frontiers | Frontiers in Communication

TYPE Original Research
PUBLISHED 30 October 2025
DOI 10.3389/fcomm.2025.1655049



OPEN ACCESS

EDITED BY
Loulou Kosmala,
Université Paris-Est Créteil Val de Marne, France

REVIEWED BY
Gaëlle Ferré,
University of Poitiers, France
Auriane Boudin,
Aix-Marseille Université, France

*CORRESPONDENCE
Marlene Böttcher
✉ mboettcher@isfas.uni-kiel.de

RECEIVED 27 June 2025
ACCEPTED 30 September 2025
PUBLISHED 30 October 2025

CITATION
Böttcher M and Rossi M (2025) The speaker's "okay" vs. the listener's "okay": exploring lexical, phonetic, and multimodal variation of backchannels and fluencemes in conversation. *Front. Commun.* 10:1655049. doi: 10.3389/fcomm.2025.1655049



# The speaker's "okay" vs. the listener's "okay": exploring lexical, phonetic, and multimodal variation of backchannels and fluencemes in conversation

Marlene Böttcher[1]* and Martina Rossi[2]

[1]Department of Linguistics and Phonetics, Institute for Scandinavian Studies, Frisian Studies and General Linguistics, Kiel University, Kiel, Germany, [2]Department of Humanities (DISTUM), Urbino University, Urbino, Italy

This research explores the use of fluencemes and backchannels in German face-to-face conversations. Fluencemes are produced by the current speaker to facilitate speech planning and to structure the speaking turn, while backchannels are produced by the current listener to signal acknowledgment and understanding of what is being said. These two conversational devices are constituted by very short utterances, often sharing the same lexical form (e.g., "ja," "okay," "genau" in German); however, they display significant differences based on their function within the interaction. This study compares the distribution of backchannels and fluencemes within the conversational turn, their acoustic form, as well as their interaction with multimodal resources, across and within dyads. We find that, while the two devices share lexical candidates, their duration, F0 and pitch contour, as well as the frequency and type of co-occurring head movements vary depending on whether the item is produced by the speaker or by the listener.

KEYWORDS

**fluencemes, backchannels, feedback signals, prosody, head movements, individual variation**

## 1 Introduction

In spontaneous interaction, the smooth flow of the conversation is the result of the joint efforts of both interlocutors (considering two-party interactions) involved in the exchange. By taking turns with each other, one interlocutor will hold the conversational floor as the current speaker, while the other one listens. In other words, one interlocutor will be occupied in "speaking activities", and one in "listening activities" (Yngve, 1970, p. 568), and both will manage the floor making sure that long silent gaps and stretches of overlapped speech will be kept appropriately low (Sacks et al., 1974; Stivers et al., 2009).

Both as a current speaker and current listener, the conversational participant has been observed to make use of conversational devices that are very short in form, do not add any new propositional content to the interaction, and facilitate turn-taking (Knudsen et al., 2020). Specifically, the current speaker may produce utterances such as "hm", "uh", "so", "okay" in order to facilitate upcoming speech planning and to hold the turn while minimising silences; in other words to signal active speakership. On their end, the listener also may make use of short utterances such as "mhm", "yes", "exactly" in order

to signal acknowledgment and understanding of what is being said in a non-interrupting way, without claiming the turn; in other words to signal active listenership.

Within the speaker's turn, these conversational devices are defined as *fluencemes*, while listener's short utterances are defined as *backchannels*. The current study explores both conversational devices in face-to-face two-party interactions in German and provides an overview of their lexical and acoustic forms, distribution within the conversation, their interaction with multimodal resources, as well as their variation across dyads. The main objective of this research is to highlight how these short utterances in conversation, in spite of sharing a few similarities in terms of their lexical forms, display striking differences in relation to the function that they have in conversation, i.e., either signaling active speakership or active listenership.

In more detail, the analysis will investigate several aspects: the distribution and frequency of fluencemes and backchannels within turn-taking structures across and within dyads, the acoustic variation of the same lexical (or non-lexical) forms when used as either backchannels or fluencemes by comparing their duration, mean fundamental frequency (F0) and pitch countour (i.e., F0 slope), and the distribution of head movements that co-occur with them.

## 1.1 Backchannels

The term *backchannel* (Yngve, 1970) refers to listener's responses produced during the current speaker's turn to display active listenership without interrupting or claiming the conversational floor. In the literature, backchannels have also been labeled as "listener feedback" or "supportive feedback" (Stubbe, 1998), "minimal responses" (Fellegy, 1995) and "acknowledgment tokens" (Jefferson, 1984) *inter alia*. When the conversation takes place face-to-face, backchannels are multimodal, i.e., they are produced both through the verbal channel, in the form of short utterances, and the visual channel, through gestures, and they are crucial for the smooth progression of conversations.

Although backchannels used to be considered optional for the development of the conversation and they are less constrained in timing than full conversational turns (Ward and Tsukahara, 2000; Roberts et al., 2015), both speakers and listeners are sensitive to their appropriate frequency, form and placement within the speaker's turn (e.g., Wehrle et al., 2023). For instance, Bavelas et al. (2000) found that the reduced use of content-relevant backchannels in dyadic interactions leads to increased disfluency and less effective storytelling. Similarly, excessively frequent or infrequent backchannels can become noticeable and irritating to the speaker (e.g., Heinz, 2003; Krause-Ono, 2004). In a study involving a virtual agent, Poppe et al. (2011) report that random distributions of verbal backchannels, and too sparse or too frequent occurrences, reduced perceived naturalness. Perceived engagement ratings also decreased when backchannels were delayed by more than one second (Boudin et al., 2024). Moreover, Wehrle et al. (2018) report that, in a perception task, German subjects judged as polite and acceptable backchannels that were produced with a rising or falling intonation, while those that had a flat contour were perceived negatively. Further investigations on German interactions from the ALICO corpus (Buschmeier et al., 2014) report that backchannels produced by attentive listeners are louder and show greater intonation variability (Malisz et al., 2012) than feedback produced by distracted listeners, who also tend to use more head movements than verbal tokens (Włodarczak et al., 2012). These findings highlight the importance of the form, modality, and frequency of backchannels for maintaining conversational flow.

In German, verbal backchannels constitute 16% of all turns in spontaneous conversation, with an average of 15% across West Germanic languages (German, English, Dutch; Knudsen et al., 2020). In German task-based dialogues, backchannel rate has been reported to average between 5.82 and 9.2 backchannels per minute, with a high degree of variability among dyads (Wehrle, 2021; Sbranna et al., 2024). As across languages, they consist of either monosyllabic or disyllabic non-lexical vocalisations, such as "mhm,", "aha" or brief lexical items like "ja" (yes), "genau" (exactly), "okay", "eben" (right), "achso" (I see), or "das stimmt" (that's right) (Knudsen et al., 2020; Liesenfeld and Dingemanse, 2022; Rossi et al., 2023; Wehrle, 2021). The prosodic form of German backchannels has been found to be closely related to their lexical composition, as well as their function. Specifically, investigating non-lexical and lexical backchannels in German dialogues, Wehrle (2021), Sbranna et al. (2022) and Sbranna et al. (2024) report that "okay" and "genau" are mostly produced with a falling intonation contour, "ja" tends to be rising, as is the non-lexical "mhm". As Wehrle and Grice (2019) observe for "mm(hm)", its rising intonation as a backchannel among German speakers could be implemented to distinguish it from the same non-lexical item used as a filled pause, for which the intonation is predominantly level.

In the visual modality, movements of the head, and in particular head nods, have been identified as the most prevalent form of backchannel (Cerrato, 2012; McClave, 2000; Paggio and Navarretta, 2011), both in isolation or accompanying verbal forms (Dittmann and Llewellyn, 1968). Besides nods (up and down movement), several other head gestures have also been observed to occur as backchannels, such as turns (left and right movement), tilts (top of the head goes in one direction and the chin in the opposite), slides (horizontal movement left and right) and protrusions (forward or backwards movement; Wagner et al., 2014; Rohrer et al., 2020; Rossi et al., 2023). Like verbal backchannels, head movements have been reported to coordinate turn-taking, signal acknowledgment and understanding of what is being said, and encourage the current speaker to continue (Cerrato and Skhiri, 2003; McClave, 2000), and are widely used by listeners in conversation. For instance, in German dialogues, Rossi et al. (2023) report that the 30% of the backchannels they identify are constituted by a head gesture on its own, and 42% of the lexical backchannels and the 56% of the non-lexical ones are accompanied by head movements.

Visual feedback signals are generally considered less disruptive to the speaker's current turn than verbal ones (Dittmann and Llewellyn, 1968; Ferré and Renaudier, 2017). This is evident in perception studies, such as that by Poppe et al. (2011), where head movements, even when randomly distributed, were judged as inappropriate less often than verbal backchannels. In fact, regarding their distribution within the current speaker's turn, results from previous studies show that verbal backchannels tend to arise after the offset of speech, during a silence, while

backchannels constituted by head movements tend to occur during speech, as they are less disruptive than verbal ones for the current stream of talk (Dittmann and Llewellyn, 1968; Ferré and Renaudier, 2017; Rossi et al., 2023; Truong et al., 2011). Rossi et al. (2023) also report a finer distinction in distribution between lexical and non-lexical backchannels in German: while the former occurs statistically more often during pauses, non-lexical expressions, which are shorter and/or less loud, are found both in overlap with the current turn or after its offset.

Variations in the frequency and distribution of feedback within conversation have been found to be influenced by social factors. For instance, different languages have specific frequency baselines for backchannels (e.g., Stubbe, 1998; Cutrone, 2011) and, within languages, studies reported some differences in the backchanneling style of male and female speakers. Some studies on gender-specific behavior in conversation report that women use backhannels more frequently, both verbally and visually (Habalet, 2019; Helweg-Larsen et al., 2004; Kjellmer, 2009; Ueno, 2004), and time them to reduce silent gaps during the interaction (Beňuš et al., 2007; Krepsz et al., 2022). However, findings on gender-specific language patterns are not always consistent and hardly generalisable, as they appear to be closely connected to the specific context in which the conversation takes place, as well as the individuals involved (Plug et al., 2021).

## 1.2 Fluencemes

The transition of speakers within conversations tends to occur with very short silences or gaps (Levinson, 2016). These gaps are reported to be shorter in natural conversations than, for example, in picture-naming tasks (Knudsen et al., 2020). This smooth transition between turns is made possible by yet another group of small discourse items, namely fluencemes. Following Götz (2013), *fluencemes* are features of speech that enable speech fluency. A broad definition of fluencemes covers temporal aspects like speech rate and lengthening, and distributional aspects like discourse markers and filler particles. Fluencemes in the latter sense can be considered short items or phrases produced during speech with no primary contribution to the propositional content of the exchange, but rather providing time for speech planning and signaling active speakership (Crible et al., 2017; Knudsen et al., 2020). The use of fluencemes in the beginning of new turns allow for enough time for the speaker to plan the upcoming speech, while also already holding the floor and ensuring small gaps between turns. The current research focuses on distributional fluencemes and includes discourse markers and filler particles in the analysis.

Prior work in this field has made a distinction between discourse markers like "well", "so", "like" in English, or "also", "genau" and "okay" in German as lexicalised particles in speech, and filler particles like "uhm" in English or "ähm" in German (also called *filled pauses* or *hesitation markers*) usually within the context of disfluency (Bortfeld et al., 2001). Both are highly polyfunctional, semantically bleached and contribute to the discourse or turn transition organisation (Crible, 2017; Fischer, 2000). Following Crible et al. (2017), we look at discourse markers and filler particles together as contributors to the speakers' fluency, and therefore to smooth turn transitions.

The most frequent fluencemes in German include both the filler particles "äh" (uh), "ähm" (uhm), "hm" as well as the discourse markers "also" (so) and "okay", "genau" (exactly) and short reply particles "ja" (yes), "nein" (no) and "nee" (nope) along with response items like "ach" and "oh" (Diewald, 2013; Sbranna et al., 2024).

Overall, fluencemes make up 10% of spoken words in spontaneous speech (Özsoy and Blum, 2023; Shriberg, 2001). The relative amount of fluencemes depends on the discourse context. In an investigation of fluencemes in a corpus of Turkish narrations, Özsoy and Blum (2023) found the majority of items in utterance initial position, and fewer instances of fluencemes produced utterance medially or finally. Complementary, in their analysis of West Germanic languages, Knudsen et al. (2020) report 2-3% of utterances beginning with a filler particle while 6-15% of utterances start with a discourse marker. In utterance initial positions, fluenceme use can be interpreted as speech planning, topic initiating, discourse structuring devices, similar to fluencemes in turn inital positions (Swerts, 1998; Staley and Jucker, 2021; Rendle-Short, 2004; Fraser, 2009; Maschler and Schiffrin, 2015). Research on dyadic conversations showed an average of 3.6 fluencemes per minute (Wehrle, 2021), only focusing on filler particles), but also revealed high degree of inter-dyad variation, similar to research on backchannel frequency (see above).

