DOI: 10.61782/fa.2023.0432

forum acusticum 2023

BACKCHANNEL BEHAVIOR IN CONVERSATIONS AND HOW IT IS AFFECTED BY HEARING LOSS, NOISE, AND HEARING AIDS

Eline Borch Petersen[1*] **Els Walravens**[1] **Anja Kofoed Pedersen**[1]
[1] WS Audiology, Lynge, Denmark

ABSTRACT

The dynamics and content of conversations between people could help us understand how hearing impairment (HI) and hearing aids affects every-day communication. We investigated how backchannel behavior is affected when people experience communication difficulties. Backchannels are utterances, such as 'mmm', made to indicate engagement and listening. Three estimates of backchannels (utterances shorter than 500 ms, overlaps occurring within another speaker's turn, and utterances with a single syllable) were automatically extracted from free triad conversations and task-oriented dyad conversations. We observed that while the occurrence of turns shorter than 500 ms better represent backchannels in dyad conversations, the occurrence of overlap within another speaker's turn more accurately represents backchannels in triad conversations. We found that normal-hearing (NH) interlocutors had more short turns relative to the HI interlocutors, and that the NH interlocutor reduced the occurrence of short turns in noise and when the HI interlocutor was not wearing a hearing aid. In noise, the HI interlocutors were observed to reduce the number of overlaps within other speakers turns. Taken together this suggests that increased communication difficulty affects backchannel behavior during triad and dyad conversations.

Keywords: *Conversation, Hearing aids, Hearing loss, Real-Life Test Method*

1. INTRODUCTION

In the quest to improve the understanding of how hearing aids affect the everyday lives of hearing-impaired (HI) users, a relevant situation to look into is communication [1]. Communication is an important and integral part of our lives, but we seldom consider the complexity involved in simultaneously listening to others while planning and delivering our own upcoming response [2]. In recent years, several studies have focused on how the dynamics of conversations are affected when one conversational partner (interlocutor) has impaired hearing. In conversations between a HI and a normal hearing (NH) interlocutor, the HI interlocutor exhibits more variability in the timing of their turn starts and speaks for longer at a time [3-5]. These changes suggest that HI interlocutors are experiencing difficulties which affect their communication dynamics. Importantly, the NH interlocutor is also affected when conversing with a HI partner and adapts the spectral content of their speech and increases the volume at which they speak [3-4, 6-7] in order to help the HI interlocutor overcome the communication difficulty.

Some of the communication difficulties can be alleviated by providing the HI interlocutor with hearing-aid (HA) amplification, causing the HI interlocutor to speak faster (increased articulation rate), initiate their turns faster, shorten their turns, and finally reduce the speech levels of both the NH and HI interlocutor and the spectral modification made by the NH interlocutor [4,8].

The above studies focus on traditional measures of conversational dynamics, e.g., speech duration, level, speed, and turn-taking timing. Another equally interesting feature of a conversation is the social aspect of providing backchannels. Backchannels are utterances that are not made with the intention of taking a turn, but rather indicate engagement and listening [9]. Backchannels usually manifest as short utterances like 'yeah' or 'mmm', but can also be short sentences, such as 'I see' or 'that's right', or even visual events such as nodding or facial expressions [10]. Common for all backchannels are that they carry little or no propositional context, and as such, they can be

[*]***Corresponding author**: Eline.Petersen_ad_wsa.com.*

delivered in a less timely manner than regular information-bearing turns.

In this paper, the backchannel behavior of NH and HI interlocutors engaging in two- and three-person conversations, denoted dyad and triad conversations, is investigated. When experiencing increased communication difficulties, we hypothesize that interlocutors will reduce the occurrence of verbal backchannels because 1) delivering backchannels requires mental energy which is better spent on following the conversations and 2) to avoid disrupting the auditory floor, thereby making listening more difficult.

