DOI: 10.17469/O2112AISV000010

SIMONA SBRANNA, MICHELINA SAVINO, FLORENCE BAILLS,
MARTINE GRICE

Vocal feedback and eye gaze patterns in Italian task-based dyadic conversations

Previous research has provided evidence of the interaction between eye gaze and turn-taking, showing that turn yielding is generally signalled with mutual gaze, whereby the primary speaker assesses the interlocutor's availability, which is subsequently confirmed by the interlocutor, as the next or secondary speaker reciprocates the gaze. However, the relation between eye gaze and turn-regulating vocal feedback signals, indicating the intention to take the conversational floor or not, can be influenced by different experimental designs. In task-based conversations, turn-regulating vocal feedback reportedly occurs predominantly when speakers avert their gaze. In this study, we investigate the relation between turn-regulating vocal feedback and eye gaze patterns, extending the binary paradigm of averted versus mutual gaze to include unilateral gaze, which can be directed towards the interlocutor when feedback signals are followed by a turn. We also find by-dyad variability, which we attribute to the fact that interlocutors can use the vocal and visual channels of communication together or separately.

Keywords: eye gaze, vocal feedback, turn-taking, multimodality, interaction.

1. *Introduction*

While the investigation of human communication has traditionally been focused on spoken language, the centrality of its multimodal nature is nowadays well recognized in linguistic studies (Paulmann, Jessen & Kotz, 2009; Perniss, 2018). Linguistic multimodality refers to the integration of a variety of information conveyed through both aural and visual channels, encompassing prosody, gestures, facial expressions, and body movements, all of which are intertwined with linguistic expression. This property of communication is particularly evident in face-to-face dialogues, where vocal and visual elements are used by interlocutors to co-construct meaning in a structured manner (Rasenberg, Pouw, Özyürek, & Dingemanse, 2022). Due to its collaborative properties, dialogue has also been compared to a Complex Dynamic System (Hollenstein, 2013), since it emerges dynamically (i.e., it is not planned) and evolves through a process of co-regulation between interlocutors based on feedback loops (Hu, Chen, 2021 for a review; Stahl, 2016 for task-based dialogues). A fundamental cooperative mechanism for structuring interactions is turn-taking, wherein speakers alternate their speech in dialogue with minimal disruption.

In spoken communication, transitions between turns have been found to occur with brief pauses and minimal overlaps, and this phenomenon is consistently

evident across individuals, methodologies, and contexts (Levinson, Torreira, 2015). This mechanism is executed with exceptional temporal precision and flexibility (Schegloff, 2020), showing that interlocutors must be able to predict in advance the endpoint of their interlocutors' turn, and respond when vocal and/or visual turn-final cues are detected (Barthel, Meyer & Levinson, 2017), allowing a participant in a dialogue to shift from recipiency to speakership with high temporal accuracy. This impressive synchronization between conversational partners has sparked scholarly interest regarding the factors governing it, and numerous studies have demonstrated the multimodal nature of turn transitions (e.g., Zellers, House & Alexanderson, 2016; Holler, Kendrick, Levinson, 2018; Kendrick, Holler & Levinson, 2023). However, the roles of eye gaze, on the one hand (e.g., Degutyte, Astell, 2021), and vocal feedback with turn-alternating functions, on the other (e.g., Drummond, Hopper, 1993), have mainly been studied independently.

This study aims to advance the understanding of how visual and vocal channels interact to create the complex feedback loops through which turn-taking is managed in real time. To do so, we conduct an exploratory analysis of the interplay between eye gaze and turn-regulating vocal feedback in signalling turn alternation in dyadic conversations. Specifically, we analyse gaze pattern frequency and temporal dynamics at the moment when turn-regulating vocal feedback occurs, based on whether the speaker takes the floor, or refrains from doing so, after producing the feedback signal. Given the multimodal nature of turn-transition feedback, we argue that such an integrated approach is particularly suitable for better accounting for conversational dynamics. Particular attention is also devoted to inter-dyadic variability.

The paper is organised as follows: Section 2 provides a review of the literature on the interplay between eye gaze and turn-taking, vocal feedback and turn-taking, and subsequently, vocal feedback and eye gaze. We conclude this section by presenting our research goal. Section 3 delineates our dataset, data collection methodology, annotation process, and measurements for the analysis. In section 4, we show the results pertaining to gaze patterns and their temporal dynamics. The results and their implications in relation to existing knowledge are discussed in section 5, followed by the conclusion in section 6.

2. *Background*

2.1 Eye gaze and turn-taking

Studies on eye gaze and turn-taking have been conducted since the 1960s with the seminal work of Kendon (1967), which significantly influenced the understanding of eye gaze dynamics at the onset and offset of a speaker's turn. Through the analysis of spontaneous dialogues, Kendon observed that a substantial majority of utterances began with the speaker looking away from the listener; conversely, they concluded with the speaker directing their gaze towards the listener. According to Kendon, at the initiation of their speaking turn, participants usually avert their gaze from their

interlocutor, serving two primary functions: 1) mitigating cognitive processing as they recall information and formulate their speech and 2) signalling the intention to take the conversational floor. For these reasons, averting gaze may constitute a strategy to signal unavailability as a listener. These observations were corroborated by subsequent studies on dyadic interactions (Cummins, 2012; Duncan, 1972; Duncan, Fiske, 1977; Ho, Foulsham & Kingstone 2015; Oertel, Włodarczak, Edlund, Wagner & Gustafson, 2012). However, some studies have challenged these findings (Beattie, 1978; Rossano, 2012; Rutter, Stephenson, Ayling & White, 1978; Streeck, 2014), suggesting that gaze aversion during early utterance production merely functions to mitigate cognitive load rather than serving as a mechanism for regulating turn-taking as such.

Regarding turn-ending, Kendon (1967) suggested that the speaker directs their gaze towards the listener to ascertain their readiness to assume the role of the next speaker, thereby signalling turn-yielding. In contrast to his claim about turn beginnings, this proposal finds strong support in subsequent studies (Novick, Hansen & Ward, 1996; Rutter et al., 1978; Lerner, 2003; Jokinen, Nishida & Yamamoto, 2009; Jokinen, Furukawa, Nishida & Yamamoto, 2013; Ho et al., 2015; Brône, Oben, Jehoul, Vranjes, & Feyaerts, 2017; Auer, 2018; Blythe, Gardner, Mushin & Stirling, 2018; Streeck, 2014). Nevertheless, Rutter et al. (1978) observed that Kendon's predictions regarding changes in conversational turns are contingent upon mutual gaze between speaker and listener. In other words, for Kendon's predictions to be substantiated, not only one, but both participants must maintain visual contact with each other. Additionally, Novik et al. (1996) identified two primary patterns of gaze direction during turn-taking: mutual-break and mutual-hold. In the mutual-break pattern, which was observed to be the most prevalent, at the conclusion of an utterance, the speaker looks at the listener, resulting in a brief period of mutual gaze until the secondary speaker initiates speech and breaks the mutual gaze. Conversely, in the mutual-hold pattern, mutual gaze is maintained while the secondary speaker begins speaking, without immediately averting their gaze. Contrary to Kendon's findings, Beattie (1978) observed that utterances ended with more immediate turn switches in the absence of mutual gaze compared to when there was mutual gaze, suggesting that gaze may not serve in determining turn-taking. In line with this finding, Kendrick et al. (2023) found that turn transitions were sped up by the use of manual gestures rather than gaze behaviour.