Prosodically, fluencemes are reported to be produced with a reduced phonetic form (Lee et al., 2020). This appears in the form of unstressed realisations when the item functions as a discourse marker, as opposed to its lexical use (Brinton, 2017), and is often accompanied by plateaued or declining pitch contours (Lee et al., 2020). However, depending on the discourse function and the lexical item, the shape of the pitch contour might vary (Ferré, 2011; Lee et al., 2020; Sbranna et al., 2024). A comparison of backchannel and fluenceme use of the same lexical items in German dyadic conversations, including "ja", "genau" and "okay", found a tendency for falling contours in the case of fluencemes in contrast to more rising pitch contours in the case of backchannels of the same lexical items (Sbranna et al., 2024). Similarly, filler particles are produced with level or declining pitch contours, while additionally being produced below the speakers mean fundamental frequency (F0; Belz and Reichel, 2015). While the research on multimodal backchannel use is well established, multimodal features of fluencemes are still relatively underexplored. Existing studies focus only on specific items, e.g., filler particles, or specific lexical items used as discourse markers. The few existing studies report contrasting findings: in some cases fluencemes appear to be associated with gesture hold phases or rest positions, but do not necessarily co-occur with manual gestures (Esposito et al., 2001; Betz et al., 2023; Kosmala, 2022). Research on head movement in overlap with fluencemes similarly does not provide clear co-occurrence patterns (Baiat et al., 2013; Ferré, 2011).

# 2 Objectives

Based on the research presented above this study explores the lexical form, prosodic aspects and the overlap with head

movements of backchannels and fluencemes in German dyadic conversations. As presented in Section 1.1 and 1.2, both backchannels and fluencemes are typically short non-lexical or short lexical items. These two conversational devices also share an inventory of lexical items in German, such as, for instance, "ja", "okay" and "genau".

Considering their distinct function in conversation as either listener or speaker resources, we expect to find differences in their location related to the speaker's turns, their prosodic shape and their multimodal use. As head movements are the most common gesture with a feedback and a discourse-structuring function, this first exploration focuses on their distribution accompanying conversational particles.

Backchannels are expected to be produced frequently in overlap with the speech by the current speaker, mostly with rising pitch contour, and with a variety of head gestures, especially head nods. Fluencemes, on the other hand, are expected to be realised as part of speaker's turn and with level or falling intonation. As prior work on fluencemes has focused on their verbal form and few studies report on their multimodal features, the current analysis of co-occurring head movements represents an exploration of the interaction of fluencemes and gestures.

Since prior work has predominantly focused on either of the two phenomena, this paper sets out to compare the distibution, lexical and phonetic form backchannels and fluencemes and their tendency to overlap with head gestures in the same set of speakers.

## 3 Materials and methods

We analysed 7 face-to-face conversations involving 14 German adult native speakers, taken from the German sub-corpus of the *DUEL Multi-lingual Multimodal Dialogue Corpus* (Hough et al., 2016). The participants, in pairs, sat across from each other, each equipped with close-range lapel microphones to enable high-quality individual audio capture, including during segments of overlapping speech; dual camera recordings were configured to capture both head movements and the manual gesture space (Hough et al., 2016). This multimodal recording setup allowed us to annotate and analyse the speech and the head movements of participants when they had the speaker's role and when they were listeners. Participants were asked to carry out an interactional task which was designed in order for the conversational partners to start speaking without having to spend too much time selecting a topic, while at the same time allowing them to interact freely, allowing a spontaneous dialogue to arise. Specifically, in the subset we investigated, participants were either involved in the "Dream Apartment" interactional scenario, or the "Film Script" one. In the former, the pair was asked to imagine and discuss the layout and furnishing of a shared apartment, having a large sum of money at their disposal; in the latter, participants were asked to imagine an embarrassing movie scene, drawing from personal experiences (Hough et al., 2016; Kousidis et al., 2013).

The subset used for this investigation included 5 same-gender conversations—3 female-female (FF) and 2 male-male (MM)—and 2 mixed-gender ones, involving one male and one female participant (MF). The speakers of 5 dyads knew each other before recording for two years or longer while the speakers of two dyads (02_MM and 18_FF) indicated low familiarity and knew each other for less than 12 months. Mean age of speakers was 23.4 years (sd = 2.7).

Within the first 5 minutes of each of the conversations verbal backchannels and fluencemes were identified in Praat (Boersma and Weenink, 2024) using the orthographic transcription of the utterances provided by the DUEL Corpus, and isolated on separate tiers within interval annotations. The analysed data set comprises 39 minutes of speech.

In the analysis the *speaker* is considered the conversational participant who holds the current speaking turn while the *listener* is identified as the conversational participant who does not hold the current speaking turn. We considered as backchannels all the short utterances produced by the listener who does not take up the turn by producing a short conversational particle. Possible backchannel candidates are defined as a direct reaction to the content of the speaker's turn, optional (i.e., answers to direct questions are not considered as backchannels), and not acknowledged by the current speaker (i.e., questions and turn-opening vocalisations are not considered as backchannels; Rossi et al., 2023; Ward and Tsukahara, 2000). We considered as fluencemes all the short utterances that structure the discourse or indicate speech planning which are produced by the speaker turn medial or turn finally or as a means to take up the turn by producing a short conversational particle turn initially. This definition includes items like filler particles and discourse markers uttered by one of the participants. The fluencemes considered in this research are semantically bleached or empty, syntactically not integrated and prosodically phrased separately.

Following these guidelines allowed us to obtain a clear distinction between what constituted a backchannel or a fluenceme, in spite of the two categories often sharing lexical forms. For instance, German "okay" is reported to occur both as a feedback response and as a discourse structuring token (Oloff, 2019), and could potentially be considered as an ambiguous case. However, using the guidelines that we defined, we were able to clearly label a token either as a backchannel or a fluenceme. For example, when "okay" was produced by the listener, with no additional speech right before or after it, letting the other interlocutor continue speaking, it was considered as a backchannel; on the other hand, when "okay" arose at the beginning of a full-fledged conversational turn, in the middle or at the end of it, meaning that was preceded, surrounded or followed by more speech by the same speaker, it was considered as a fluenceme. Moreover, our distinction was established by first identifying backchannels and fluencemes in the data, and then determining their lexical forms. Shared forms did not present ambiguities, as they emerged directly from the data rather than from a top-down classification.

After isolating all phenomena that corresponded to the above described guidelines, we defined the annotation labels of backchannels and fluencemes based on their lexical form. Even though part of the analysis will be focused on the most frequent forms that emerged from the data, we did not exclude any item during the annotation and the exploration phases. Following Zellers' (2021) labels for backchannels, we annotated both backchannels and fluencemes as: *mlx* for monosyllabic lexical

items (e.g., "ja", yes), *dlx* for disyllabic items (e.g., "genau", exactly), *non* for non-lexical expressions (e.g., "mh" or "mhm"), *cplx* for lexical phrases (such as "das stimmt", that's right), and *cmbn* for the different combinations of the previous forms. Head movements were manually identified and annotated in ELAN (Wittenburg et al., 2006) using video files only, without audio. The Multimodal and M3D Multidimensional Labeling Scheme (Rohrer et al., 2020, 2023), based on Wagner et al. (2014)'s classification, provided the movement type tier (nod, slide, tilt, protrusion). All instances of head movements that arose were annotated. The labels were then imported into Praat on a separate "Head" tier. Co-occurring head movements were extracted starting from the vocal element. Specifically, when the gesture annotation interval overlapped the interval of the verbal backchannel or fluenceme, it was extracted as a co-occurring head movement.

The annotation was carried out by both authors. Specifically, one focused on the annotation of backchannels and one on the annotation of fluencemes. Both authors then jointly discussed individual cases and labels that could potentially pose ambiguities.

The rate of conversational devices is calculated as items per minute of dialogue for each of the dyadic conversations in line with prior work on phonetic aspects of backchannels and fluencemes by Sbranna et al. (2022, 2024) and Wehrle (2021), Wehrle et al. (2023, 2024). The relevant distributional and prosodic aspects of the annotated items were extracted in Praat using a script. The pitch range was set to 60–400 Hz for male speakers and to 100–600 Hz for female speakers. F0 was measured in semitones (ST) in the beginning and in the end of the annotated interval considering the central 80% of the item, as well as the mean of the whole annotated interval. In cases where no F0 measurements were obtained in the beginning or the end, F0 was measured in the cental 70–60% of the item [see method described in Wehrle (2021)]. Excursion was calculated as the difference in F0 between the measurement in the beginning and the end. An excursion size larger than 1 ST is considered falling and an excursion size smaller than −1 ST is considered rising. The script additionally extracted the duration of the annotated items, as well as the overlap with head movements. We also extracted the possible overlap of the item with the other interlocutor's speech. Specifically, if the start of the backchannel or the fluenceme arose in overlap with speech by the other interlocutor, the item was labeled as "turn internal"; if the start of the item arose during a silent gap, it was labeled as "turn external".

Data analysis and visualisation was carried out in R Studio (R Core Team, 2023; Posit team, 2023) using the ggplot2 package (Wickham, 2016).

# 4 Results

## 4.1 Distribution of backchannels and fluencemes

### 4.1.1 Frequency of backchannels and fluencemes

Overall there are 634 conversational devices in the analysed dataset of which 151 classify as backchannels (BCs) and 483 as fluencemes (FLs). As illustrated in Figure 1 not only the absolute number of BCs is lower compared to FLs, but also the rate per minute is lower, with a mean of 2.3 BCs per minute of dialogue (sd = 0.9) compared to a mean FL rate of 7.3 per minute of dialogue (sd = 2.8).

A closer look at the rates of BCs and FLs within and across dyads reveals individual variation, presented in Figure 2. For some dyads (i.e., 16_MM and 18_FF) the rates of both BCs and FLs are very similar for both dyad participants. In general, it appears that the frequency of BCs tends to remain similar between conversational partners, with no striking differences. On the other hand, it is more common for FLs to show bigger differences in rate between speakers within the same dyad (i.e., 13_MF, 10_FF, 15_FF). There is a tendency for female interlocutors to produce higher rates of BCs ($\bar{X}$ = 2.2 per min, sd = 0.7) and FLs ($\bar{X}$ = 6.6 per min, sd = 2.6) compared to male speakers (BCs: $\bar{X}$ = 1.5 per min, sd = 1.0; FLs: $\bar{X}$ = 5.7, sd = 3.2). Yet, dyad 13_MF shows the reverse pattern, with more FLs produced by the male speaker compared to his female interlocutor.

A Welch's t-test shows that the difference in rate per minute between BCs and FLs in general is significant ($t = -5.47$, $p < 0.001$, 95% CI [−6.01, −2.65], Cohen's $d = -2.07$). The difference in BC and FL rate between male and female conversational participants, however, is not significant.

The general distribution across speaker turns shows expected pattern with 66% of BCs produced turn internally, i.e. during the speaker's turn and in overlap with the speech by the current speaker (see Table 1). This is illustrated in example (1) where person B is narrating a scene and person A produces the backchannel "ah ja" (ah yes) in overlap with the narration. The relevant items of backchannels and fluencemes in the examples (1)–(5) are highlighted in bold, overlaps of speech are indicated by [] for the respective speakers A and B and silent pauses are indicated in () providing the duration in brackets (following the guidelines in Selting et al., 2009).

(1) dyad 10_FF (72 s into the conversation of planning a funny film script)

B sie fragt ob er die pfandflaschen weggebracht hat und er sagt so ja, hab ich gemacht [und dann]
"She asks whether he has returned the reusable bottles and he goes 'yes, I did' "[and then]"

A [**ah ja**]
"[**ah yes**]" (backchannel)

B macht sie den schrank auf und dann kommen ihr die ganzen pflandflaschen entgegen
"she opens the cupboard and all the reusable bottles fall out"

Fluencemes, on the other hand, are produced predominantly turn externally (73%), i.e. without any overlaps. In example (2) the BCs "ja" (yeah) are again produced turn internally and in overlap with the interlocutor speech, whereas the fluenceme "ähm" (uhm) is produced after a silence and is also followed by silence of 500 ms, before speaker A takes up the floor again and continues talking about his idea.