Based on the voice activity detected from each interlocutor, three measures indicative of backchannel behavior are automatically extracted: As backchannels are often short utterances, one estimate was extracted as 1) the occurrence of utterances shorter than 500 ms. As backchannels are not meant to take the floor, they are socially acceptable to deliver in overlap with other speakers turn and 2) the occurrence of overlaps made within another interlocutor's speech is also extracted as a measure of backchannel behavior. Finally, short utterances can be quantified by computing the number of articulations and extracting 3) the occurrence of utterances containing a single syllable. Based on these three measures of backchannel behavior, we investigated how increased communication difficulty in the form of being hearing impaired, adding background noise to the conversation, and providing amplification to the HI interlocutor affected the occurrence of backchannels in dyad and triad conversations.

2. METHODS

The current paper presents results from two different experiments: One where a NH and a HI interlocutors collaborate on a spot the difference task [4] and one study in which two NH and one HI interlocutor interact in a guided, but free conversation [11]. Below is a brief description of each experiment and the steps taken to extract the three backchannelling features.

2.1 Conversational Data

To compare data from the two experiments, only a subset of the originally tested experimental conditions is used: The conversations held in quiet/near-quiet where the HI interlocutor received HA amplification or not (denoted Quiet Aided and Quiet Unaided, respectively) and the conversations in the presence of background noise where the HI interlocutor was wearing HAs with an omni-directional sound processing scheme (denoted Noise Aided). In both experiments the HI interlocutor was fitted with binaural Signia Pure 312 7X receiver-in-the-canal HAs.

2.1.1 Two-Talker Conversations

Conversations consisted of 11 pairs, of a younger NH (mean age 25.3 years) and an elderly HI interlocutor (mean age 74.1) with symmetrical mild-to-moderate hearing-impairment, solving the Danish version of the Diapix tasks [12], [13]. Conversations were recorded during five experimental conditions repeated twice. The interlocutors were given a maximum of 10 minutes to identify 10 differences between the image each of them was provided with. In the conditions included into the current study, the pairs conversed in complete quiet, or in the presence of babble noise presented at 70 dB SPL from two loudspeakers positioned between the interlocutors at +/- 45 degrees azimuth.

2.1.2 Three-Talker Conversations

Conversations from 25 triads consisting of one elderly HI (mean age 75.8) and two NH interlocutors were recorded. The two NH interlocutors were recruited to be either younger (mean age 27.2) or middle-aged (mean age 54.8). The analysis provided here will not distinguish between younger and middle-aged NH interlocutors.

Natural and free conversations were initiated using images with keywords, and questions requiring a consensus decision [11]. The interlocutors sat equally spaced in a circle while performing twelve 5-minutes conversations under four different experimental conditions repeated three times. During all conversations, noise from a canteen scene was presented from three loudspeakers placed in front of each interlocutor. The noise was presented at 50 dBC SPL (denoted quiet) and at 75 dBC SPL.

In both the dyad and triad conversations, the speech of each talker was recorded using individual directional headsets (DPA 4088 microphones). Besides the number of interlocutors, the dyad and triad conversations differed in other important ways: 1) The quiet condition of the triad conversations was conducted in a low level of background noise. 2) The triad conversations were free, while the dyad conversations were held under time-pressure to solve the task as fast as possible. 3) While the dyad conversations were held in a sound-isolated lab, the triad conversations were conducted in a meeting room.

2.2 Processing of the Speech Data

2.2.1 Voice Activity Detection

To quantify the backchannel behavior, the voice activity of each individual interlocutor was determined. For the dyad conversations this was done automatically using individual power thresholds. However, due to the higher level of background noise used for the triads, the position of the noise sources, and the more abundant presence of crosstalk, the same automatic procedure could not be reliably applied to detect the voice activity. Instead, the speech produced by each interlocutor in the triads was marked manually.