It is important to note that variations in findings regarding turn boundaries and gaze behaviour may stem from differences in experimental design (Degutyte, Astell, 2021), including the choice of interaction types in the analysis (such as dyadic vs. triadic/multiparty interaction or question/answer sequences vs. other types of adjacency pairs) and the task being performed (task-oriented dialogue vs. free conversation). Moreover, the methodology employed can also affect results in terms of the level of granularity (eye tracking vs. video). Furthermore, numerous individual factors can influence gaze behaviour during conversation, including gender (Argyle, Dean, 1965; Myszka, 1975; Bissonnette, 1993), social status (Myszka, 1975;

Foulsham, Cheng, Tracy, Henrich & Kingstone, 2010), acquaintance status (Rutter et al., 1978; Bissonnette, 1993), and cultural background (Rossano et al., 2009). Thus, considering and comparing these variables can help clarify some contradictory statements found in the literature.

Despite some variability and different interpretations[1], overall, it still appears that either gaze towards the listener or mutual gaze might be associated with turn yielding, while averted gaze might be related to turn initiation.

2.2 Vocal feedback and turn-taking

Speakers' contribution to a conversation significantly depends on listeners' reactions to their speech, conveyed through feedback signals. These vocal and/or visual signals enhance fluency in social interactions (Amador-Moreno, McCarthy & O'Keeffe, 2013) by supporting the ongoing turn of the interlocutor, revealing listeners' attitude towards its content, and structuring conversations (Kraut, Lewis & Swezey, 1982; Sacks, Schegloff & Jefferson, 1974; Schegloff, 1982) by guiding floor management (Jefferson, 1983).

Vocal feedback signals have been largely investigated in spoken interaction and variously named as there is no consensus in the literature regarding their definition. Fries (1952) is recognized as one of the earliest to identify these "signals of attention", which do not interrupt the speaker's discourse in telephone conversations. Over time, various terms have been used to describe this phenomenon, such as "accompaniment signals" (Kendon, 1967), "receipt tokens" (Heritage, 1984), "minimal responses" (Fellegy, 1995), "reactive tokens" (Clancy, Thompson, Suzuki & Tao, 1996), "response tokens" (Gardner, 2001), "engaged listenership" (Lambertz, 2011), and "active listening responses" (Simon, 2018). Yngve (1970) introduced the concept of "backchannel communication" to distinguish the primary channel used by the speaker from that used by the listener (i.e., the back channel), to convey essential information without actively taking the floor. In other words, backchannels are described as vocal tokens expressing acknowledgment and understanding, thereby encouraging the current speaker to continue. As backchannels are generally lexical and/or non-lexical tokens such as 'yes' or 'yeah', okay', 'mh-mh', 'uh-huh', etc., Jefferson (1983) first suggested that the use of specific tokens could additionally cue whether the listener intends to transition from a listening to a speaking role after having signalled acknowledgment/understanding to the current speaker. She observed that 'mh-mh' generally shows Passive Recipiency (PR), indicating that the current speaker should continue speaking, while 'yeah' is more often used by the listener to signal Incipient Speakership (IS), that is, the intention of taking the floor (see also Drummond, Hopper, 1993). This pragmatic distinction between PR and IS was adopted in later studies addressing dialogue act classification (e.g.,

[1] Note that there are also different interpretations of eye gaze functions, such as the one proposed by Rossano, Brown and Levinson (2009) and Streeck (2014), who claimed that gaze does not support turn-taking, but rather the organization of complex actions, which can stretch over several turns.

Jurafsky, Shriberg, Fox & Curl, 1998), and in research investigating the lexical/non-lexical and intonational characterisation of PR and IS tokens in languages other than English in task-oriented dialogues, notably in Italian and German (Savino, 2010, 2011, 2014; Savino, Refice, 2013 for Italian; Spaniol, Janz, Wehrle, Vogeley & Grice, 2023, Janz, 2023, Wehrle, 2023 for German; Sbranna, Möking, Wehrle & Grice, 2022 and Sbranna, Wehrle & Grice, in press for native Italian and German, and in L2 German by Italian learners).

In this study, the same distinction is adopted between PR tokens (produced without the speaker taking the floor, i.e., backchannels as conceptualised by Yngve, 1970, or continuers by Schegloff, 1981), and IS tokens (produced to acknowledge the interlocutor's turn before initiating their own, facilitating a smooth transition of turns) for classifying backchannels produced by participants in dyadic game-based sessions, to investigate their interplay with gaze behaviour during interaction.

2.3 Vocal feedback and eye gaze

It has been proposed that speakers might provide certain cues to invite listeners to produce feedback (see "backchannel relevant spaces" by Heldner, Hjalmarsson & Edlund, 2013). These invitation signals, often investigated under the term "backchanneling inviting cues", can be conveyed through different cues, such as pausing, intonation, head movements, or gaze direction adjustments. However, there is evidence that listeners do not actually fill all backchannel relevance spaces, as they actively and intentionally select feedback positions to support speakers in structuring their speech (Heldner et al., 2013). Furthermore, evidence suggests that a single channel is insufficient to fully account for the timing of feedback (Hjalmarsson, Oertel, 2012), suggesting the complex and multimodal nature of these inviting cues.

Several studies examining the interaction between vocal feedback and eye gaze have demonstrated that the vocal and non-vocal feedback signals which are not followed by a turn and are only used to convey acknowledgment and understanding to the speaker (backchannels as defined by Yngve, 1970; or PR tokens as defined by Jefferson, 1983), tend to occur during mutual gaze between interlocutors (Kendon, 1967; Bavelas, Coates & Johnson, 2002; Eberhard, Nicholson, 2010; Cummins, 2012; Oertel et al., 2012). In particular, Bavelas et al. (2002) found that speakers actively seek a response at crucial moments in their speech by establishing eye contact with the listener, who reciprocates, leading to brief intervals of mutual gaze termed "gaze windows", which are interrupted by the listener who averts their gaze shortly afterwards. They noted that during these gaze windows, the listener was highly likely to react, such as with a vocal "mhm" or a nod, prompting the speaker to rapidly avert their gaze and resume speaking. The authors explain this behaviour by referring to the cooperative nature of interaction (Hall, 1995; Jacoby, Ochs, 1995), suggesting a cooperative relationship between the speaker's acts and the listener's responses, which are coordinated by eye gaze.

Mutual gaze is widely recognized as a strong indicator of backchannels, regardless of their vocal or visual nature (Ferré, Renaudier, 2017; Hjalmarsson, Oertel, 2012;

Poppe, Truong & Heylen, 2011). However, a distinction has been highlighted between vocal and visual feedback signals, with visual signals being more frequently observed in conjunction with mutual gaze compared to vocal ones (Ferré, Renaudier, 2017; Eberhard, Nicholson, 2010; Truong, Poppe, Kok, & Heylen, 2011). Additionally, gaze tends to be sustained more consistently throughout the entire duration of a visual backchannel (e.g., a head nod) compared to a verbal one (Ferré, Renaudier, 2017). These findings can be explained by the fact that the presence of vocal signals does not require visual contact. The strong correlation between being looked at and gestural backchannels, but not vocal ones, has further been motivated by the hypothesis that gaze may establish a mode of communication between interlocutors, whereby a gesture responds to another gesture. Indeed, in contexts where the speaker does not make eye contact with the listener, greater variability across gestural and vocal backchannels is observed (Bertrand, Ferré, Blache, Espesser & Rauzy, 2007). This hypothesis may modulate the strong expectation that vocal feedback signals tend to cooccur with mutual gaze, supported by the aforementioned findings, partially explaining the variability in results across studies.