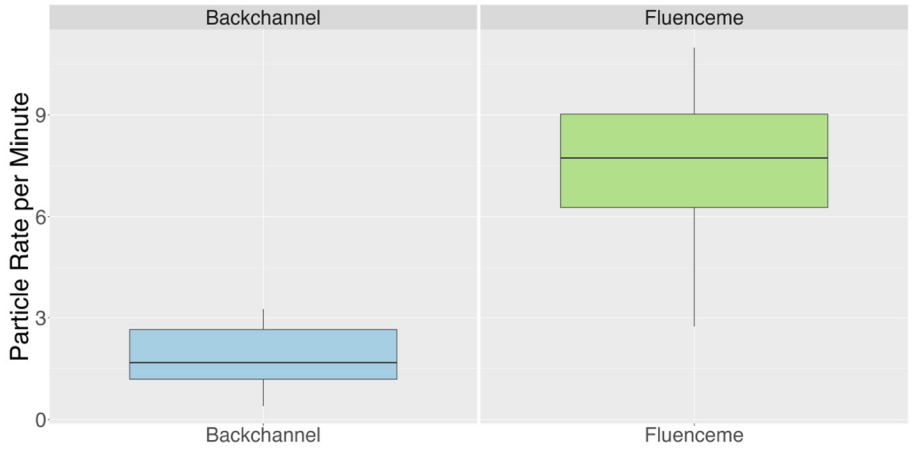


FIGURE 1
Rates per minute of dialogue for backchannels and fluencemes.

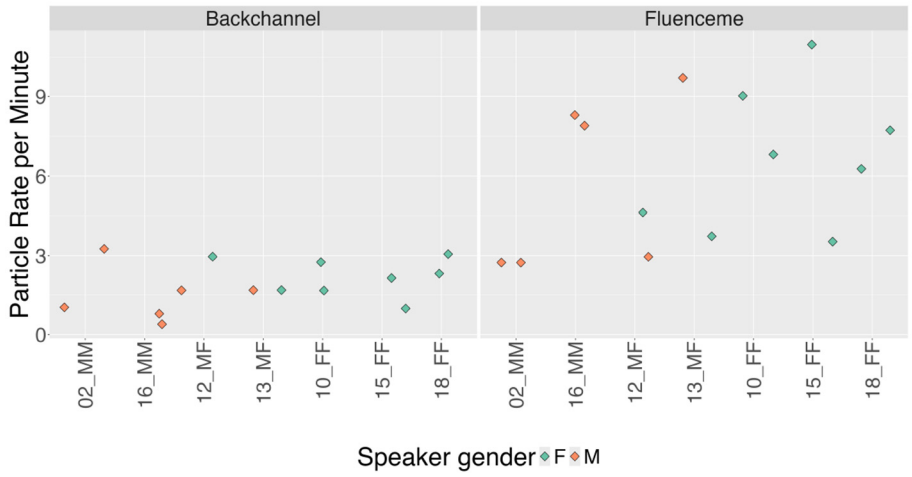


FIGURE 2
Rates per minute of dialogue for backchannels and fluencemes produced by speakers within the respective dyad. Dyads are ordered according to gender of the interlocutors (male = M, female = F), with mixed gender dyads in the middle.

(2) dyad 02_MM (36 s into the conversation, of planning a dream apartment)

B [**ja**]
"[**yeah**]" (backchannel)
A [zum] vorsaufen oder sowas [<laughter]>
"[to] get drunk or so (laughing)"
B [**ja**]
"[**yeah**]" (backchannel)
A (0.5) **ähm**
"(0.5) **uhm**" (fluenceme)
A (1.2) und von daher müsste das wohnzimmer aber ich hab da jetzt gar keine dimension was man da (0.9) wie groß
"(1.2) and so the livingroom would need to be but I don't know about dimensions what one (0.9) how large"

In some cases, FLs overlap the speech by the other interlocutor. Example (3) illustrates such a case where the two interlocutors speak in overlap, and speaker A produces the fluenceme "äh" (uh) to negotiate who will continue, and succeeds in taking the floor. After making a suggestion of introducing an embarrassing situation, speaker B jumps in and produces the fluenceme "so" (like) again in overlap, to make a suggestion of her own. In this case, FLs are produced turn internally, i.e. in overlap, and are used as turn managing devices to negotiate the floor. This is possible both while the other interlocutor is still talking to signal the intention to take the next turn, or to negotiate if both speakers start talking in

**TABLE 1** Distribution of backchannels and fluencemes produced turn internally i.e., during interlocutors' turns or turn externally as independent turns.

| Particle type | Location | n | Percent |
|---|---|---|---|
| Backchannel | Turn external | 52 | 34.00 |
| | Turn internal | 99 | 66.00 |
| Fluenceme | Turn external | 353 | 73.00 |
| | Turn internal | 130 | 27.00 |

overlap. In these cases a fluenceme was being produced within the turn or at the end of the speaker's turn.

(3) dyad 10_FF (132 s into the conversation of planning a funny film script)

B [die postkarte vielleicht]
"A postcard maybe."
A [aber es muss ja irgendwas **äh**]
"But it needs to be something **uhm**" (fluenceme)
A sein wo man sich vielleicht mit seiner dummheit blamiert (0.8) du brauchst [irgendein irgend]"
"an embarassing stupid situation(0.8) you need [something some]"
B [**so** (.) die mu]tter
"**Like** (.) the mother" (fluenceme)

In some cases, BCs might be used to negotiate speaker turns, such as when they occur turn externally, i.e., during a silent gap, as they can be employed to refuse the speaking turn. In both examples (4) and (5) the BCs "mhm" and "ja" (yes) are produced during silences and are part of the negotiation for who will take the next turn. In both these cases, it is the current speaker, not the person producing the backchannel, who will take up the next turn. In these cases backchannels are used to signal that the channel remains open, or to refuse the conversational floor (Cutrone, 2005; Ward and Tsukahara, 2000).

(4) dyad 02_MM (133 s into the conversation of planning a dream apartment)

A das badezimmer sollte auf jeden fall ne badewanne noch haben (-) zusätzlich zum zum zu der dusche
"The bathroom definitely needs a bathtub (-) additionally to the the the shower."
B (0.6) **mhm**
"(0.6) **mhm**" (backchannel)
A (0.9) wenn wir das einrichten
"(0.9) When we decorate it."

(5) dyad 12_MF (299 s into the conversation of planning a funny film script)

B wir haben ja zehn minuten zeit wir können uns ganz viele Ideen ausarbeiten
"We have plenty of time to work on our ideas"
A (.) **ja**
"**Yes**" (backchannel)
B (0.7) und dann am ende brainstorm welche
"And we can brainstorm which one in the end."

A $\chi^2$ test confirms the existence of a relationship between the conversational device type and the distribution within the interaction ($\chi^2 = 72.803$, df = 1, $p < 0.001$). The Pearson residuals for the test indicate that BCs are observed significantly more frequently turn internally (r = 6.02) and less frequently turn externally (r = −4.52), while FLs arise significantly more frequently than expected turn externally (r = 2.53) and less frequently turn internally (r = −3.36).

As can be seen in Figure 3 this tendency holds across dyads, with some variation. While female speakers in female-female (FF) dyads produce BCs almost exclusively turn internally ($n = 62$, 81%) and FLs turn externally ($n = 170$, 66%), male speakers in same-gender dyads produce BCs equally turn externally ($n = 17$, 55%) and turn internally ($n = 14$, 45%) but show the same preference for FLs in turn external position ($n = 88$, 77%). In mixed dyads the number of BCs produced turn externally is increased in female speakers ($n = 13$, 52%), and reduced in male speakers ($n = 7$, 39%) while the preference for turn external FLs holds for these dyads as well. A $\chi^2$ test indicates the existence of a significant relationship between BCs and the gender of the listener ($\chi^2 = 5.87$, df = 1, $p < 0.05$), with no significant deviations from the expected values indicated by the Pearson residuals, and between FLs and the gender of the speaker ($\chi^2 = 8.15$, df = 1, $p < 0.01$), with male participants producing FLs less frequently than expected in turn internal location (r = −1.99).

As illustrated in the example (4) and (5) above, in these cases BCs are produced between speaker turns to negotiate the next turn. Overall, FLs are more frequently used in this turn position, i.e., turn externally, yet, female speakers in FF dyads show higher numbers of FLs produced in overlap with an interlocutor's turn.

### 4.1.2 Lexical type

In the whole dataset BCs and FLs were most frequently monosyllabic lexical items (*mlx*; $n = 221$, 35%) or non-lexical forms (*non*; $n = 186$, 29%), followed by disyllabic lexical items (*dlx*; $n = 123$, 19%) and complex utterances (*cplx*, $n = 32$, 5%). Combinations of items of different form make up to 11% of the analysed dataset (*cmbn*, $n = 72$). The combinations of items are most frequently combinations of other forms with *mlx* ($n = 43$, 59%) or combinations of multiple *mlx* forms ($n = 20$, 28%). Of these combinations with *mlx*, those with *dlx* ($n = 24$) are more frequent than combinations with *non* ($n = 11$) or combinations with *cplx* ($n = 8$). There are also a few instances of combinations of *dlx* and *non* ($n = 5$) and of multiple *dlx* ($n = 2$) or multiple *non* ($n = 2$). Table 2 presents an overview of occurrences of these lexical types across BCs and FLs along with a list of examples for each category. In general, BCs are most frequently produced in *mlx* form, while FL are most frequently *non* or *mlx*. A $\chi^2$ test reveals a significant relationship between type of conversational device and the lexical form ($\chi^2 = 19.03$, df = 2, $p < 0.001$). In particular, Pearson residuals indicate that *mlx* are significantly more likely to occur as BCs in our data (r = 1.98) and less likely to occur as FLs (r = −1.11) while *non* are significantly more likely to occur als FLs (r = 1.2) and less likely as BCs (r = −2.15).

The distribution of lexical type of the two conversational devices is comparable across BCs and FLs, both turn internally and turn externally. Turn internally, i.e., during interlocutors' turns, combinations of BCs are more frequent compared to turn externally, where *mlx* are more frequent. FLs, on the other hand, are more frequently *mlx* during interlocutors' turns compared to turn externally, where *non* are produced more often (see Figure 4). Only for FLs, a $\chi^2$ test reveals a significant association between lexical type and location ($\chi^2 = 20.61$, df = 2, $p < 0.001$).

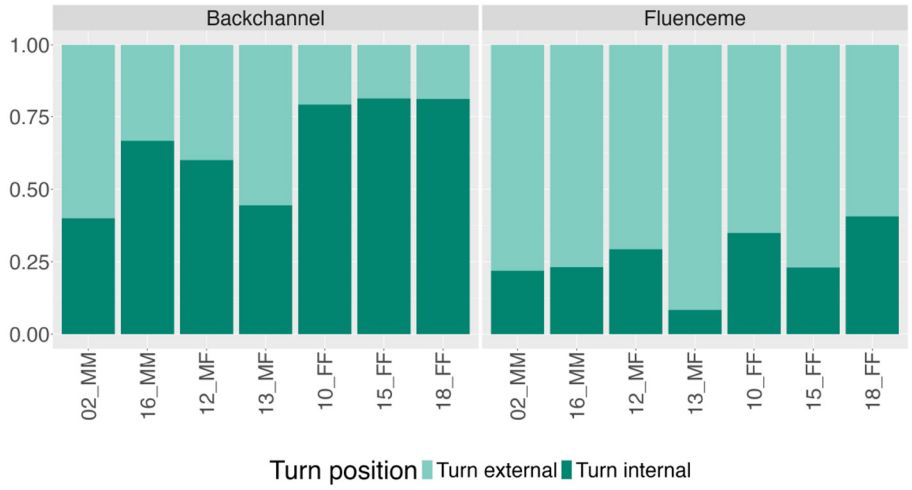


FIGURE 3
Proportion of backchannels and fluencemes produced within the interlocutor turn (turn internal) or turn externally within the respective dyads ($n$ = 634). Dyads are ordered according to gender and gender of the interlocutor (male = M, female = F) with mixed gender dyads in the middle.

**TABLE 2** Different lexical types of backchannels and fluencemes as non-lexical (non), monosyllabic lexical (mlx) and disyllabic lexical items (dlx), complex items like phrases (cplx) and combinations of items (cmbn) along with a non-exhaustive list of examples presented both in raw numbres ($n$) and percentages (percent).

| Particle type | Lexical type | Examples | n | Percent |
|---|---|---|---|---|
| Backchannel | non | *mhm, hm, ach* | 30 | 20.00 |
| | mlx | *ja, doch, gut* | 67 | 44.00 |
| | dlx | *okay, genau, bestimmt, achso* | 21 | 14.00 |
| | cplx | *alles klar, natürlich nicht, naja klar* | 8 | 5.00 |
| | cmbn | *ah ja, achso okay, ja eben* | 25 | 17.00 |
| Fluenceme | non | *äh, ähm, hm, oh* | 156 | 32.00 |
| | mlx | *ja, ne, so* | 154 | 32.00 |
| | dlx | *also, okay, genau* | 102 | 21.00 |
| | cplx | *ich meine, keine ahnung* | 24 | 5.00 |
| | cmbn | *ja genau, achso äh, also halt, ich weiß nicht* | 47 | 10.00 |

Specifically, Pearson residuals indicate that *mlx* and *non* occur respectively more and less frequently than expected turn internally (r = 3.02; r = −2.54).