2.2.2 Post-Processing and Feature Extraction

The voice activity detection indicated inter-pausal units (IPU), i.e. stretches of speech surrounded by pauses longer than 180 ms [2]. A turn can consist of a single IPU, but often contains multiple IPUs including the pauses between them. To make the voice activity detection reflect turns rather than IPUs, pauses between IPUs from the same interlocutor that were shorter than 1 second were marked as voice activity. In this study, IPUs uttered completely overlapping with another interlocutor's speech were not considered or included into turns.

By comparing the post-processed voice activity detection of all interlocutors in the conversation, each utterance was classified either as a turn, or as an overlap with another interlocutor's speech. From this classification two of the three measures of backchannel behavior were directly extracted as: 1) The number of turns shorter than 500 ms and 2) the number of overlaps within another interlocutor's speech. Additional steps were taken to identify the 3) number of utterances (turns and overlaps) containing a single syllable: Firstly, the PRAAT software was used to detect all syllables in each recording [14]. Based on this, utterances containing a single syllable was determined. All backchannel measures are computed as the occurrence per minute conversation.

2.3 Statistical Analysis

For each of the three measures of the backchannels, two mixed-effects regression models were generated to separately investigate the effects of HA amplification and background noise. Both models included the main effects of group size (dyad and triad), hearing status (HI and NH), and condition. In one model the condition contrasts HA amplification of the HI interlocutor (Quiet Unaided and Quiet Aided), while another model was generated to test the effect of adding background noise when the HI interlocutor was wearing HAs (Quiet Aided vs Noise Aided). All second-order interactions were included into the model, while person nested within conversations group was included as a random effect. Backwards stepwise elimination of non-significant factors was used to reduce the final model. Post-hoc testing was done by computing pairwise difference in least-squares means.

3. RESULTS

Prior to presenting the effects of altering HA amplification, background noise, and group size on the estimated backchannel behavior, an insight into the three measures of backchannelling will be provided.

3.1 Validation of the Measures of Backchannels

The three measures: 1) Number of utterances shorter than 500 ms per minute, 2) number of utterances overlapping within another interlocutor's speech per minute and 3) number of utterances with a single syllable per minute were extracted with the expectation that they quantify different aspects of backchannelling. To evaluate this expectation, a manual classification of a subset of the utterances was made. A total of 75 sound files were sampled from each of the two experiments (dyad and triad) such that the conditions (unaided quiet, aided quiet, aided noise), talkers (NH and HI) and individual groups were as equally represented as possible. From each of the 150 sound files one example of each of the three measures was extracted and manually classified as either a backchannel or a turn. An utterance was marked as a turn if informational context was provided, or an attempt to take the turn was made.

The result in **Table 1** shows that >80% of the utterances shorter than 500 ms and of the overlaps represented backchannels. It is worth noting that while 90.7% of the turns shorter than 500 ms were backchannels during the dyad conversations, it was only 72.2% during triad conversations. This picture is reversed for the overlaps of which 88% were backchannels for the triad conversations, but only 76% for the dyads.

Only 62% of the single-syllable utterances was identified as backchannelling, suggesting that this is not a reliable estimation. Due to this poor result for both dyad and triad conversations, this measure of backchannels was not included in further analysis.

Table 1: Percentage of the 150 utterances extracted for each of the three backchannel measures manually classified as backchannels. The numbers in italics indicate the occurrence of each measure per interlocutor per minute conversation in quiet, averaged across hearing status and HA amplification.

Group size	Turns < 500ms	Overlaps	Turns with one syllable
Dyad	90.7% *1.9/min*	76.0 % *2.9/min*	65.3% *0.85/min*
Triad	72.2 % *0.7/min*	88.0 % *3.6/min*	58.7 % *0.96/min*
Overall	81.3%	82%	62%

The average occurrence per minute conversation of each of the three measures per interlocutor is also provided in **Table 1**. An estimate of the occurrence of a backchannel (short turn or overlap) can be estimated by first correcting the occurrence by the probability of it actually being a backchannel (percentage provided in **Table 1**), adding the occurrence of short turns and overlaps, and finally multiplying with the number of interlocutors in the conversation. When doing this, the dyad conversations have 8.0 estimated backchannels per minute conversation, while the triad conversations have 11.8 backchannels per minute conversation.