Further information relative to the temporal details of the "gaze windows" comes from Ondáš et al. (2023), who examined the duration of the interval between the initiation of the gaze directed towards the listener and the subsequent backchannel in interviews. They found that in most cases, the moderator receives vocal feedback from their guest within 500 ms of initiating eye contact, while the second most frequent duration value ranged from 1500 ms to 2000 ms. The interval of 1500 ms to 2000 ms was the most probable duration when the guest receives feedback from the moderator.

Despite providing relevant evidence for the interplay between eye-gaze and vocal feedback signals in general, these studies defined feedback as acknowledgments (as in Yngve, 1970 or Schegloff, 1981), whereas less is known about the interplay between their turn-regulating function (as, by contrast, in Jefferson, 1983) and gaze patterns, which was not considered in the aforementioned studies.

2.4 The interplay of turn-regulating vocal feedback and eye gaze

To the best of our knowledge, there is only one recent study (Spaniol et al., 2023) conducted on four German dyads, which examined the interaction between gaze and turn-regulating vocal feedback, taking the PR- and IS-specific functions of backchannels into consideration, under mutual and averted gaze conditions. Participants' behaviour was compared in three different and subsequent communicative contexts: 1) casual conversation for mutual acquaintance, 2) engagement in a Tangram game, and 3) free discussion regarding the game. They found that during task-based interactions, participants predominantly produced vocal feedback while averting their gaze from their interlocutor, irrespective of whether it was followed by a change in turn. Interestingly, this finding contrasts with prior research claiming that feedback is elicited by a gaze window (Bavelas et al., 2002; Novik et al., 1996) and is in line with the theory that interlocutors might respect a mutual "gesture mode" of communication (Ferré, Renaudier, 2017;

Bertrand et al., 2007). Spaniol et al. attribute their result to the presence of a visual distractor during the game, that is, the materials they had to examine to accomplish the task. In the remaining two conversational contexts, PR feedback signals were more often associated with mutual gaze than to IS feedback signals, indicating a different distribution of turn-regulating backchannels and gaze behaviour patterns according to their function and to the context of the interaction.

However, in most previous studies reviewed, only two gaze patterns are analysed: directed vs. averted gaze by the backchannel producer in the Spaniol et al. study and mutual vs. averted by both speakers in the studies in section 2.3. In the former case, these binary categories do not capture the interactive dynamic of gaze established by both interlocutors, and in the latter case, the distinction between instances in which only one participant looks at the interlocutor from instances in which both participants avert their gaze. Consequently, a binary distinction risks losing information about the search for eye contact by each speaker, which might reveal that even in task-based settings speakers do seek eye contact despite a visual competitor.

Therefore, in the present paper, we aim to expand the exploration of the relation between eye gaze and turn-regulating vocal feedback beyond the binary gaze pattern distinction of mutual vs. averted gaze to include unilateral gaze. In particular, we aim to (1) assess the relation between a specific gaze pattern and the production of turn-regulating vocal feedback, and (2) examine the extent to which the duration of this specific gaze pattern expedites the production of the feedback utterance and the potential subsequent turn alternation.

3. *Method*

To address our research questions, we analysed the gaze pattern (mutual, averted, unilateral) of participants while producing vocal feedback expressions classified as PR or IS backchannels (i.e., followed or not by a turn change) by pairs interacting in a Tangram game-based dialogue.

3.1 Data

The analysed data consist of six audio-video recorded sessions including pairs of Italian participants involved in Tangram game-based dialogues (as described in Savino, Lapertosa & Refice, 2018).

The use of Tangram-based dialogues was intended to create a controlled setting in which we could assume a certain homogeneity in behaviour across speakers. By contrast, free conversation, might include personal topics with emotional components, potentially influencing eye gaze unpredictably across dyads (Adams & Kleck, 2005; Liang, Zou, Liang, Wu & Yan, 2021). Additionally, Tangram-based games also reflect real-world communication contexts involving a visual object as the topic of conversation (e.g., discussing food choices from a menu or choosing directions while looking at a navigation device, among many other examples) with the advantage of not involving any emotional component. Finally, this approach

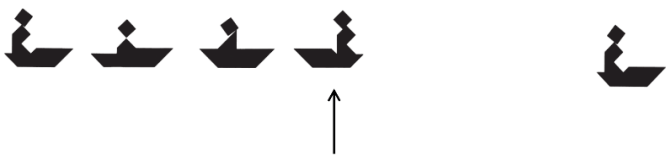
has already been used to study the interplay between eye gaze and turn alternation, providing some comparability (Spaniol et al., 2023). At the time of recording, participants were all students at the University of Bari (Italy) and shared the same geographic area of origin, gender, and acquaintance status, as they were all from the Bari metropolitan area, were female (aged 21–25), and knew each other as they were classmates. Speakers of each dyad sat at desks facing each other, separated by a panel which allowed eye contact while inhibiting the view of Tangram figures. Participants wore head-mounted microphones connected to a digital recorder, whereas camcorders were placed behind each participant, capturing their upper body and face frontally.

In each game round, participants were assigned roles as either Director or Matcher and given sets of Tangram figures accordingly. The Director received a set of four figures, in which one figure was marked with an arrow, and the Matcher received only one figure extracted from the Director's set. The objective of each round was to ascertain whether the marked figure on the Director's sheet and the figure on the Matcher's sheet were the same (an example of Director's and Matcher's sets is shown in Figure 1). Instructions prompted the Director to initially describe the marked figure; afterwards, participants were meant to interact collaboratively until reaching consensus about whether the two figures were the same or not. They verified their decision by lifting up their sheet and showing their figures to each other.

Each session includes 22 rounds of the Tangram game. Participants alternated their role as Director or Matcher in each round so that the distribution of role types was balanced between partners across the entire recording session.

We analysed rounds 1, 2, 9, 10, 21, and 22 of each recorded session. This selection was decided to account for the beginning (rounds 1-2), the medial (rounds 9-19), and the final part (rounds 21-22) of the recording sessions, with the aim of ascertaining the role of round progressions on the gaze patterns[2]. The total amount of analysed interactions corresponds to 45 minutes.

Figure 1 - *Example of a Tangram set of figures used in the Bari Italian game dialogues. The sheet on the left was given to the participant playing the Director role in that round, the sheet on the right to the participant playing the Matcher role. The aim of each round was to collaboratively ascertain whether the arrowed figure on the left matched that on the right*



[2] We also accounted for the influence of round progression on gaze and vocal feedback behaviour. However, the analysis did not reveal any consistent changing pattern as rounds progressed, instead we found a highly variable picture with dyad-specific preferences. This is why we show results pooled across rounds.

3.2 Annotations

The data contain transcripts of all game dialogues, where the beginning and end of each game round are reported, along with the participant's role (Director/Matcher) in that round. Multilevel annotations of the speech signal are also provided, which include game round intervals, as well as inter-pausal units (IPUs – speech intervals separated by at least 100 ms of silence), phonological words, and syllables produced by each of the two participants in each dyad. An additional level for vocal feedback annotation is also provided, as described in the following section.