The higher use of *mlx* FLs turn externally and the preference for *non* FLs turn internally can also be observed in the distribution of lexical types across the different dyads presented in Figure 5. For BCs, dyads show more internal variation than for FLs, especially for female speakers (FF) as they tend to produce a variety of lexical types as BCs independent of the location within the current turn. Male speakers in MM dyads predominantly use *mlx* forms also independent of the location within the current turn. The variation in BC forms employed by female speakers adds to the variation in the mixed-gender dyads. Yet, in these dyads, the use of *non* as a BC is more frequent compared to the same-gender dyads. In spite of these observed differences in BC form, a $\chi^2$ test did not show significant associations between BC's lexical type and gender. On the other hand, we found a significant relationship between FL's lexical type and gender ($\chi^2 = 10.05$, df = 2, $p$ <0.01), with *dlx* forms being used less frequently than expected by male speakers (r = −2.004).

### 4.1.3 Most frequent lexical items

The most frequent lexical items are "ja" ($n$ = 159), "okay" ($n$ = 30) and "genau" ($n$ = 18) for both BCs and FLs. The most frequent *non* backchannel "mhm" ($n$ = 26), the most frequent *non* fluenceme "äh(m)" ($n$ = 135), and the FLs "also" ($n$ = 65) account for 69% of the data (see Table 3). The "other" categoy includes the remainder of FLs and BCs including single instances of lexical items and also combinations of frequent items and phrases. Many of these occurred only once in their individual sequence produced by one of the speakes in a dyad.

A closer look at the use of these lexical forms across the analysed dyadic conversations in Figure 6 reveals that "genau" and "okay" are used as BCs only in the three dyadic conversations with female speakers (FF). In the mixed gender dyad 13_MF, both the female and the male speaker produce "okay" and "genau" as backchannel forms. Male-male dyads almost exclusively employ the backchannel "ja" and also show less variation in fluenceme form, sticking predominantly to "ja" and "ähm".

## 4.2 Prosodic form of backchannels and fluencemes

To investigate the acoustic differences of BCs and FLs, this study analyses their duration and F0, as well as pitch contour. In 12 items of the overall dataset no reliable pitch measurements were

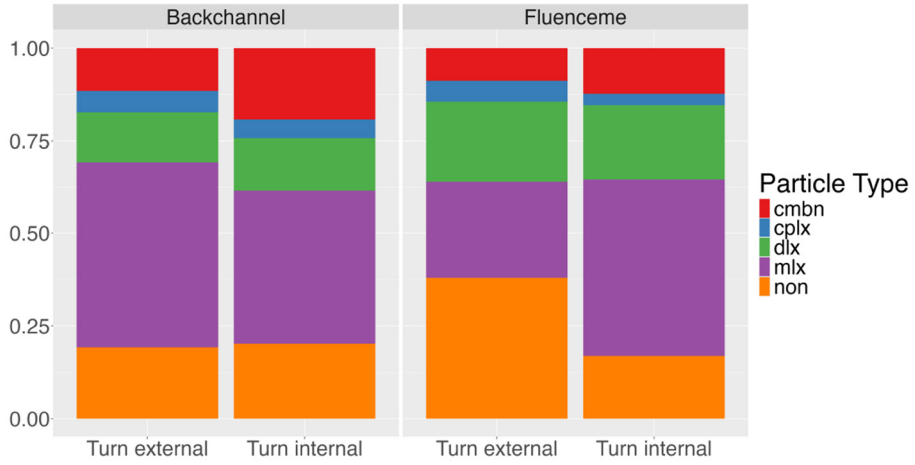


FIGURE 4
Type of conversational device produced turn internally, i.e. during interlocutors' turns, or turn externally as independent turns ($n = 634$). Types include non-lexical (non), monosyllabic (mlx), disyllabic (dlx), phrases (cplx) and combinations (cmbn).

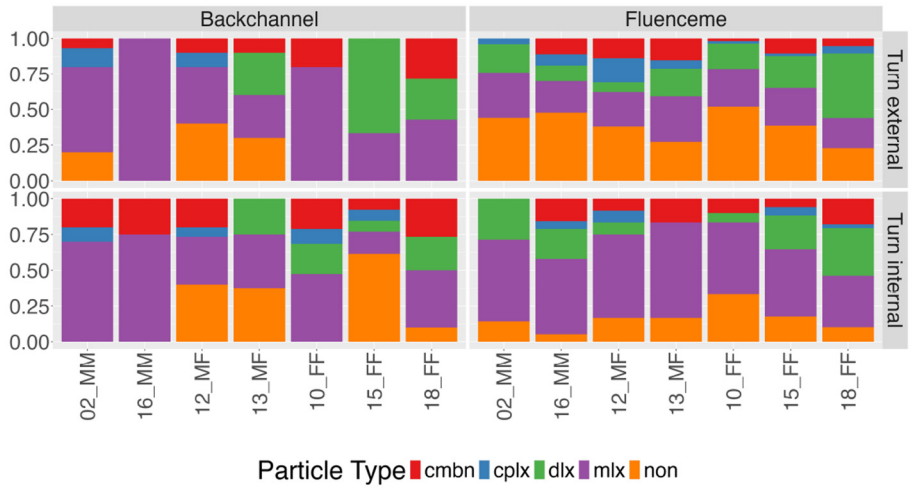


FIGURE 5
Type of conversational device produced turn internally, i.e. during interlocutors' turns, or turn externally as independent turns within the dyads ($n = 634$). Types include non-lexical (non), monosyllabic (mlx), disyllabic (dlx), phrases (cplx) and combinations (cmbn). Dyads are ordered according to gender and gender of the interlocutor (male = M, female = F) with mixed gender dyads in the middle.

possible. The analysis of F0 presented below, therefore, includes the remaining 622 items. In the whole dataset, *mlx* items are the shortest, with a mean duration of 280 ms. As it can be expected, disyllabic items are produced with a longer mean duration, i.e., 308 ms. The non-lexical items share a similar syllable structure to the monosyllabic items, yet they are produced with a longer mean duration of 395 ms. Combinations of forms and complex items (both categories comprise several items or whole phrases) measure even longer, with mean durations of 648 ms and 548 ms respectively.

As the variation in segmental strings might influence the phonetic form of the items, the following analysis only includes the most frequent lexical forms of BCs and FLs. As presented in Table 4, the mean duration of the monosyllabic "ja" items and the disyllabic "also" items are comparable in length, while the monosyllabic non-lexical items compare in length to the disyllabic lexical items.

Considering the different functions of the forms in conversation, the data presented in Figure 7 reveals that items used as FLs are overall shorter in duration (overall mean of lexical conversational devices: $\bar{X} = 0.27$ s, sd = 0.11) than the same form used as a BC (overall mean: $\bar{X} = 0.34$ s, sd = 0.13). The only exception is represented by the non-lexical FL "äh(m)" ($\bar{X} = 0.39$ s), which tends to be longer than the non-lexical BC "mhm" ($\bar{X} = 0.32$ s). A Welch's t-test shows that the difference in duration is significant between the disyllabic forms used as either BC or FL ($t = 4.84$, $p < 0.001$, 95% CI [0.07, 0.19], Cohen's $d = 1.442$), between the monosyllabic "ja" as either BC or FL ($t = 2.97$, $p < 0.001$, 95% CI [0.0209, 0.1041], Cohen's $d = 0.486$), and between the *non* forms as BC or FL ($t = -3.46$, $p < 0.001$, 95% CI [−0.105, −0.028],

Cohen's d = −0.42). Including gender, we observe a significant difference in the duration of the BCs "ja" and "mhm", which are longer when produced by female listeners ("ja": t = 3.18, $p$ <0.05, 95% CI [0.0303, 0.193], Cohen's d = 1.022; "mhm": t = 2.32, $p$ <0.05, 95% CI [0.004, 0.118], Cohen's d = 1.065).

A closer look at the individual dyads and the mean duration of lexical items within them in Figure 8 reveals some variation in duration, yet, confirms the overall tendency of items being longer when produced as BC and shorter as FL. That is, the mean duration of "okay" produced as a FL in dyad 02_MM ($\bar{X}$ = 0.45 s, sd = 0.10) is longer than duration of "okay" produced as BC in dyad 13_MF ($\bar{X}$ = 0.38 s, sd = 0.09). Yet, "okay" produced as a FL in dyad 13_MF ($\bar{X}$ = 0.30 s, sd = 0.03) is shorter than the same item as BC.

Comparing the F0 measured within the items (Figure 9) provides a less clearly distinct pattern, but suggests lower F0 within FLs compared to BCs of the same lexical type. That is, non-lexical "mhm" used as BC tends to have a higher F0 ($\bar{X}$ = 89.05 ST, sd = 6.76) compared to non-lexical "äh(m)" used as FL ($\bar{X}$ = 88.09 ST, sd = 4.39). A similar tendency can be seen with "okay" (see Table 4) Additionally, in general, disyllabic items seem to be produced with a higher F0 compared to the monosyllabic items. However, these qualitative differences in F0 between BCs and FLs do not turn out to be statistically different in a Welch's t-test. When gender is included in the t-test, all BC and FL forms display significantly higher F0 values for female speakers than for male speakers, which can be related to anatomical factors.

Tapping into individual variation within dyads in Figure 10 shows gender differences, with higher F0 in conversational devices produced by female speakers compared to male speakers, while the lower F0 in FLs compared to BCs shows more variance across dyads and does not hold for all speakers or all items.

Within the analysed 622 items F0 measurements were not possible in the beginning ($n$ = 67) and/or the end ($n$ = 55) resulting in missing values for F0 contour in 106 cases. Linear trajectories of F0 within the remaining 528 were predominantly falling F0 ($n$ = 237), while level contours ($n$ = 160) and rising contours were less frequent ($n$ = 131). The falling contour is used with similar frequency across BCs ($n$ = 38, 43%) and FLs ($n$ = 25,44%). Rising contours are more frequently associated with BCs ($n$ = 29, 33%) compared to FLs ($n$ = 68, 24%), while level contours arise more often with FLs ($n$ = 93, 33%) compared to BCs ($n$ = 21, 24%). A $\chi^2$ test did not show any significant association between the conversational device type and the contour trajectory.

Comparing pitch contours (see Figure 11), we can see that the non-lexical form confirms the general pattern, with a higher percentage of rising contours when used as BC ("mhm"), while the intonation is predominantly level or falling when used as FL ("äh(m)"). Similarly, "okay" is produced with a rising or level contour when it is a BC, and more frequently falling when used as a FL. However, "ja" displays a different pattern, as shown in Figure 11: in fact, it appears to be predominantly falling both

TABLE 3 Occurrence of most frequent lexical types for both backchannels and fluencemes.

| Particle type | Lexical type | Lexical item | n | Percent |
|---|---|---|---|---|
| Backchannel | non | mhm | 26 | 17.00 |
| | mlx | ja | 60 | 40.00 |
| | dlx | genau | 7 | 5.00 |
| | dlx | okay | 10 | 7.00 |
| | other | | 46 | 31.00 |
| Fluenceme | non | äh(m) | 135 | 28.00 |
| | mlx | ja | 99 | 20.00 |
| | dlx | genau | 11 | 2.00 |
| | dlx | okay | 20 | 4.00 |
| | dlx | also | 65 | 13.00 |
| | other | | 153 | 32.00 |

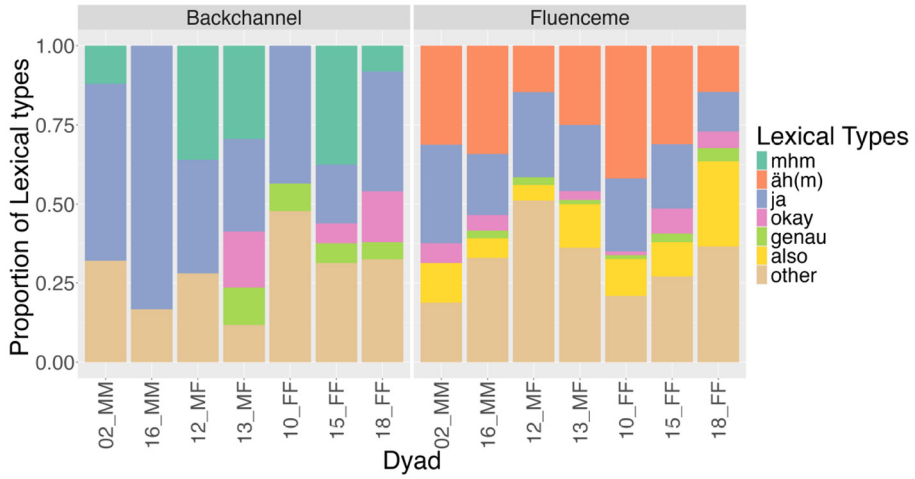


FIGURE 6
Distribution of most frequent lexical types of backchannels and fluencemes in the analysed dataset produced within dyads ($n$ = 433). Dyads are ordered according to gender and gender of the interlocutor (male = M, female = F) with mixed gender dyads in the middle.