3.2 Effects of HA Amplification in Quiet

All variables significantly affecting the two measures of backchannels are presented in **Table 2**. In **Figure 1**, the occurrence of the two backchannel measures is shown for the NH and HI interlocutors in the dyad and triad conversations. Although presented together in **Figure 1**, only two of the conditions (Quiet Unaided vs Quiet Aided) are included into the statistical analysis of the effects of hearing status, group size, and providing the HI interlocutor with HA amplification.

The statistical analysis revealed that in the dyad conversations, utterances shorter than 500 ms occurred 1.2 times more per minute than during the triad conversations ($p < 0.001$, **Figure 1A**). Hearing status also affected the occurrence of short turns since NH interlocutors provided short-turn backchannels 0.4 times more per minute relative to HI interlocutors ($p < 0.001$, **Figure 1A**).

Finally, an interaction between hearing status and HA processing ($p = 0.02$, **Figure 2A**) revealed that while being aided had no effect on the backchannelling behavior of the HI interlocutor, the NH interlocutor uttered 0.14 fewer short

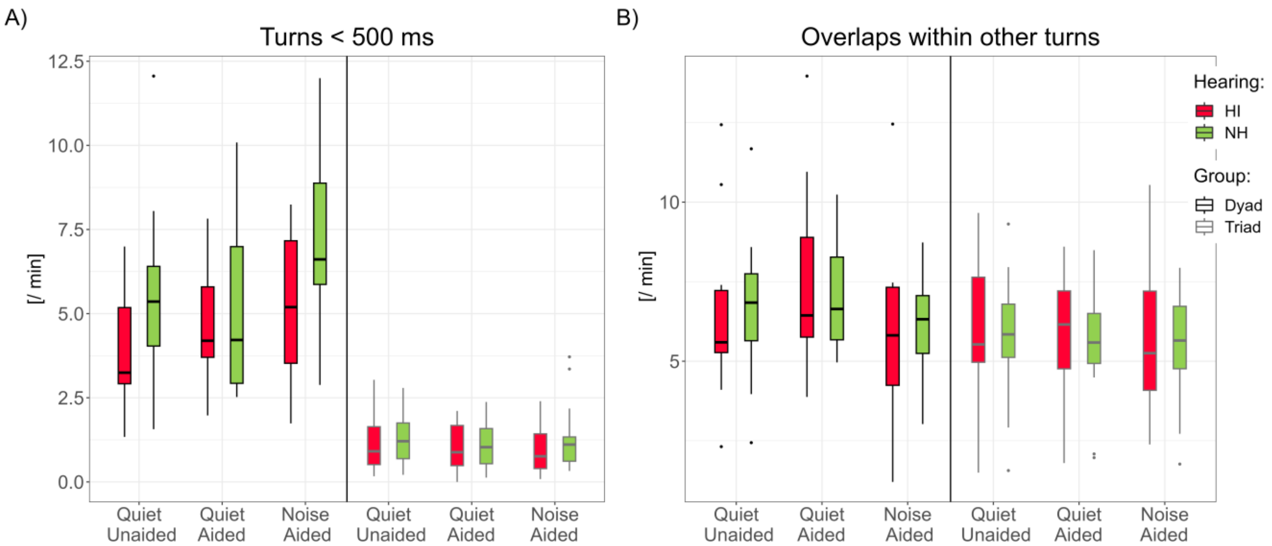

Figure 1: Boxplots of two measures of backchannels A) occurrence of turns shorter than 500 ms and B) occurrence of overlaps made within the turn of another interlocutor. The backchannels are extracted from three experimental conditions in dyad (black boxes) and triad conversation (grey boxes). Boxes belonging to the dyad and triad conversations are separated by a vertical black line and for the normal-hearing (NH, green) and hearing-impaired (HI, red) interlocutors.