3.2.1 Annotation of vocal feedback signals

Vocal feedback entailing lexical and non-lexical tokens conveying acknowledgment to the speaker are annotated, along with their function in regulating turns. As described in section 2.2, when the secondary speaker refrains from taking the floor after providing vocal feedback, allowing the primary speaker to continue, we categorised the feedback as PR. Conversely, when the secondary speaker takes the floor immediately after the production of vocal feedback, we annotated it as IS (Savino, 2010; 2011; 2014, also adopted by Sbranna et al., 2022; in press, and by Spaniol et al. 2023).

An example of both categories is provided in the following example from the data[3]:

Speaker A: allora la mia figura
(Eng.: so my figure)
Speaker B: sì **[vocal feedback-PR]**
(Eng.: yes)
Speaker A: ha la forma di un vaso e la parte superiore assomiglia ad un diamante
(Eng.: has a vase shape and the upper part looks like a diamond)
Speaker B: <m> **[vocal feedback-PR]**
(Eng.: <m>)
Speaker A: immagina che alla base ci sia un triangolo e sopra ci sia un diamante rove+ un diamante
(Eng.: Imagine that at the base there is a triangle, and on top, there is a diamond upsided+, a diamond)
Speaker B: <m> **[vocal feedback-IS]** da quante<ee> parti è composta?
(Eng.: <m> how many parts it is composed of?)

3.2.2 Annotation of eye gaze directions

We annotated gaze in ELAN (The Language Archive, 2023; Wittenburg, Brugman, Russel, Klassmann & Sloetjes, 2006). Following the methodologies established in prior research (Kendon, 1967; Beattie, 1978; Goodwin, 1980; Egbert, 1996; Novik et al., 1996; Jokinen et al., 2009; Streeck, 2014; Auer, 2018; Blythe et al., 2018), we annotated the intervals of directed gaze. These intervals denote periods during which participants consistently gaze in one direction without shifts, with interval boundaries marked by gaze transitions in different directions.

[3] The annotation convention "+" indicates an interruption. Vocalisations and nasalisations are indicated in angle brackets "< >".

We annotated gaze as directed towards "Task", "Addressee", or "Other". Gaze labelled as "Task" entailed the participant looking at their desk where the Tangram figures were placed, while "Addressee" involved the participant looking directly at their conversational partner. With "Other" we coded cases when participants gazed at the experimenter; however, since such cases were rather rare, they were not included in the analysis. Notably, participants showed very homogeneous gaze behaviour in that they either looked at the desk or directly at their partner.

We did not establish a minimum duration threshold for eye gaze annotation, but annotated all perceivable changes between gaze directions. However, since some previous studies used minimum duration thresholds (Beattie, 1979; Jokinen et al., 2009; Brône et al., 2017; Zima et al., 2019; Bavelas et al., 2002), we report the shortest duration of gaze direction intervals found in our data, which is 0.069 seconds, a value below most of the previously used thresholds (see Degutyte, Astell, 2021 for a review).

Twenty percent of the data were independently annotated by another linguist trained in eye gaze annotation. Inter-annotator reliability was assessed using Staccato in ELAN (Lücking, Ptock & Bergmann, 2011), a reliability measure of the temporal overlap between annotations based on Thomann's technique (2001). The result shows a mean degree of overlap of 70 %, indicating substantial reliability. No disagreements in category labels were found across annotators. The annotations by the first author were used for analysis.

3.3 Measurements

From the annotations, we derived the following categories and measurements for our analysis:

- *Mutual gaze:* when intervals of gaze to "Addressee" from both participants overlap, that is, they look at each other;
- *Averted gaze:* stands for overlapping intervals of gaze to "Task" from both participants, i.e., when both participants do not look at each other but at the task materials;
- *Unilateral gaze:* derives from one single interval of gaze to "Addressee", whereas the other speaker features gaze to "Task", i.e., one participant is looking at the other, while the interlocutor is looking at the task materials.

Given these three categories, which we will refer to as gaze patterns, we analysed: 1) the proportion of occurrences of PR and IS feedback signals under the different gaze pattern conditions, that is, at the moment when the feedback is uttered; 2) the length of this gaze pattern before vocal feedback is uttered, that is, how long a specific gaze pattern had already been established when the vocal feedback signal is produced. This time window was operationalised as a time interval (i.e., duration) calculated from the onset of the gaze pattern to the onset of the vocal feedback.

Given the exploratory nature of this study, we do not provide any inferential statistics. Instead, a detailed analysis at a by-dyad level is presented to account for

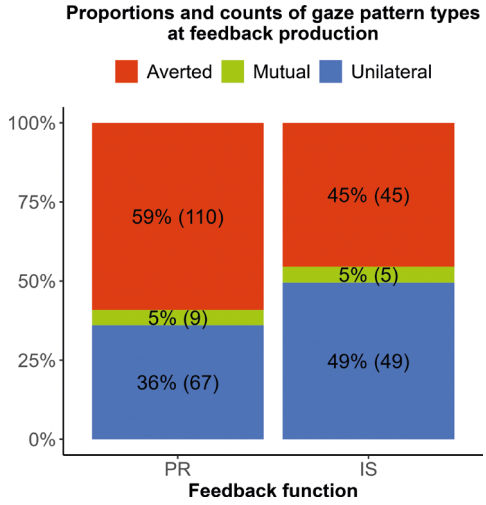
the individual conversational dynamic of each pair. The data tables and the code used to perform the analysis are openly available in the accompanying repository at https://osf.io/qzgwj/?view_only=270e19b9c36b4070ae41b6444f0220a9.

4. *Results*

4.1 Occurrences of gaze pattern types at feedback production

Figure 2 shows that both IS and PR vocal feedback are rarely produced in co-occurrence with mutual gaze (5 %). There is a prevalence of averted gaze co-occurring with PR feedback (59 %), in which both members of the dyad look at the task materials. However, speakers do not devote their visual attention exclusively to the task: unilateral gaze co-occurs with PR feedback in 36 % of cases and with IS vocal feedback in 49 % of cases, showing that speakers still use gaze to visually check the interlocutor across both feedback functions, prevalently before turn transitions.

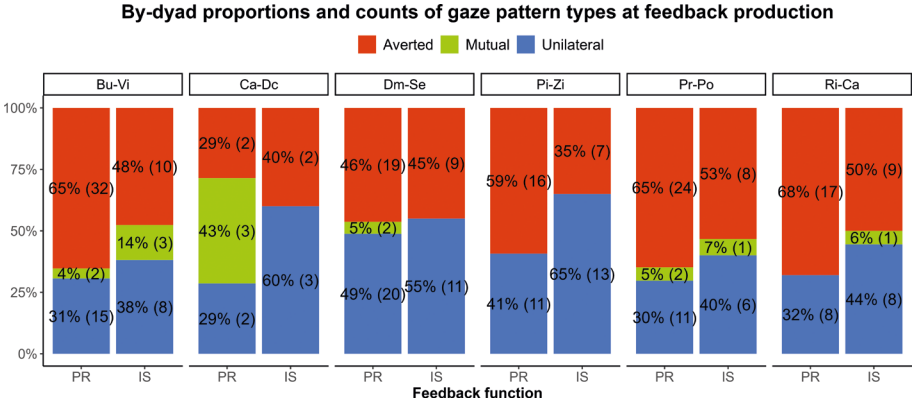
Figure 2 - *Proportions (and counts) of gaze pattern types when PR and IS vocal feedback signals were produced by all dyads in the Tangram game rounds*