TABLE 4 Mean duration and F0 along with the standard deviations (sd) for the most frequent lexical types of backchannels and fluencemes.

| Particle type | Lexical item | Lexical type | Mean duration (s) | sd (duration) | Mean F0 (St) | sd (F0) |
|---|---|---|---|---|---|---|
| Backchannel | mhm | non | 0.32 | 0.06 | 89.05 | 6.76 |
| | ja | mlx | 0.32 | 0.12 | 87.81 | 5.28 |
| | genau | dlx | 0.44 | 0.12 | 89.97 | 3.82 |
| | okay | dlx | 0.39 | 0.11 | 91.64 | 3.74 |
| Fluenceme | äh | non | 0.35 | 0.15 | 87.85 | 4.42 |
| | ähm | non | 0.45 | 0.19 | 88.42 | 4.36 |
| | ja | mlx | 0.26 | 0.13 | 88.31 | 4.79 |
| | genau | dlx | 0.34 | 0.14 | 89.35 | 3.32 |
| | okay | dlx | 0.33 | 0.09 | 89.99 | 5.05 |
| | also | dlx | 0.24 | 0.07 | 88.84 | 4.53 |

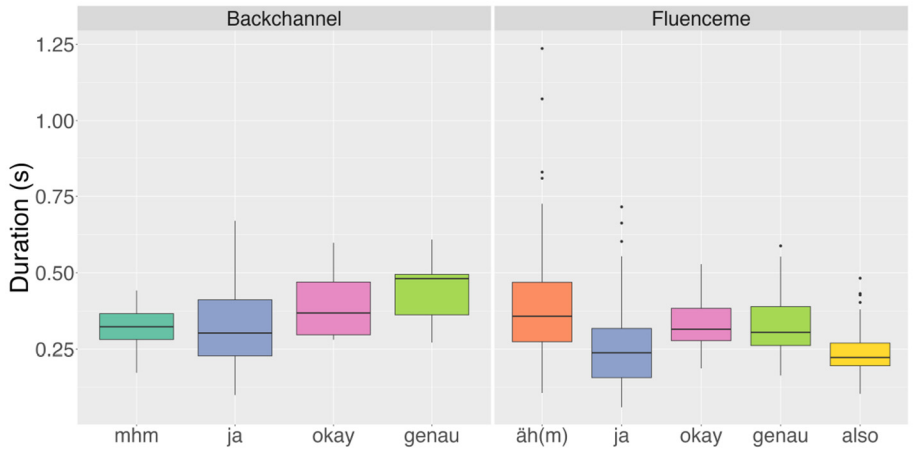


FIGURE 7
Duration of the most frequent lexical types of backchannels and fluencemes in the analysed data ($n = 433$).

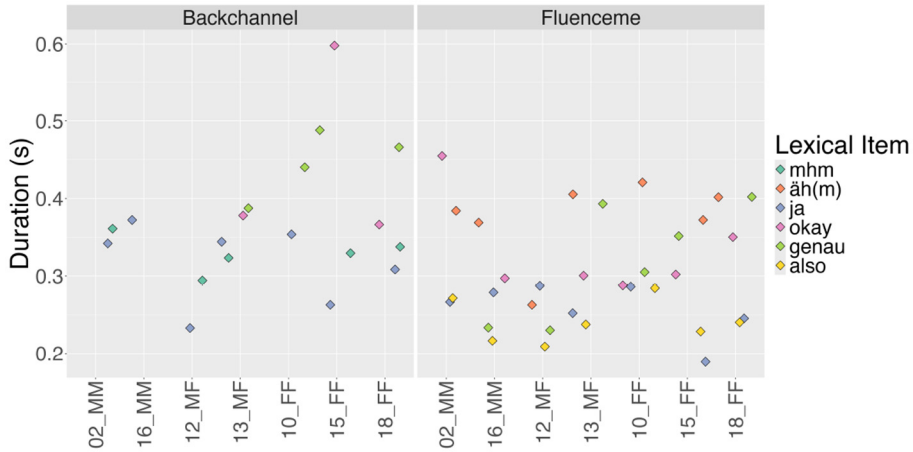


FIGURE 8
Mean duration of the most frequent lexical types of backchannels and fluencemes in the analysed data acoss dyads ($n = 433$). Dyads are ordered according to the gender of the interlocutors (male = M, female = F), with mixed gender dyads in the middle.

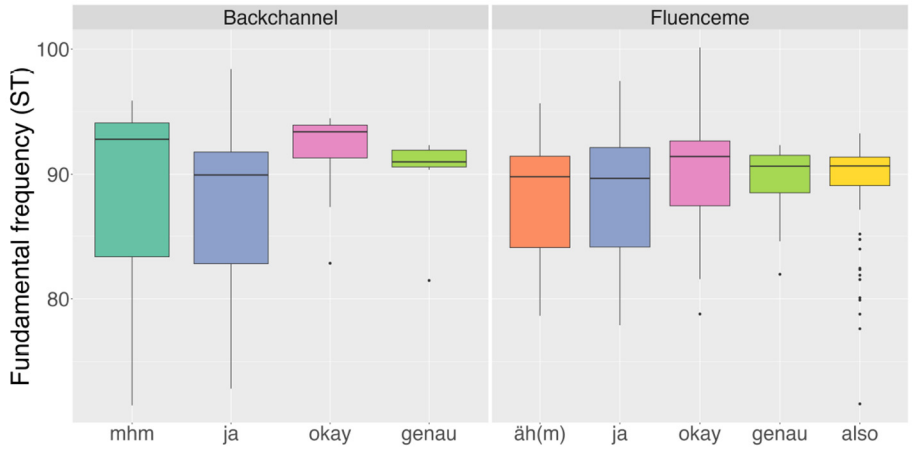


FIGURE 9
Mean F0 measured within the most frequent lexical types of backchannels and fluencemes in the analysed data ($n = 433$).

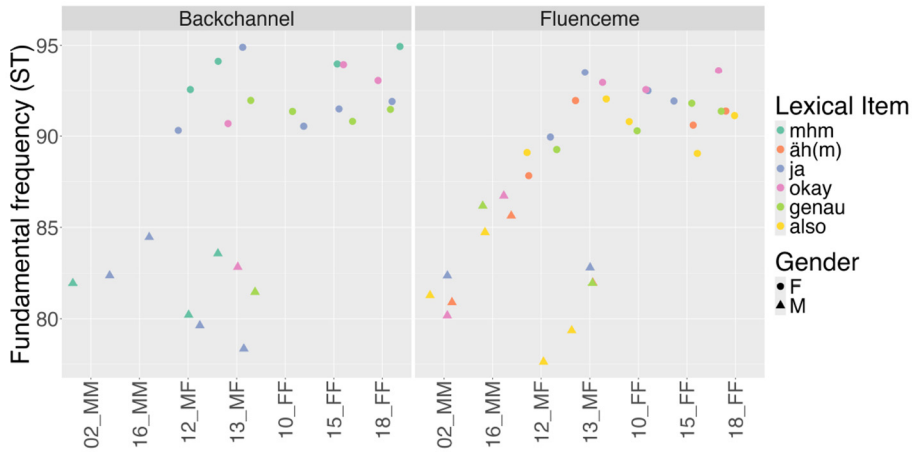


FIGURE 10
Mean F0 of the most frequent lexical types of backchannels and fluencemes in the analysed data across dyads ($n = 433$). Dyads are ordered according to the gender of the interlocutors (male = M, female = F), with mixed gender dyads in the middle.

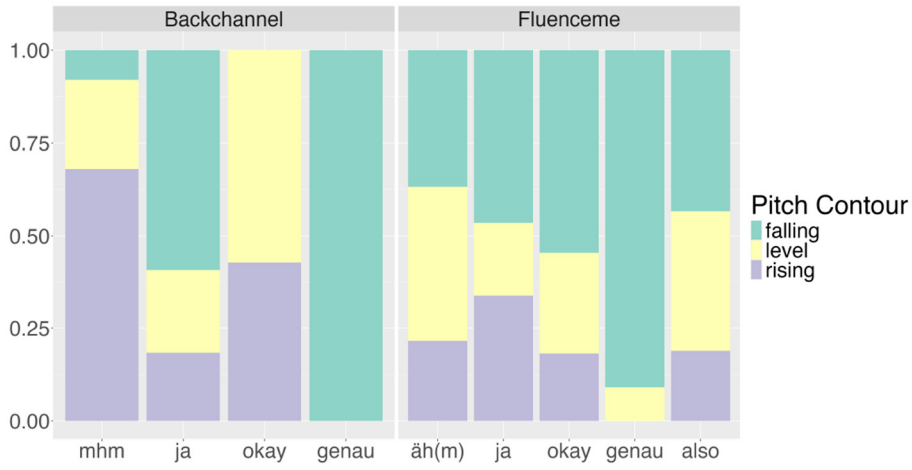


FIGURE 11
Pitch contours classified based on linear trajectories between beginning and end of the most frequent lexical types of backchannels and fluencemes in the analysed data ($n = 374$).

as a BC and a FL, with slightly higher percentages of level contours when it is used as a FL. Similarly, "genau" is produced with a falling intonation irrespective of its interactional function within the conversation. A $\chi^2$ test confirms the existence of a relationship between the lexical form and the contour type of BCs ($\chi^2 = 23.46$, df = 4, $p$ <0.001) and FLs ($\chi^2 = 15.12$, df = 4, $p$ <0.01). For BCs, "mhm" occurs less frequently with a falling contour than expected (r = −2.69) and more frequently with rising intonation (r = 3). For FLs, "ja" is less likely to have a level intonation than expected (r = −2). No significant relationships between gender and contour type emerges from our data.

## 4.3 Co-occurring head movements

Of the 634 conversational devices, 91 were produced with co-occurring head movements, of which the most frequent ones were nods ($n = 51$), followed by tilts ($n = 17$), and fewer instances of turns ($n = 11$), protrusions ($n = 9$), and slides ($n = 3$). BCs are produced with a co-occurring head movement ($n = 46$, 30%) more frequently than FLs ($n = 45$, 9%).

Considering the most frequent lexical items in overlap with head gestures, Figure 12 shows that BCs are almost exclusively accompanied by nods, while FLs overlap with a variety of head gestures, including nods. The relationship between the conversational device type and the head movement type is significant ($\chi^2 = 16.106$, df = 4, $p$ <0.01).

This variation in overlapping head gestures with FLs can also be seen by looking at the individual dyads (see Figure 13). On the other hand, BCs are predominantly produced with head nods across dyads. The dyad 02_MM stands out, as the two speakers in this dialogue show a high degree of variation in the types of head movement that they use with BCs. Notably, in this dyad, nods are not the predominant head movement type with BCs. No significant relationship between

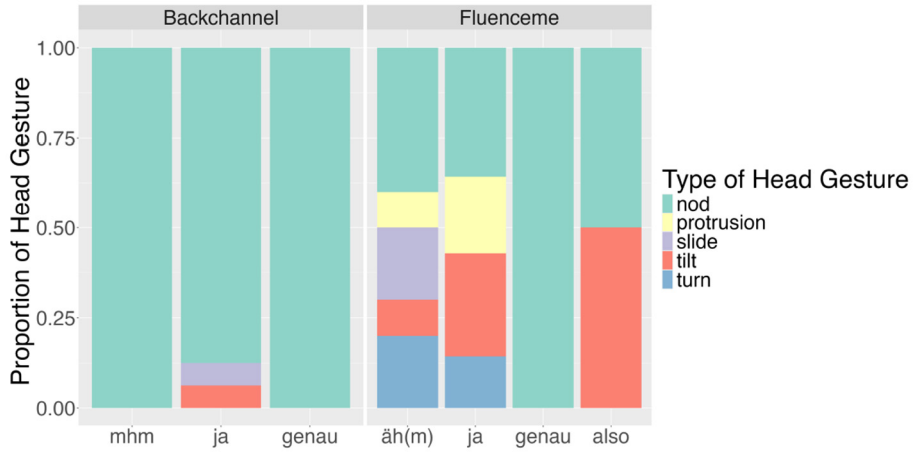


FIGURE 12
Type of head movements produced in overlap with the most frequent lexical types of backchannels and fluencemes in the analysed data ($n = 56$).