Table 2: Statistically significant effects of the two LMER models after backwards stepwise elimination. The full model, prior to the elimination, included all variables and their second order interactions.

	Effects of amplification in quiet			**Effects of noise**		
Turns < 500 ms	Group	F(1,38) = 70	p < 0.001	Group	F(1,38) = 92	p < 0.001
	Hearing	F(1,58) = 12	p < 0.001	Hearing	F(1,62) = 19	p < 0.001
	Hearing * HA	F(1,433) = 6	p = 0.02	Hearing * noise	F(1,433) = 8	p < 0.01
Overlaps				Group	F(1,43) = 8	p < 0.001
	Group	F(1,44) = 5	p = 0.04	Noise	F(1,438) = 16	p < 0.001
	Hearing * HA	F(1,438) = 5	p = 0.02	Group * noise	F(1,438) = 5	p = 0.02
				Hearing * noise	F(1,438) = 6	p = 0.01

turns per minute when the HI interlocutor was aided.

The occurrence of overlaps within other talkers turns also differed with group size, however opposing the above result, the triad conversations had 0.9 overlaps more than the dyad conversations (p = 0.02, **Figure 1B**). HA processing also affected the overlaps in interaction with hearing status (p = 0.02, **Figure 2C**), with the post-hoc test revealing that a (non-significant) reduction in overlaps for the HI interlocutor without HAs (p = 0.075), combined with a (non-significant) increase (p = 0.1) for the NH interlocutor, causes the difference to be statistically significant (p < 0.01).

3.3 Effects of Noise

The effect of increasing the background noise on the backchannel behavior was investigated by comparing conversations where the HI interlocutor was wearing HAs (Quiet Aided vs Noise Aided in **Figure 1**).

The group size affected the occurrence of turns shorter than 500 ms, as well as the occurrence of overlaps in the same way as for the conversations in quiet: In the dyads, short turns occurred 2.2 times more per minute compared to triads (p < 0.001, **Figure 1A**), while overlaps occurred 1.1 times more per minute in triads compared to dyads (p < 0.01, **Figure 1B**).

The occurrence of overlaps was reduced by 0.5 per minute when adding noise (p < 0.001, **Figure 1B**). owever an interaction between noise and group size (p < 0.001, **Figure 2D**) revealed that the occurrence of overlaps in the dyad conversations were significantly more affected by noise (reduction of 0.85/min, p < 0.001) than in triad conversations (reduction of 0.24/min, p = 0.03). This resulted in a higher occurrence of overlaps for triads in noise compared to dyads in noise (p < 0.001), while the occurrence did not differ between group size in quiet (p = 0.075). Furthermore, the interaction between hearing status and noise (p < 0.01, **Figure 2E**), revealed that while the HI interlocutor decreased the number of overlaps in noise (0.79/min, p < 0.001), adding noise had no effect on the occurrence of overlaps for the NH interlocutors (p = 0.067). For the short turns, the same interaction effect between hearing status and noise (p = 0.015, **Figure 2B**) showed the opposite effect: While NH interlocutors increase the occurrence of short turns (0.2/min, p < 0.001), noise did not affect the HI interlocutors to change backchannel behavior (p = 0.64)

4. DISCUSSION

The backchannel behavior was investigated using three measures estimating the backchannels in dyad and triad conversations between NH and HI interlocutors in different listening conditions. Below we will discuss how well the three measures quantified backchannel behavior, as well as the observed effect of the noise, HA amplification, hearing status, and group size on these measures.