A partially different view is shown in Figure 3, which displays the same proportions presented by dyad. Here, a higher percentage of unilateral gaze than averted gaze connected to the production of IS feedback is confirmed in three dyads only (Ca-Dc; Dm-Se; Pi-Zi). This suggests that dyad-specific variability is involved. It is worth noting that in these three dyads, no mutual gaze is present at all for IS, whereas the other three dyads (Bu-Vi; Pr-Po; Ri-Ca), with values of averted gaze higher than unilateral gaze, present a low percentage of mutual gaze. When summing the two values for mutual and unilateral gaze for the latter three dyads, results show a very similar distribution of "no gaze involved" vs "some gaze involved". As for PR feedback signals, the general trend is confirmed in four dyads out of six (Bu-Vi; Pi-Zi; Pr-Po;

Ri-Ca), that is a prevalence of averted gaze. However, while the specific percentage values are dyad-dependent, an overall trend can be identified: PR feedback shows a higher co-occurrence of averted gaze than IS, and IS shows a higher co-occurrence of unilateral gaze than PR.

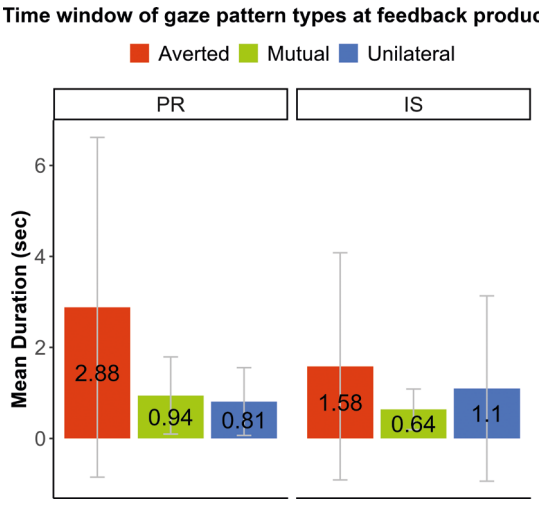
Figure 3 - *Proportions (and counts) of gaze pattern types at PR and IS vocal feedback productions, pooled by dyad*



4.2 Time window of gaze pattern types before feedback production

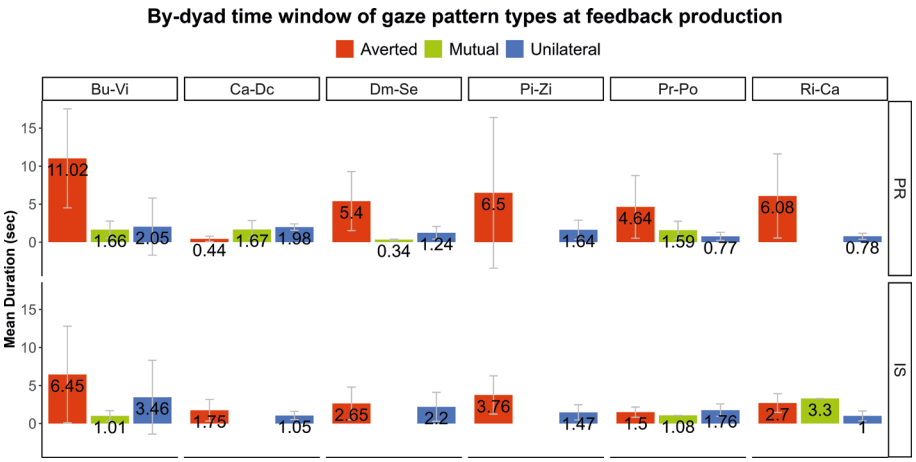
Figure 4 displays the time window for the gaze pattern preceding feedback production. It reveals that the duration of averted gaze up to the moment of feedback production is longer before PR feedback than IS feedback. Conversely, unilateral gaze is longer before IS feedback. The mutual gaze period is shorter before a turn initiation (that is, following an IS vocal feedback) than in cases in which feedback signals do not entail a change of speaker (PR vocal feedback). This shows that if a "gaze window" is activated and there is willingness to take the floor, this happens more quickly than when a speaker is only willing to acknowledge the other's speaker speech. However, given the very few data points of mutual gaze compared to averted and unilateral gaze (see raw numbers in Fig. 2), this observation should be considered only preliminary and requires further investigation.

Figure 4 - *Mean duration of gaze pattern window, for each gaze pattern types, when PR and IS vocal feedback are produced*



By-dyad values in Figure 5 show, again, some variability with respect to the observed duration of the gaze patterns preceding the different feedback functions, albeit less than for the occurrences of gaze pattern types (Fig. 3). Indeed, the averaged trends observed above hold in four dyads out of six (Bu-Vi; Dm-Se; Pr-Po; Ri-Ca, with the latter not presenting any instance of mutual gaze co-occurring with PR feedback, impeding a comparison across feedback functions). The other two dyads diverge from the general trends to varying extents. Ca-Dc shows shorter averted gaze before PR than IS feedback signals and longer unilateral gaze before PR than IS feedback signals, contradicting the general trend for both gaze patterns, while Pi-Zi differs from the trend only due to slightly longer unilateral gaze before PR than IS feedback signals. Importantly, looking at individual by-dyad values enables us to refine the observations on mutual gaze duration found in the averaged data. While averaged data (Fig. 4) display values under one second, by-dyad values are in all cases but one (Dm-Se for PR) above one second. Still, where data points are available for a comparison across feedback functions (Bu-Vi; Pr-Po), the statement that mutual gaze preceding PR feedback signals is longer than when a turn alternation (IS) is involved holds true.

Figure 5 - *Mean duration of gaze pattern window, for each gaze pattern types, when PR and IS vocal feedback are produced, pooled by dyad*



5. *Discussion*

Research question 1 assessed the relation between a specific gaze pattern and the production of turn-regulating vocal feedback. Our findings indicate that both averted and unilateral gaze were prevalent, with averted gaze being more frequently associated with PR, and unilateral gaze with IS. This somewhat unexpected result might be explained by the speakers' differential use of intonational cues for turn-taking purposes. Previous studies on backchannel intonation contours in Italian task-oriented dialogues (Savino, 2010, 2011, 2014; Sbranna et al., 2021; Sbranna et al., in press) have shown that continuers (PR vocal feedback) predominantly exhibit a rising intonation contour, whereas IS vocal feedback tokens are predominantly realised with a falling intonation. Therefore, the rising intonation characteristic of PR feedback production may be readily interpreted by interlocutors, thus reducing the need for additional visual confirmation. In contrast, the fact that IS vocal feedback co-occurred more frequently with unilateral gaze (than PR feedback) suggests that IS feedback may require additional visual confirmation to distinguish it from a possible falling PR vocal signal, and may serve as a request for joint attention to clarify the intention of a subsequent turn transition both vocally and visually.

We also observed that mutual gaze rarely co-occurs with vocal feedback signals, regardless of their turn-taking regulating function, confirming previous findings in a similar task-based study in German (Spaniol et al., 2023). This phenomenon might be attributed to the task-based setting and the presence of task materials, which necessitate visual attention to achieve the communicative goal of the conversation. It can be assumed that other interactional contexts – e.g. a free conversation without such a strong a visual competitor – might reveal different distributions of gaze direction and vocal feedback combinations than those emerging from our data.