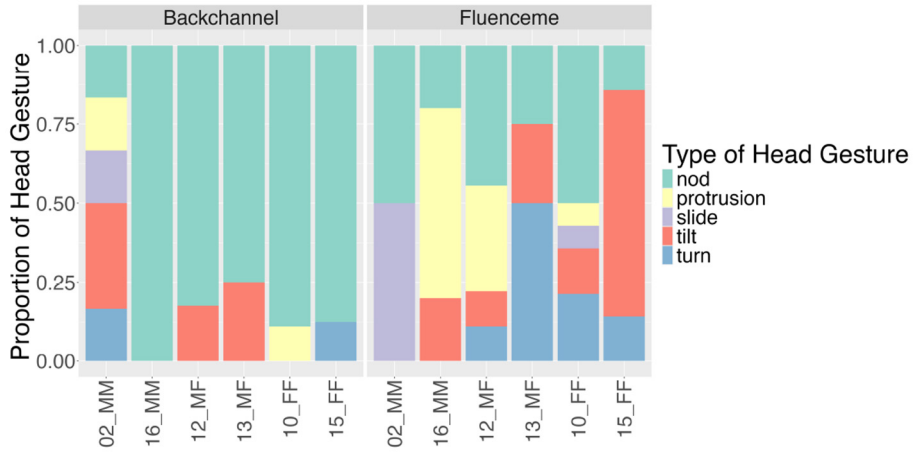


FIGURE 13
Type of head movements produced in overlap with backchannels or fluencemes across dyads ($n = 56$). Dyads are ordered according to the gender of the interlocutors (male = M, female = F), with mixed gender dyads in the middle.

gender and head movement type was observed for either BCs or FLs.

## 5 Discussion

In this paper, we analysed German face-to-face conversation to explore the variation of backchannels and fluencemes. In our analysis we focused on two conversational devices that share lexical forms, but that have different functions within the conversation. Our main aims were to explore their distribution across and within dyads, and to observe whether the different interactional functions of the same items are reflected in their acoustic form (i.e., duration, F0 and pitch contour) and by the use of multimodal resources (i.e., co-occurring head movements). Specifically, we observed that backchannels tend to have a significantly lower rate per minute than fluencemes (2.3 vs. 7.3 occurrences per minute of dialogue). Backchannels tend to be produced while the current speaker is talking, while fluencemes tend to not involve speech overlaps with the other interlocutor. In terms of their form, shorter items are generally preferred: the most common items for both conversational devices were the monosyllabic and non-lexical ones, followed by disyllabic forms. In particular, the most frequent conversational devices that emerged from this dataset are "ja", "okay", "genau", "mhm" as backchannels and, "ja", "okay", "genau", "äh(m)" and "also" as fluencemes. Comparing the phonetic configuration of these most frequent forms, we find that monosyllabic items are the shortest in duration, while non-lexical ones are the longest. By comparing the two groups, we observed that, when they function as a fluenceme, the same forms are significantly shorter than when they occur as a backchannel, with the exception of non-lexical items, which are longer as fluencemes. The differences in F0 found across backchannels and fluencemes are not statistically significant and seem to be an artifact of speakers' biological sex. In terms of their intonation contour, backchannels tend to be more frequently rising, while fluencemes are mostly level. In particular, we find that the rising contour is mostly observed for the non-lexical backchannel "mhm", while the fluenceme "äh(m)" mostly arises with a level or falling intonation; similarly, the intonation of "okay" is mostly rising or level as backchannel, while falling as a fluenceme; 'ja" and "genau" tend to display a falling intonation for both conversational devices. Finally, we find that, in this dataset, backchannels are more frequently accompanied by a head movement compared to fluencemes. For backchannels, the head movement is predominantly a nod, while fluencemes are accompanied by a wider variety of gestures.

The frequency of backchannels in the analysed data is lower and the rate per minute is also inferior than that reported by other studies (Sbranna et al., 2022, 2024; Wehrle et al., 2023). While this could be resulting from individual variation for the analysed dyads, the fact that this study only looked at 5 minutes per dyad might have skewed our results. Another aspect related to the frequency of backchannels is familiarity. While this was not discussed in the current paper, all the conversational partners except in one dyad were already familiar with each other, and BCs tend to be more frequent when the speakers are unfamiliar to each other (Bodur et al., 2022). The rate of fluencemes, on the other hand, is higher compared to prior work. As backchannel rate influences fluency, and therefore the frequency of fluencemes, our results on backchannel and fluenceme rate can be related. In fact, low backchannel rate has been shown to decrease fluency and could therefore relate to higher fluenceme rate (Bavelas et al., 2000). The high frequency of fluencemes could also be an artifact of our methods, as this study includes a large variety of particles often not investigated together, which makes it difficult to interpolate a fluenceme rate. Prior work on different types of fluencemes in corpus research also uses different measurements (e.g. in relation to word frequency, per 100 words in Bortfeld et al., 2001, or per 1000 words in Crible, 2017). Therefore, more comparable analyses are necessary to relate these findings to larger corpus research.

Within dyads, the conversational partners are more similar to each other in BC rate than they are in FL rate. This result can be explained by the fact that BCs are a conversational device that is not only used with a turn-taking function, but also with an interactive and social function of entrainment and rapport building between the speakers, and to show active affiliation and understanding (Bavelas et al., 2000; Cutrone, 2005; Dideriksen et al., 2023). Fluencemes and fluency may be modulated by contextual factors such, as the genre or speech register (Kosmala and Crible, 2022; Böttcher and Zellers, 2024), but are also related to idiosyncrasy (Kosmala and Crible, 2022; Özsoy and Blum, 2023). Additionally, their use is functionally more diverse and connected with the turn-taking function of starting, maintaining, closing the turn and structuring the topics and the turn (Swerts, 1998; Staley and Jucker, 2021; Rendle-Short, 2004; Fraser, 2009; Maschler and Schiffrin, 2015), rather than building affiliation with the other interlocutor, as is the case for backchannels.

The result on distribution in overlap or during a gap within the current speaker's turn is largely in line with prior work, with backchannels predominantly used in overlap or turn internally, and fluencemes during gaps or turn externally (Knudsen et al., 2020). This supports the interpretation that backchannels function as a listener's tool for signaling active listenership, while fluencemes serve as a speaker's device for managing speech planning and discourse organisation, and demonstrating active speakership. The analysed data also reveal some tendencies that are possibly gender-related in how these conversational devices are used. In our dataset, for instance, female speakers tend to produce more fluencemes within their own turns, often overlapping with the interlocutor, and they also use more overlapping backchannels. In contrast, male speakers produce more frequently both fluencemes and backchannels outside of turns, typically during pauses or gaps. If further analyses confirmed similar patterns in a bigger dataset, these results could provide further evidence of gender-specific strategies for managing conversational turns (Rossi, 2022).

The analysis of lexical types showed a preference for the monosyllabic and lexical item "ja" (yes). These semantically more specific forms for backchannels might still be connected to their function of displaying understanding and active listenership i.e., confirming the current listener-speaker situation. For fluencemes, the non-lexical forms "ähm" (uhm) are used almost similarly to short lexical items like "ja". These short items can be used to structure the turn in different ways. The non-lexical forms are more frequently used turn externally where a less specific form can encode turn initiation and signal speech planning (Clark and Fox Tree, 2002; De Jong, 2016; Heritage, 2018), while during the

turn a range of discourse and turn managing functions need to be encoded e.g., structuring the topics, maintaining the turn and closing the turn (Swerts, 1998; Staley and Jucker, 2021; Rendle-Short, 2004; Fraser, 2009; Maschler and Schiffrin, 2015). This diversity in function might therefore allow for more diversity in form.

While this variation in the type of item for fluencemes can be observed for all dyads, the use of backchannels shows more variation in type in the female-female dyads compared to male-male dyads. The use of a wide variety of backchannel forms is connected to display active attention: to avoid signaling boredom or inattention, (female) listeners might be diversifying their verbal backchannels throughout a conversation rather than repeatedly using the same forms (McCarthy, 2003; Schegloff, 1982). Within our dataset, female speakers, or interlocutors in mixed-gender dyads, might be more attentive to this unconscious practice.

Our findings on the most frequent lexical forms in this dataset for backchannels confirm previous studies (Sbranna et al., 2022, 2024; Wehrle, 2021). The analysed lexical forms for fluencemes are shorter than the same lexical forms for backchannels. This suggests that the different interactional function of forms with the same phonological structure might be encoded in the variation of their acoustic detail (e.g., Drager, 2011; Martinuzzi and Scherz, 2022), although further research is needed to confirm this. Fluencemes are predominantly produced as part of speaker turns, aligning temporally with their speech rate, and they tend to exhibit reduced prosodic prominence relative to the propositional elements of the utterance, which may result in their being realised with shorter duration. Only non-lexical fluencemes are longer in duration than the other lexical fluenceme forms. This can also be linked to their positional placement, i.e., either at the onset of a speaker's turn or phrased as separate units, where they are more likely to be prolonged. Such lengthening may function as a turn-holding cue and indicate cognitive planning processes, which are sometimes interpreted as hesitations (Clark and Fox Tree, 2002; Shriberg, 2001). Backchannels are produced as stand-alone utterances rather than as components of longer turns, and convey acknowledgment and active listenership. Since listeners producing backchannels are not constrained by speech rate, they may extend the duration of these signals to ensure perceptibility by the speaker. Given that short vocal particles such as fillers are easily missed in perception (Niebuhr and Fischer, 2019), longer backchannel realisations may serve to enhance their effectiveness as feedback cues.

The contour differences among backchannels and between the non-lexical conversational devices in this dataset confirm the results from previous studies (Wehrle and Grice, 2019; Sbranna et al., 2022, 2024; Wehrle, 2021). Additionally, in our data the lexical item "genau" (exactly) shows to be less intonationally specific when considering the linear intonation contour between backchannel and fluenceme use. In both cases "genau" is produced with a falling contour. This can be explained by the existence of a lexicalised pitch contour i.e., an intonation contour associated with the lexical item that is stored and produced as a prosodic exemplar. These kind of lexicalised pitch contours have also been observed in the case of other short phrases and collocations that frequently occur in discourse (Calhoun and Schweitzer, 2012). This implies that, while some conversational devices vary in prosodic form possibly in relation to their function in conversation (backchannel vs. fluenceme), other items such as "ja" might be stored and (re-)produced as prosodic units with lexicalised intonation contours. In the case of "ja", the accompanying head gesture then contributes to a functional disambiguation. It is possible that the conversational activity also had an influence on the results. Dialogues in the DUEL corpus are in fact loosely task-directed (Hough et al., 2016), involving an interactional activity requiring a conversational goal to be reached by the two participants (see Section 3), and we do not exclude that this might have played a role in the configuration of backchannels and fluencemes in this dataset. Previous research indicates that, in task-based interactions, conversational devices exhibit greater intonational variation with identifiable form-specific contours, and backchannels are more likely to arise with rising intonation, as it was the case in our dataset. In contrast, spontaneous interactions tend to show more limited intonational variation, with feedback being more often lexical and associated with falling or level contours (Spaniol et al., 2024; Wehrle, 2021; Sbranna et al., 2022). While the assessment of task influence on backchannel and fluenceme configuration was not a direct objective of the current study, a future comparison of this dataset with a corpus of spontaneous interactions might provide more evidence on the impact of conversational activity.

Furthermore, the two conversational devices show different co-occurrence patterns with head movements: head nods co-occur more frequently with backchannels, while FLs are accompanied by a wider variety of head movements, even though less frequently than backchannels. This is in line with prior work on fluencemes with a hesitation function accompanied by few or no gestures (Betz et al., 2023; Baiat et al., 2013), and with the diverse findings of research on different types of fluencemes (Ferré, 2011). This lower gestural specificity is accompanied by the lexical form diversity also highlighted by our results. We therefore argue for a general lower specificity of fluencemes when compared to backchannels. Further research will need to address whether this can also be observed in manual gestures, or whether patterns of turn-initial versus turn-medial fluencemes can be identified.