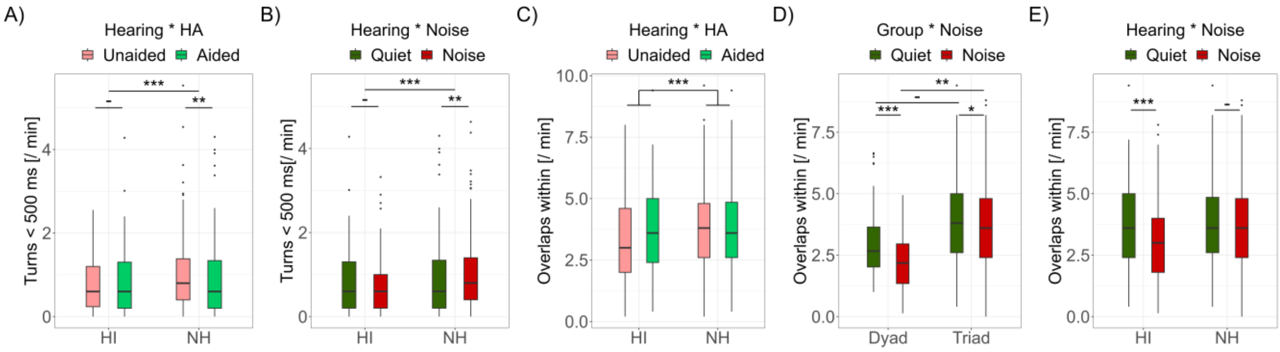

Figure 2: Boxplots of the statistically significant interaction effect, as indicated by the title of each plot. Asterisks indicate level of statistical significance: * p < 0.05, ** p < 0.01, ** p < 0.001, while a dash (-) indicates no statistical difference.

4.1 Automatic Estimation of Backchannel Behavior is Possible

The manual classification of 150 examples of turns shorter than 500 ms and overlaps showed that >80% of the classified examples were backchannels (**Table 1**). Nevertheless, it is interesting to note that the turns shorter than 500 ms better represent backchannels in the dyad (90.7%) compared to the triad conversations (72.2%), while the overlaps better represent backchannels in the triad (88.0%) than in the dyad conversations (76.6%). This suggests that there are some inherent differences in backchannel behavior between the dyad and triad conversations.

The exchange of facts required to solve the Diapix task of the dyad conversations could cause interlocutors to provide backchannels as shorter turns spoken when the entirety of the information has been provided, i.e., after the turn stopped. Consequently, utterances provided within a turn are used to alert the conversational partner of a potential difference in the Diapix images, thereby causing a higher percentage of the overlaps for dyad conversations to be turns rather than backchannels. Taken together, the nature of the backchannels provided in the dyad conversation seem to be short (<500 ms) statements confirming or disconfirming the information just provided by the conversational partner. Indeed, task-oriented conversations have been found to induce more backchannels relative to free conversations because it requires better alignment between interlocutors [15].

A limitation of the recording method of this study is that backchannelling was defined as verbal backchannels only, whereas in real life, backchannels in conversations consist of verbal and visual signals [16].

4.2 Conversation Task, Rather Than Group Size, Affects Backchannel Behavior

The difference observed for the manual classification of the backchannel measures also manifests in the significant effects of group size on the two backchannel measures. Whereas turns shorter than 500 ms occurred around two times more per minute in dyads compared to triads (**Figure 1A**), the number of overlaps occurred one time more per minute for the triads compared to the dyads (**Figure 1B**). The fact that dyads have more short turns per minute and that the short turns also represent backchannel behavior better (**Table 1**) and vice versa for the overlaps of the triad conversations, emphasizes that the two groups size significantly differ in the type and usage of backchannels.

Indeed, when listening to the triad conversations it is clear that backchannels provided in these conversations are of a different nature compared to the dyad conversations: The backchannels provided in triads seem to be longer and include more sentence-like backchannels such as 'I see' or 'yes, that's right' which are longer than 500 ms and therefore not included in the current analysis. If including utterances up to 1000 ms as backchannels and manually classifying them (similar to the result in **Table 1**), 62% of these longer utterances were backchannels in the triad conversations, while only 38.7% of them were backchannels in the dyad conversations. This indeed suggests that backchannels have different characteristics in the dyad and triad conversations, more likely caused by the difference in conversation task rather than number of interlocutors.

