taking, feedback and joint attention in situated human-robot interaction.*