## 5.1 Conclusion and outlook

The present research explored the distribution, lexical type and form, prosodic shape and multimodal variation of backchannels and fluencemes, and provides a comparative analysis of these two conversational items. The analysed data largely confirms prior work by highlighting a tendency for backchannels to be produced by the listener in overlap with the current speaker turn, with rising intonation contours and accompanied by nodding head gestures, while fluencemes are part of speaker turns and less often produced in overlap, predominantly with level or falling intonation and fewer but more varied co-occurring head gestures. The results on dyad-specific and individual conversational behavior provide further evidence for possible gender-specific tendencies with more overlapping speech and more diverse inventory of both verbal and multimodal items employed by female speakers, while male speakers tend to produce more gaps and stick to a more specific inventory of verbal and non-verbal backchannels. It is important to highlight, however, that while we interpreted some of the emerged dyad-specific tendencies in light of gender

performance, the use of both fluencemes and backchannels is highly influenced by several more individual factors that were not taken into account for the current study. For instance, backchannels have been found to be influenced by the age of the speakers, by the degree of familiarity among interlocutors, their level of attention, as well as their personality measured using the the Big Five traits (Blomsma et al., 2024; Engwall et al., 2023; Buschmeier et al., 2014; Vinciarelli et al., 2015); similarly, fluenceme rate has been reported to be influenced by familiarity, by the role of the interlocutor in the conversation (Vinciarelli et al., 2015), and also by individual strategies of dealing with social and psycholinguistic demands and individual levels of tolerance for hesitations in speech (Schettino et al., 2021; Cataldo et al., 2019). For this reason, it would be interesting, in a further development of this study, to highlight how different identity variables of the individual interlocutors are performed through their use of conversational devices in both the role of the speaker and the role of the listener in the interaction.

In our study we focused on head movements as one of the most common conversational feedback gestures. Yet, further analyses need to take other bodily articulators into consideration (including manual and other non-manual gestures) to account for a holistic representation of communicative feedback in dyadic conversations (e.g. Bauer et al., 2025).

This study is limited by a small sample of speakers, which makes it difficult to generalise the findings to a larger population, particularly with regard to the gender-specific tendencies found across dyads. However, the observed trends are consistent with prior research on backchannels and fluencemes, suggesting their potential relevance despite the sample size. Future research will need to investigate whether these tendencies hold also with a larger set of conversations. The data is also restricted with regard to the number of speakers within a conversation. Backchannel and fluenceme use in triadic conversations might in fact show different strategies and patterns.

While this study is exploratory in nature we show how beneficial a comparative look at backchannels and fluencemes is for the understanding of the complex mechanisms of conversation management, and that the speaker's "okay" is different than the listener's "okay" in terms of distribution, phonetic form and its interaction with head movements.

## Data availability statement

The data analysed in this study is subject to the following licenses/restrictions: The dataset is only partially available online (only transcriptions and annotations, not audio files), and can be found at https://github.com/clp-research/DUEL. Requests to access these datasets should be directed to https://github.com/clp-research/DUEL.

## Author contributions

MB: Software, Writing – review & editing, Writing – original draft, Formal analysis, Methodology, Data curation, Visualization, Conceptualization, Investigation. MR: Writing – original draft, Methodology, Investigation, Software, Validation, Conceptualization, Writing – review & editing, Data curation.

## Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

## Acknowledgments

The authors would like to thank Margaret Zellers for her helpful feedback and support throughout the preparation of this work.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

## Publisher's note



## References

Baiat, G. E., Coler, M., Pullen, M., Tienkouw, S., and Hunyadi, L. (2013). "Multimodal analysis of "well" as a discourse marker in conversation: A pilot study," in *2013 IEEE 4th International Conference on Cognitive Infocommunications (CogInfoCom)* (Budapest: IEEE), 283–288.

Bauer, A., Gipper, S., Herrmann, T.-A., and Hosemann, J. (2025). *Rethinking linguistic Feedback: A Modality-Agnostic and Holistic Approach to Multimodal Addressee Signals in Spoken and Signed Dyadic Interaction* [Preprint]. doi: 10.31219/osf.io/t8czb_v1

Bavelas, J. B., Coates, L., and Johnson, T. (2000). Listeners as co-narrators. *J. Personal. Soc. Psychol.* 79:941. doi: 10.1037//0022-3514.79.6.941

Belz, M., and Reichel, U. D. (2015). "Pitch characteristics of filled pauses," in *Proceedings of Disfluency in Spontaneous Speech (DiSS). The 7th Workshop on Disfluency in Spontaneous Speech*, 1–4. Available online at: https://real.mtak.hu/32442/1/BelzReichelDiss2015.pdf (Accessed July 10, 2025).

Beňuš, ; S., Gravano, A., and Hirschberg, J. (2007). "The prosody of backchannels in American English," in *Proceedings of the International Congress of Phonetic Sciences (ICPhS 2007)* (London: International Phonetic Association), 1065–1068.

Betz, S., Bryhadyr, N., Türk, O., and Wagner, P. (2023). Cognitive load increases spoken and gestural hesitation frequency. *Languages* 8:71. doi: 10.3390/languages8010071

Blomsma, P., Vaitonytė, J., Skantze, G., and Swerts, M. (2024). Backchannel behavior is idiosyncratic. *Lang. Cognit.* 16, 1158–1181. doi: 10.1017/langcog.2024.1

Bodur, K., Nikolaus, M., Fourtassi, A., and Prévot, L. (2022). "Backchannel behavior in child-caregiver zoom-mediated conversations," in *Proceedings of the 44th Annual Meeting of the Cognitive Science Society* (San Francisco: Cognitive Science Society).

Boersma, P., and Weenink, D. (2024). "Praat: doing phonetics by computer," in *Computer Software*. Version 6.4.21. Available online at: https://www.praat.org (Accessed September 24, 2024).

Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., and Brennan, S. E. (2001). Disfluency rates in conversation: effects of age, relationship, topic, role, and gender. *Lang. Speech* 44, 123–147. doi: 10.1177/00238309010440020101

Böttcher, M., and Zellers, M. (2024). Do you say uh or uhm? A cross-linguistic approach to filler particle use in heritage and majority speakers across three languages. *Front. Psychol.* 15:1305862. doi: 10.3389/fpsyg.2024.1305862

Boudin, A., Rauzy, S., Bertrand, R., Ochs, M., and Blache, P. (2024). How is your feedback perceived? An experimental study of anticipated and delayed conversational feedback. *JASA Express Letters* 4:075201. doi: 10.1121/10.0026448

Brinton, L. J. (2017). *The Evolution of Pragmatic Markers in English: Pathways of Change*. Cambridge: Cambridge University Press.

Buschmeier, H., Malisz, Z., Skubisz, J., Wlodarczak, M., Wachsmuth, I., Kopp, S., et al. (2014). "ALICO: a multimodal corpus for the study of active listening," in *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)* (Reykjavik: European Language Resources Association (ELRA)), 3638–3643.

Calhoun, S., and Schweitzer, A. (2012). Can intonation contours be lexicalised? Implications for discourse meanings. *Prosody Mean.* 2012, 271–328. doi: 10.1515/9783110261790.271

Cataldo, V., Schettino, L., Savy, R., Poggi, I., Origlia, A., Ansani, A., et al. (2019). "Phonetic and functional features of pauses, and concurrent gestures, in tourist guides' speech," in *Gli archivi sonori al crocevia tra scienze fonetiche, informatica umanistica e patrimonio digitale, volume 6 of Studi AISV* (Milan: Officinaventuno), 205–231.

Cerrato, L. (2012). *Investigating Communicative Feedback Phenomena across Languages and Modalities* (PhD thesis). Royal Institute of Technology, Stockholm, Sweden.

Cerrato, L., and Skhiri, M. (2003). "Analysis and measurement of head movements signalling feedback in face-to-face human dialogues," in *Proceedings of the First Nordic Symposium on Multimodal Communication* (Copenhagen), 43–52.

Clark, H. H., and Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition* 84, 73–111. doi: 10.1016/S0010-0277(02)00017-3

Crible, L. (2017). *Discourse markers and (dis) fluency across registers - A Contrastive Usage-Based Study in English and French* (dissertation). Louvain: University of Louvain.

Crible, L., Degand, L., and Gilquin, G. (2017). The clustering of discourse markers and filled pauses: A corpus-based French-English study of (dis) fluency. *Lang. Contrast* 17, 69–95. doi: 10.1075/lic.17.1.04cri

Cutrone, P. (2005). A case study examining backchannels in conversations between Japanese-British dyads. *Multilingua* 24, 237–274. doi: 10.1515/mult.2005.24.3.237

Cutrone, P. (2011). "Politeness and face theory: implications for the backchannel style of Japanese L1/L2 speakers," in *Language Studies Working Papers. University of Reading: Reading*, eds. D. Giannoni, and C. Ciarlo (Reading: University of Reading).

De Jong, N. H. (2016). Predicting pauses in L1 and L2 speech: The effects of utterance boundaries and word frequency. *Int. Rev. Appl. Linguist. Lang. Teach.* 54, 113–132. doi: 10.1515/iral-2016-9993

Dideriksen, C., Christiansen, M. H., Tylén, K., Dingemanse, M., and Fusaroli, R. (2023). Quantifying the interplay of conversational devices in building mutual understanding. *J. Exp. Psychol.: General* 152, 864–889. doi: 10.1037/xge0001301

Diewald, G. (2013). "Same same but different. Modal particles, discourse markers and the art (and purpose) of categorization," in *Discourse markers and modal particles: Categorization and Description* (Amsterdam, Netherlands: John Benjamins Publishing Company), 19–46.

Dittmann, A. T., and Llewellyn, L. G. (1968). Relationship between vocalizations and head nods as listener responses. *J. Personal. Soc. Psychol.* 9, 79–84. doi: 10.1037/h0025722

Drager, K. (2011). Sociophonetic variation and the lemma. *J. Phonet.* 39, 694–707. doi: 10.1016/j.wocn.2011.08.005

Engwall, O., Cumbal, R., and Majlesi, A. R. (2023). Socio-cultural perception of robot backchannels. *Front, Robot, AI* 10:988042. doi: 10.3389/frobt.2023.988042

Esposito, A., McCullough, K. E., and Quek, F. (2001). "Disfluencies in gesture: Gestural correlates to filled and unfilled speech pauses," in *Proceedings of IEEE Workshop on Cues in Communication* (Princeton, NJ: Citeseer), 1–6.

Fellegy, A. M. (1995). Patterns and functions of minimal response. *Am. Speech* 70, 186–199. doi: 10.2307/455815

Ferré, G. (2011). "Multimodal analysis of discourse markers "donc", "alors" and "en fait" in conversational French," in *Proceedings of the International Congress of Phonetic Sciences (ICPhS)* (London: International Phonetic Association), 671–674.

Ferré, G., and Renaudier, S. (2017). Unimodal and bimodal backchannels in conversational English. *Proc. SEMDIAL* 2017, 27–37. doi: 10.21437/SemDial.2017-3

Fischer, K. (2000). Discourse particles, turn-taking, and the semantics-pragmatics interface. *Revue de sémantique et pragmatique* 8:111–132.

Fraser, B. (2009). An account of discourse markers. *Int. Rev. Pragmat.* 1, 293–320. doi: 10.1163/187730909X12538045489818

Götz, S. (2013). *Fluency in Native and Nonnative English Speech*. Amsterdam: John Benjamins Publishing Company.

Habalet, B. (2019). The impact of the listener's gender on the frequency and diversity of the English audible backchannels. *Human Sci. J.* 2019, 59–68. doi: 10.34174/0079-000-052-020

Heinz, B. (2003). Backchannel responses as strategic responses in bilingual speakers' conversations. *J. Pragmat.* 35, 1113–1142. doi: 10.1016/S0378-2166(02)00190-X

Helweg-Larsen, M., Cunningham, S. J., Carrico, A., and Pergram, A. M. (2004). To nod or not to nod: An observational study of nonverbal communication and status in female and male college students. *Psychol. Women Quart.* 28, 358–361. doi: 10.1111/j.1471-6402.2004.00152.x

Heritage, J. (2018). "Turn-initial particles in English: the cases of oh and well," in *Between Turn and Sequence* (Amsterdam: John Benjamins Publishing Company), 155–190.

Hough, J., Tian, Y., de Ruiter, L., Betz, S., Kousidis, S., Schlangen, D., et al. (2016). "DUEL: a multi-lingual multimodal dialogue corpus for disfluency, exclamations and laughter," in *Proceedings of LREC* (Paris: ELRA European Language Resources Association), 1784–1788.