On the other hand, other studies have also argued that mutual gaze is more closely associated with visual feedback in the form of head and hand gestures, rather than vocal feedback (Ferré, Renaudier, 2017; Bertrand et al., 2007). This might explain the low association between gaze with vocal feedback in our finding, although this hypothesis cannot be tested in the present study, as head and hand gestures are not analysed. Therefore, irrespective of the potential role of the task, this result is in line with previous findings, supporting the hypothesis that gestural acts are more readily followed by gestural responses.

Research question 2 aimed to examine whether the duration of a gaze pattern expedites the production of feedback utterance and potential subsequent turn alternation. Here, we observed longer averted gaze patterns preceding PR feedback production, whereas longer unilateral gaze patterns were observed preceding IS feedback production. One possible interpretation is that a prolonged averted gaze may signal a preference by the secondary speaker to maintain the listener role, visually indicating their unavailability for a turn transition and encouraging the primary speaker to continue. This situation may occur when one interlocutor describes their picture and the other participant is expected to listen carefully in order to complete the Tangram-game task. Conversely, when the secondary speaker intends to take the floor (i.e., when mostly longer unilateral gaze patterns occur), they may first look at the primary speaker to visually signal their intention. If this gaze is not reciprocated, resulting in a unilateral gaze, the secondary speaker eventually initiates the turn after a vocal feedback signal without having obtained visual attention. However, if the gaze is reciprocated, creating a period of mutual gaze, turn initiation after feedback occurs more rapidly than when no turn transition is intended after feedback (PR), possibly benefitting from visual contact as an additional channel to communicate the imminent turn alternation. This result contrasts with the findings of Kendrick et al. (2023), who observed that eye gaze does not speed up turn transitions.

Finally, these general trends appear to be relatively consistent across dyads, although the exact proportions and duration values exhibit substantial dyad-specific variability. This underscores the importance of exploring multimodal dynamics at the dyadic level to complement and refine interpretations based on averaged data. In essence, each interaction outcome is contingent upon the dynamics negotiated and co-constructed by the participants.

6. *Conclusion*

We explored the relationship between eye gaze patterns and turn-regulating vocal feedback (continuers/Passive Recipiency, turn-initiating feedback/Incipient Speakership) based on the occurrences and the time window of the gaze patterns, prior to and concurrent with the utterance of vocal feedback.

Our findings indicate that a binary classification of gaze patterns (mutual vs. averted) is not sufficient to capture the interplay between gaze and turn-regulating vocal feedback. In fact, despite averted gaze co-occurring with both types of vocal

feedback signals, unilateral gaze towards the interlocutor is found predominantly when a turn transition is imminent. We also observed that averted gaze is longer when preceding PR than when preceding IS. In contrast, unilateral gaze is longer before IS than before PR, suggesting that speakers may attempt to establish and wait for gaze contact before taking the turn, which they eventually indicate with a vocal feedback signal if their interlocutor fails to reciprocate their gaze. Finally, we found dyad-specific variability, as expected, due to the availability of two communication channels (vocal and visual), which can be used either in combination or independently. Moreover, it is worth remembering that each interaction can be regarded as a complex system resulting from the dynamics created by the participants in the conversation through a cycle of actions and responses, which inevitably results in unique detailed aspects beyond the more general common trends. Thus, we emphasise the importance of examining intra-dyadic details to obtain a more comprehensive understanding of interactional phenomena.

The application of this methodology to a larger sample of dyads and a comparative analysis with dialogue formats without visual competitors, will help further elucidate the phenomena underscored by these preliminary yet insightful results.

Acknowledgments

This research was funded by the Cluster Development Program Language Challenges, a funding line within the Excellent Research Support Program of the University of Cologne, and the German Research Foundation (DFG), grant number 281511265, SFB 1252 Prominence in Language.

Credit Author Statement

Simona Sbranna: Conceptualization, Data curation, Formal Analysis, Methodology, Visualization, Writing – original draft, Writing – review & editing.

Michelina Savino: Data curation, Investigation, Methodology, Resources, Writing – review & editing.

Florence Bailles: Methodology, Writing – review & editing.

Martine Grice: Methodology, Supervision, Writing – review & editing.

References

Adams, R.B. Jr., Kleck, R.E. (2005). Effects of direct and averted gaze on the perception of facially communicated emotion. In *Emotion*, 5, 3-11. doi: 10.1037/1528-3542.5.1.3

Amador-Moreno, C.P., Mccarthy, M. & O'keeffe, A. (2013). Can English Provide a Framework for Spanish Response Tokens?. In Romero-Trillo, J. (Eds.), *Yearbook of Corpus Linguistics and Pragmatics 2013*, vol 1. Springer, Dordrecht. https://doi.org/10.1007/978-94-007-6250-3_9

Argyle, M., Dean, J. (1965), Eye-contact, distance and affiliation. In *Sociometry*, 289-304, doi:102307/2786027

Auer, P. (2018). Gaze, addressee selection and turn-taking in three-party interaction. In Brône, G., Oben, B. (Eds.), *Eye-tracking in interaction: Studies on the role of eye gaze in dialogue*. John Benjamins, 197-232. http://doi.org/10.1075/ais.10.09aue

Barthel, M., Meyer, A.S. & Levinson, S.C. (2017). Next speakers plan their turn early and speak after turn-final "go-signals". In *Frontiers in psychology*, 8, 239095.

Bavelas, J.B., Coates, L. & Johnson, T. (2002). Listener responses as a collaborative process: The role of gaze. In *Journal of Communication*, 52, 566-580. https://doi.org/10.1111/jcom.2002.52.issue-3

Beattie, G.W. (1978). Floor apportionment and gaze in conversational dyads. In *British Journal of Social and Clinical Psychology*, 17, 7–15. http://doi.org/10.1111/j.2044-8260.1978.tb00889.x

Beattie, G.W. (1979). Contextual constraints on the floor-apportionment function of speaker gaze in dyadic conversation. In *British Journal of Social and Clinical Psychology*, 18, 391-392.

Bertrand, R., Ferré, G., Blache, P., Espesser, R. & Rauzy, S. (2007). Backchannels revisited from a multimodal perspective. In *Auditory-visual Speech Processing*, 1-5.

Bissonnette, V.L. (1993). *Interdependence in dyadic gazing* [Doctoral dissertation, The University of Texas at Arlington]. The University of Texas at Arlington ProQuest Dissertations Publishing.

Blythe, J., Gardner, R., Mushin, I. & Stirling, L. (2018). Tools of engagement: selecting a next speaker in Australian aboriginal multiparty conversations. In *Research on Language and Social Interaction*, 51, 145-170. http://doi.org/10.1080/08351813.2018.1449441

Brône, G., Oben, B., Jehoul, A., Vranjes, J. & Feyaerts, K. (2017). Eye gaze and viewpoint in multimodal interaction management. In *Cognitive Linguistics*, 28, 449-483. http://doi.org/10.1515/cog-2016-0119

Clancy, P.M., Thompson, S.A., Suzuki, R. & Tao, H. (1996). The conversational use of reactive tokens in English, Japanese, and Mandarin. In *Journal of Pragmatics*, *26* (3), 355-387.

Cummins, F. (2012). Gaze and blinking in dyadic conversation: a study in coordinated behaviour among individuals. In *Language and Cognitive Processes*, 27, 1525-1549. http://doi.org/10.1080/01690965.2011.615220

Degutyte, Z., Astell, A. (2021). The role of eye gaze in regulating turn taking in conversations: A systematized review of methods and findings. In *Frontiers in Psychology*, 12, 616471. http://doi.org/10.3389/fpsyg.2021.616471

Drummond, K., Hopper, R. (1993). Back channels revisited: Acknowledgment tokens and speakership incipiency. In *Research on Language and Social Interaction*, 26, 157-177. http://doi.org/10.1207/s15327973rlsi2602_3

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

Duncan, S. (1973). Toward a Grammar for Dyadic Conversation. In *Semiotica*, 9(1), 29-46. https://doi.org/10.1515/semi.1973.9.1.29

Duncan, S., Fiske, D.W. (1977). *Face-to-face interaction: Research, methods, and theory*. L. Erlbaum Associates.

Eberhard, K.M., Nicholson, H. (2010). Coordination of understanding in face-to-face narrative dialogue. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, *32*, 2051-2056. eScholarship, University of California. https://escholarship.org/uc/item/6bk8h630

Egbert, M.M. (1996). Context-sensitivity in conversation: Eye gaze and the German repair initiator bitte?. In *Language in Society*, 25(4), 587-612.

ELAN (Version 6.7) [Computer software]. (2023). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Retrieved from https://archive.mpi.nl/tla/elan

Fellegy, A.M. (1995). Patterns and functions of minimal response. In *American Speech International Journal of Educational Best Practices*, 2 (1), 186-199.

Ferré, G., Renaudier, S. (2017). Unimodal and Bimodal Backchannels in Conversational English. In *Proceedings of SemDial*, Saarbrücken, Germany, 27-37.

Foulsham, T., Cheng, J.T., Tracy, J.L., Henrich, J. & Kingstone, A. (2010). Gaze allocation in a dynamic situation: effects of social status and speaking. In *Cognition*, 117, 319–331. http://doi.org/10.1016/j.cognition.2010.09.003

Fries, C.C. (1952). *The structure of English*. London: Longmans, Green; Company.

Gardner, R. (2001). *When listeners talk: Response tokens and listener stance*. Amsterdam: John Benjamins Publishing Company.

Goodwin, C. (1981). *Conversational organization: Interaction between speakers and hearers*. Academic Press.

Hall, S. (1995). Review of the Observer 3.0. In *Journal of Child Psychology & Psychiatry*, 36,1495-1498.

Heldner, M., Hjalmarsson, A. & Edlund, J. (2013). Backchannel relevance spaces. In Asu, E.L., Lippus, P. (Eds.), *Nordic Prosody: Proceedings of the XIth Conference, Tartu 2012*. Peter Lang, 137-146. http://doi.org/10.3726/978-3-653-03047-1

Heritage, J. (1984). A change-of-state token and aspects of its sequential placement. In *Structures of Social Action: Studies in Conversation Analysis*, 299-345.

Hjalmarsson, A., Oertel, C. (2012). Gaze direction as a back-channel inviting cue in dialogue. In *IVA 2012 workshop on realtime conversational virtual agents* (Vol. 9). Citeseer.

Ho, S., Foulsham, T. & Kingstone, A. (2015). Speaking and listening with the eyes: gaze signaling during dyadic interactions. In *PLoS ONE., 10*, Article e0136905. http://doi.org/10.1371/journal.pone.0136905

Hollenstein, T. (2013). State Space Grids. In *State Space Grids*. Springer, Boston, MA. https://doi.org/10.1007/978-1-4614-5007-8_2

Holler, J., Kendrick, K.H. & Levinson, S.C. (2018). Processing language in face-to-face conversation: Questions with gestures get faster responses. In *Psychonomic bulletin & review*, 25(5), 1900-1908. https://doi.org/10.3758/s13423-017-1363-z

Hu, L., Chen, G. (2021). Trajectories of Idea Emergence in Dialogic Collaborative Problem Solving: Toward a Complex Dynamic Systems Perspective. In *Frontiers in Psychology*, *12*, 735534. https://doi.org/10.3389/fpsyg.2021.735534

Jacoby, S., Ochs, E. (1995). Co-Construction: An Introduction. In *Research on Language and Social Interaction*, 28(3), 171-183. https://doi.org/10.1207/s15327973rlsi2803_1

Janz, A. (2022). Navigating common ground using feedback in conversation – A phonetic analysis. Cologne, Germany: University of Cologne. (MA thesis).

Jefferson, G. (1983). Two explorations of the organization of overlapping talk in conversation: Notes on some orderlinesses of overlap onset. In *Tilburg Papers in Language and Literature, 28*, 1–28. https://liso-archives.liso.ucsb.edu/Jefferson/Onset.pdf

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, 12. http://doi.org/10.1145/2499474.2499481

Jokinen, K., Nishida, M. & Yamamoto, S. (2009). Eye-gaze experiments for conversation monitoring. In *Proceedings of the 3rd International Universal Communication Symposium*. ACM, 303-308. http://doi.org/10.1145/1667780.1667843

Jurafsky, D., Shriberg, E., Fox, B. & Curl, T. (1998). Lexical, prosodic, and syntactic cues for dialog acts. In Stede, M., Wanner, L. & Hovy, E. (Eds.), *Discourse Relations and Discourse Markers*. ACL Anthology, 114-120. https://aclanthology.org/W98-0319.pdf

Kendon, A. (1967). Some functions of gaze direction in social interaction. In *Acta Psychologica*, 26, 22-63. https://doi.org/10.1016/0001-6918(67)90005-4

Kendrick, K.H., Holler, J. & Levinson, S.C. (2023). Turn-taking in human face-to-face interaction is multimodal: gaze direction and manual gestures aid the coordination of turn transitions. In *Philosophical Transaction of the Royal Society B, 378*, Article 20210473. http://doi.org/10.1098/rstb.2021.0473

Kraut, R.E., Lewis, S.H. & Swezey, L.W. (1982). Listener responsiveness and the coordination of conversation. In *Journal of Personality and Social Psychology*, 43, 718-731.

Lambertz, K. (2011). Back-channelling: The use of yeah and mm to portray engaged listenership. In *Griffith Working Papers in Pragmatics and Intercultural Communication*, 4, 11-18.

Lerner, G.H. (2003). Selecting next speaker: The context-sensitive operation of a context-free organization. In *Language in Society*, 32, 177-201. http://doi.org/10.1017/S004740450332202X

Levinson, S.C., Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. In *Frontiers in Psychology*, 6, 731. https://doi.org/10.3389/fpsyg.2015.00731

Liang, J., Zou, Q., Liang, Y., Wu, W. & Yan, J. (2021). Emotional Gaze: The Effects of Gaze Direction on the Perception of Facial Emotions. In *Frontiers in Psychology*, 12. https://doi.org/10.3389/fpsyg.2021.684357

Lücking, A., Ptock, S. & Bergmann, K. (2011). Assessing agreement on segmentations by means of staccato, the segmentation agreement calculator according to thomann. In *International gesture workshop*. Berlin, Heidelberg: Springer Berlin Heidelberg, 129-138. https://doi.org/10.1007/978-3-642-34182-3_12

Myszka, T.J. (1975). Situational and intrapersonal determinants of eye contact, direction of gaze aversion, smiling and other non-verbal behaviors during an interview [Doctoral dissertation, University of Windsor]. Electronic Thesis and Dissertations University of Windsor. https://scholar.uwindsor.ca/etd/3477