Jefferson, G. (1984). Notes on a systematic deployment of the acknowledgement tokens "yeah" and "mm hm". *Paper Linguist.* 17, 197–216. doi: 10.1080/08351818409389201

Kjellmer, G. (2009). Where do we backchannel? On the use of mm, mhm, uh huh and such like. *Int. J. Corpus Linguist.* 14, 81–112. doi: 10.1075/ijcl.14.1.05kje

Knudsen, B., Creemers, A., and Meyer, A. S. (2020). Forgotten little words: how backchannels and particles may facilitate speech planning in conversation? *Front. Psychol.* 11:593671. doi: 10.3389/fpsyg.2020.593671

Kosmala, L. (2022). Exploring the status of filled pauses as pragmatic markers: the role of gaze and gesture. *Pragmat. Cognit.* 29, 272–296. doi: 10.1075/pc.21020.kos

Kosmala, L., and Crible, L. (2022). The dual status of filled pauses: Evidence from genre, proficiency and co-occurrence. *Lang. Speech* 65, 216–239. doi: 10.1177/00238309211010862

Kousidis, S., Pfeiffer, T., and Schlangen, D. (2013). Mint. tools: Tools and adaptors supporting acquisition, annotation and analysis of multimodal corpora. *Proc. Interspeech*, 2013, 2649–2653. doi: 10.21437/Interspeech.2013-609

Krause-Ono, M. (2004). *Change in Backchanneling Behaviour: The Influence from L2 to L1 on the Use of Backchannel Cues* (PhD thesis). Muroran Institute of Technology, Muroran, Japan.

Krepsz, V., Horváth, V., Hámori, Á., Gyarmathy, D., and Dér, C. I. (2022). Backchannel responses in Hungarian conversations: a corpus-based study on the effect of the partner's age and gender. *Linguistica Silesiana* 2022, 113–140. doi: 10.24425/linsi.2022.141220

Lee, L., Jouvet, D., Bartkova, K., Keromnes, Y., and Dargnat, M. (2020). "Correlation between prosody and pragmatics: case study of discourse markers in French and English," in *Proceedings of Interspeech*. doi: 10.21437/Interspeech.2020-2204

Levinson, S. C. (2016). Turn-taking in human communication-origins and implications for language processing. *Trends Cognit. Sci.* 20, 6–14. doi: 10.1016/j.tics.2015.10.010

Liesenfeld, A., and Dingemanse, M. (2022). "Bottom-up discovery of structure and variation in response tokens (backchannels) across diverse languages," *Proceedings of Interspeech*, 1126–1130. doi: 10.21437/Interspeech.2022-11288

Malisz, Z., Wlodarczak, M., Buschmeier, H., Kopp, S., and Wagner, P. (2012). "Prosodic characteristics of feedback expressions in distracted and non-distracted listeners," in *Proceedings of The Listening Talker: An Interdisciplinary Workshop on Natural and Synthetic Modification of Speech in Response to Listening Conditions,*

*Edinburgh, UK*, 36–39. Retrieved from https://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-185478

Martinuzzi, C., and Scherz, J. (2022). Sorry, not sorry: the independent role of multiple phonetic cues in signaling the difference between two word meanings. *Lang. Speech* 65, 143–172. doi: 10.1177/0023830921988975

Maschler, Y., and Schiffrin, D. (2015). "Discourse markers language, meaning, and context," in *The Handbook of Discourse Analysis* (Hoboken, NJ: Wiley), 189–221.

McCarthy, M. (2003). Talking back: small interactional response tokens in everyday conversation. *Res. Lang. Soc. Interact*. 36, 33–63. doi: 10.1207/S15327973RLSI3601_3

McClave, E. (2000). Linguistic functions of head movements in the context of speech. *J. Pragmat*. 32, 855–878. doi: 10.1016/S0378-2166(99)00079-X

Niebuhr, O., and Fischer, K. (2019). "Do not hesitate!-unless you do it shortly or nasally: How the phonetics of filled pauses determine their subjective frequency and perceived speaker performance," in *Proceedings of Interspeech* (Rotterdam: International Speech Communication Association), 544–548.

Oloff, F. (2019). Okay as a neutral acceptance token in german conversation. *Lexique* 25, 197–225. doi: 10.54563/lexique.924

Özsoy, O., and Blum, F. (2023). Exploring individual variation in Turkish heritage speakers' complex linguistic productions: Evidence from discourse markers. *Appl. Psycholinguist*. 44, 534–564. doi: 10.1017/S0142716423000267

Paggio, P., and Navarretta, C. (2011). "Head movements, facial expressions and feedback in Danish first encounters interactions: a culture-specific analysis," in *Universal Access in Human-Computer Interaction. Users Diversity - 6th International Conference, volume 6766 of Lecture Notes in Computer Science*, eds. C. Stephanidis (Berlin: Springer), 583–590.

Plug, I., Stommel, W., Lucassen, P. L., olde Hartman, T. C., van Dulmen, S., and Das, E. (2021). Do women and men use language differently in spoken face-to-face interaction? A scoping review. *Rev. Commun. Res*. 9, 43–79. doi: 10.12840/issn.2255-4165.026

Poppe, R., Truong, K., and Heylen, D. (2011). "Backchannels: quantity, type and timing matters," in *Proceedings of the Intelligent Virtual Agents International Conference* (Cham: Springer), 228–239.

Posit team (2023). *RStudio: Integrated Development Environment for R*. Boston, MA: Posit Software, PBC.

R Core Team (2023). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rendle-Short, J. (2004). Showing structure: Using UM in the academic seminar. *Pragmat. Quart. Publicat. Int. Pragmat. Assoc*. 14, 479–498. doi: 10.1075/prag.14.4.04ren

Roberts, S. G., Torreira, F., and Levinson, S. C. (2015). The effects of processing and sequence organization on the timing of turn taking: a corpus study. *Front. Psychol*. 6:509. doi: 10.3389/fpsyg.2015.00509

Rohrer, P., Tütüncübasi, U., Vilà-Giménez, I., Florit-Pons, J., Esteve-Gibert, N., Ren-Mitchell, A., et al. (2023). *The MultiModal MultiDimensional (M3D) Labeling System*. doi: 10.17605/OSF.IO/ANKDX

Rohrer, P., Vila-Giménez, I., Florit-Pons, J., Gurrado, G., Esteve Gibert, N., Ren, A., et al. (2020). "The multimodal multidimensional (M3D) labelling scheme for the annotation of audiovisual corpora," in *Proceedings of GESPIN*. Available online at: https://www.researchgate.net/publication/344243029_The_MultiModal_MultiDimensional_M3D_labelling_scheme_for_the_annotation_of_audiovisual_corpora (Accessed July 10, 2025).

Rossi, M. (2022). "Gender influence on phonetic turn-taking cues at potential transition locations in German," in *The Position of the Speaker in Interaction: Attitudes, Intentions, and Emotions in Verbal Communication*, eds. R. Orrico, and L. Schettino (Milan: Officinaventuno), 127–147.

Rossi, M., Schröer, M., Ludusan, B., and Zellers, M. (2023). "A multimodal account of listener feedback in face-to-face interactions," in *Proceedings of the International Congress of Phonetic Sciences (ICPhS)* (London: International Phonetic Association), 4121–4125.

Sacks, H., Schegloff, E., and Jefferson, G. (1974). A simplest systematics for the organisation of turn-taking for conversation. *Language* 50, 696–735. doi: 10.1353/lan.1974.0010

Sbranna, S., Möking, E., Wehrle, S., and Grice, M. (2022). "Backchannelling across Languages: Rate, Lexical Choice and Intonation in L1 Italian, L1 German and L2 German," in *Proceedings of Speech Prosody 2022*, 734–738. doi: 10.21437/SpeechProsody.2022-149

Sbranna, S., Wehrle, S., and Grice, M. (2024). A multi-dimensional analysis of backchannels in l1 german, l1 italian and l2 german. *Lang. Interact. Acquisit*. 15, 243–277. doi: 10.1075/lia.00026.sbr

Schegloff, E. (1982). Discourse as an interactional achievement: some uses of 'uh huh' and other things that come between sentences. *Analyz. Discour.: Text and talk* 71, 71–93.

Schettino, L., Betz, S., Cutugno, F., and Wagner, P. (2021). "Hesitations and individual variability in Italian tourist guides' speech," in *Speaker Individuality in Phonetics and Speech Sciences: Speech Technology and Forensic Applications* (Milan: Officinaventuno), 243–262.

Selting, M., Auer, P., Barth-Weingarten, D., Bergmann, J. R., Bergmann, P., Birkner, K., et al. (2009). *Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). Gesprächsforschung: Online-Zeitschrift zur verbalen Interaktion*.

Shriberg, E. (2001). To 'errrr' is human: ecology and acoustics of speech disfluencies. *J. Int. Phonetic Assoc*. 31, 153–169. doi: 10.1017/S0025100301001128

Spaniol, M., Wehrle, S., Janz, A., Vogeley, K., and Grice, M. (2024). "The influence of conversational context on lexical and prosodic aspects of backchannels and gaze behaviour," in *Proceedings of Speech Prosody* 2024, 607–611.

Staley, L., and Jucker, A. H. (2021). "The uh deconstructed pumpkin pie" : The use of uh and um in Los Angeles restaurant server talk. *J. Pragmat*. 172, 21–34. doi: 10.1016/j.pragma.2020.11.004

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., et al. (2009). Universals and cultural variation in turn-taking in conversation. *Proc. Nat. Acad. Sci*. 106, 10587–10592. doi: 10.1073/pnas.0903616106

Stubbe, M. (1998). Are you listening? Cultural influences on the use of supportive verbal feedback in conversation. *J. Pragmat*. 29, 257–289. doi: 10.1016/S0378-2166(97)00042-8

Swerts, M. (1998). Filled pauses as markers of discourse structure. *J. Pragmat*. 30, 485–496. doi: 10.1016/S0378-2166(98)00014-9

Truong, K., Poppe, R., de Kok, I., and Heylen, D. (2011). "A multimodal analysis of vocal and visual backchannels in spontaneous dialogues," in *Proceedings of Interspeech*, 23–25.

Ueno, J. (2004). Gender differences in Japanese conversation. *Intercult. Commun. Stud*. 13, 85–100.

Vinciarelli, A., Chatziioannou, P., and Esposito, A. (2015). When the words are not everything: the use of laughter, fillers, back-channel, silence, and overlapping speech in phone calls. *Frontiers in ICT* 2:4. doi: 10.3389/fict.2015.00004

Wagner, P., Malisz, Z., and Kopp, S. (2014). Gesture and speech in interaction: an overview. *Speech Commun*. 57:209–232. doi: 10.1016/j.specom.2013.09.008

Ward, N., and Tsukahara, W. (2000). Prosodic features which cue backchannel responses in English and Japanese. *J. Pragmat*. 32, 1177–1207. doi: 10.1016/S0378-2166(99)00109-5

Wehrle, S. (2021). *A Multi-Dimensional Analysis of Conversation and Intonation in Autism Spectrum Disorder* (Phd thesis). University of Cologne, Cologne, Germany.

Wehrle, S., and Grice, M. (2019). "Function and Prosodic Form of Backchannels in L1 and L2 German," in *ISPhonCog 2019: Hanyang International Symposium on Phonetics & Cognitive Sciences of Language 2019, Seoul, Korea*.

Wehrle, S., Grice, M., and Vogeley, K. (2024). Filled pauses produced by autistic adults differ in prosodic realisation, but not rate or lexical type. *J. Autism Dev. Disord*. 54, 2513–2525. doi: 10.1007/s10803-023-06000-y

Wehrle, S., Roettger, T., and Grice, M. (2018). "Exploring the dynamics of backchannel interpretation: The meandering mouse paradigm," in *ProsLang: Workshop on the Processing of Prosody across Languages and Varieties, Wellington, New Zealand*. doi: 10.13140/RG.2.2.14707.30248

Wehrle, S., Vogeley, K., and Grice, M. (2023). Backchannels in conversations between autistic adults are less frequent and less diverse prosodically and lexically. *Lang. Cognit*. 2023, 1–26. doi: 10.31234/osf.io/tbr3c

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Cham: Springer-Verlag New York.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., and Sloetjes, H. (2006). "ELAN: a professional framework for multimodality research," in *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC 2006)* (Genoa: European Language Resources Association (ELRA)), 1556–1559.

Włodarczak, M., Buschmeier, H., Malisz, Z., Kopp, S., and Wagner, P. (2012). "Listener head gestures and verbal feedback expressions in a distraction task," in *Proceedings of the Interdisciplinary Workshop on Feedback Behaviours in Dialogue*, 93–96. doi: 10.5281/zenodo.3234893

Yngve, V. H. (1970). *On Getting a Word in Edgewise. CLS-70*, 567–578.

Zellers, M. (2021). An overview of forms, functions, and configurations of backchannels in Rurụuli/Lunyala. *J. Pragmat*. 175, 38–52. doi: 10.1016/j.pragma.2021.01.012