Novick, D., Hansen, B. & Ward, K. (1996). Coordinating turn-taking with gaze. In *Proceedings of Fourth International Conference on Spoken Language Processing* (ICSLP-96). IEEE Xplore, 1888-91. http://doi.org/10.1109/ICSLP.1996.608001

Oertel, C., Włodarczak, M., Edlund, J., Wagner, P. & Gustafson, J. (2012). Gaze patterns in turn-taking. In Sproat, R. (Ed.), *Proceedings of Interspeech 2012*. ISCA, 2246-2249. http://doi.org/10.21437/Interspeech.2012-132

Ondáš, S., Kiktová, E., Pleva, M. & Juhár, J. (2023). Analysis of Backchannel Inviting Cues in Dyadic Speech Communication. In *Electronics*, 12, 3705. https://doi.org/ 10.3390/electronics12173705

Paulmann, S., Jessen, S. & Kotz, S.A. (2009). Investigating the multimodal nature of human communication. In *Journal of Psychophysiology*, 23, 63-76. https://doi.org/10.1027/0269-8803.23.2.63

Perniss, P. (2018). Why we should study multimodal language. In *Frontiers in Psychology*, *9*, 342098. https://doi.org/10.3389/fpsyg.2018.01109

Poppe, R., Truong, K.P. & Heylen, D. (2011). Backchannels: Quantity, type and timing matters. In *Intelligent Virtual Agents: 10th International Conference, IVA 2011, Reykjavik, Iceland, September 15-17, 2011. Proceedings 11*. Springer Berlin Heidelberg, 228-239.

Rasenberg, M., Pouw, W., Özyürek, A. & Dingemanse, M. (2022). The multimodal nature of communicative efficiency in social interaction. In *Scientific Report*, 12, Article 19111. https://doi.org/10.1038/s41598-022-22883-w

Rossano, F. (2012). Gaze behaviour in face-to-face interaction. PhD Dissertation, Radboud University, Nijmegen. https://hdl.handle.net/11858/00-001M-0000-000F-ED23-5

Rossano, F., Brown, P. & Levinson, S.C. (2009). Gaze, questioning and culture. In *Conversation Analysis, 27*, 187-249. http://doi.org/10.1017/CBO9780511635670.008

Rutter, D.R., Stephenson, G.M., Ayling, K. & White, P.A. (1978). The timing of looks in dyadic conversation. In *British Journal of Social and Clinical Psychology*, 17, 17-21. http://doi.org/10.1111/j.2044-8260.1978.tb00890.x

Sacks, H., Schegloff, E. & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. Lg. 50.696-735.

Savino, M. (2010). Prosodic strategies for backchanneling in Italian Map Task dialogues. In Botinis, A. (Ed.), *Proceedings of the ISCA Tutorial and Research Workshop on Experimental Linguistics (Exling 2010)*. ISCA, 157-160. http://doi.org/.36505/ExLing-2010/03/0040/000160

Savino, M. (2011). The intonation of backchannels in Italian task-oriented dialogues: cues to turn-taking dynamics, information status and speaker's attitude. In *Proceedings of the 5th Language and Technology Conference: "Human Language Technology as a Challenge for Computer Science and Linguistics*. Poznan (Poland), 25-27 November 2011, 370-374.

Savino, M. (2014). The intonation of backchannel tokens in Italian collaborative dialogues. In Vetulani, Z., Mariani, J. (Eds.), *Human Language Technology Challenges for Computer Science and Linguistics*, Springer LNAI Series n. 8387, 17-28. https://doi.org/10.1007/978-3-319-08958-4_3

Savino, M., Refice, M. (2013). Acknowledgement or reply? Prosodic features for disambiguating pragmatic functions of the Italian token 'sì'. In *Proceedings of the 7th International*

Conference on Speech Technology and Human-Computer Interaction (SpeD 2013). IEEE, 117-112. http://doi.org/10.1109/SpeD.2013.6682660

Savino, M., Lapertosa, L. & Refice, M. (2018). Seeing or not seeing your conversational partner: the influence of interaction modality on prosodic entrainment. In Karpov, A., Jokisch, O. & Potapova, R. (Eds.), *Speech and Computer, Proceedings of the 20th SPECOM International Conference*, Leipzig 18-22 September 2918, LNCS series n. 11096, Springer, 574-584. DOI: 10.1007/978-3-319-99579-3

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. In Frota, S., Vigário, M. (Eds.), *11th International Conference on Speech Prosody*. http://doi.org/10.21437/SpeechProsody.2022-149

Sbranna, S., Wehrle, S. & Grice, M. (in press). A multi-dimensional analysis of backchannels in L1 German, L1 Italian and L2 German. In *Language, Interaction, and Acquisition*. 15.2.

Schegloff, E.A. (1982). Discourse as an interactional achievement: Some uses of "uh huh" and other things that come between sentences. In Tannen, D. (Ed.), *Analyzing discourse: Text and talk, Georgetown University Roundtable on Languages and Linguistics* 1982, 71–93. Washington, DC: Georgetown University Press.

Schegloff, E.A. (2020). Interaction: The infrastructure for social institutions, the natural ecological niche for language, and the arena in which culture is enacted. In *Roots of human sociality*. Routledge, 70-96.

Simon, C. (2018). The functions of active listening responses. In *Behavioural Processes*, 157, 47–53.

Spaniol, M., Janz, A., Wehrle, S., Vogeley, K. & Grice, M. (2023). Multimodal signalling: The interplay of oral and visual feedback in conversation. In Skarnitzl, R., Volín, J. (Eds.), *Proceedings of the 20th International Congress of Phonetic Sciences*. Guarant International, 4110-4114.

Stahl, G. (2016). *Constructing dynamic triangles together: The development of mathematical group cognition*. Cambridge: Cambridge University Press.

Streeck, J. (2014). Mutual gaze and recognition. Revisiting Kendon's Gaze direction in two-person conversation. In Seyfeddinipur, M., Gullberg, M. (Eds.), *From Gesture in Conversation to Visible Action as Utterance*. John Benjamins, 35–55. http://doi.org/10.1075/z.188.03str

Thomann, B. (2001). Observation and judgment in psychology: Assessing agreement among markings of behavioral events. In *Behavior Research Methods, Instruments, & Computers*, 33, 339-348.

Truong, K.P., Poppe, R., Kok, I.D. & Heylen, D. (2011). A multimodal analysis of vocal and visual backchannels in spontaneous dialogs. *Proc. Interspeech 2011*, 2973-2976, DOI: 10.21437/Interspeech.2011-744

Wehrle, S. (2023). *Conversation and intonation in autism: A multi-dimensional analysis*. Studies in Laboratory Phonology 14. Berlin: Language Science Press. DOI: 10.5281/zenodo.10069004

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A. & Sloetjes, H. (2006). *ELAN: a Professional Framework for Multimodality Research*. In *Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation*.

Yngve, V. (1970). On getting a word in edgewise. In *Papers from the sixth regional meeting, Chicago Linguistic Society*, 567-577. Chicago: Chicago Linguistic Society.

Zellers, M., House, D. & Alexanderson, S. (2016). Prosody and hand gesture at turn boundaries in Swedish. In *Speech Prosody*, 831-835.

Zima, E., Weiss, C. & Brône, G. (2019). Gaze and overlap resolution in triadic interactions. In *Journal of Pragmatics*, 140, 49-69.