4.3 Backchannel Behavior is Affected by Communication Difficulty

In the introduction we speculated that increasing the communication difficulty causes interlocutors to provide fewer backchannels to avoid disrupting listening and because fewer mental resources are available to produce them. Several of our findings concur with this hypothesis, specifically the findings that: 1) The occurrence of short utterances was higher in NH than HI interlocutors across noise levels (**Figure 1A**). 2) Both short utterances and overlaps are more frequent in quiet relative to noise (**Figure 1A+B**). 3) The HI interlocutor reduced the occurrence of overlaps when noise was added (**Figure 2E**). And finally, 4) when being aided the increase in overlaps of the HI interlocutor opposed the decrease observed for the NH interlocutors (**Figure 2C**).

On the other hand, the NH interlocutor increased the number of short utterances in quiet when the HI interlocutors was not receiving HA amplification (**Figure 2A**), an observation that cannot be explained by the hypothesis of increased communication difficulty, as the NH interlocutor does not experience any change in the acoustic scenario in this situation. However, as a conversation is an interactive interplay between interlocutors, it is possible that communication difficulties will not only affect the interlocutor experiencing it, but also the communication partner. One example is the increase in speech levels by NH interlocutors upon not providing HA amplification to the HI interlocutor in both triad and dyad conversations [4,11].

In conclusion, we have shown that it is possible to proximate backchannel behavior by the occurrence of turns shorter than 500 ms and overlaps within another interlocutor's turn. However, these measures are approximations as the nature of the conversation (free or task bound) and number of interlocutors can have a significant influence on the backchannel behavior. Nevertheless, the current work shows that the estimated backchannel behavior is sensitive to changes in the background noise, hearing status, and providing hearing aid amplification to the HI interlocutor. This suggests that although backchannels are often produced and delivered without much consideration, they can convey important information regarding the communication difficulty experienced by the interlocutors.

5. REFERENCES

[1] G. Keidser *et al.*, "The Quest for Ecological Validity in Hearing Science : What It Is , Why It Matters , and How to Advance It," *Ear Hear.*, vol. 41, no. Suppl 1, pp. 5S-19S, 2020.

[2] S. C. Levinson and F. Torreira, "Timing in turn-taking and its implications for processing models of language," *Front. Psychol.*, vol. 6, pp. 1–17, 2015.

[3] A. Sørensen, "The effects of noise and hearing loss on conversational dynamics," Technical University of Denmark, 2021.

[4] E. B. Petersen, E. N. Macdonald, and A. Sørensen, "The Effects of Hearing Aid Amplification and Noise on Conversational Dynamics between Normal-Hearing and Hearing-Impaired Talkers," *Trends Hear.*, vol. 26, pp. 1–18, 2022.

[5] H. Lu, M. F. McKinney, T. Zhang, and A. J. Oxenham, "Investigating age, hearing loss, and background noise effects on speaker-targeted head and eye movements in three-way conversations," *J. Acoust. Soc. Am.*, vol. 149, no. 3, pp. 1889–1900, 2021.

[6] T. Beechey, J. M. Buchholz, and G. Keidser, "Hearing impairment increases communication effort during conversations in noise," *J. Speech, Lang. Hear. Res.*, vol. 63, no. 1, pp. 305–320, 2020.

[7] V. Hazan *et al.*, "How do aging and age-related hearing loss affect the ability to communicate effectively in challenging communicative conditions?," *Hear. Res.*, vol. 369, pp. 33–41, Nov. 2018.

[8] T. Beechey, J. M. Buchholz, and G. Keidser, "Hearing aid amplification reduces communication effort of people with hearing impairment and their conversation partners," *J. Speech, Lang. Hear. Res.*, vol. 63, no. 4, pp. 1299–1311, 2020.

[9] V. Yngve, "On getting a word in edgewise," in *Chicago Linguistics Society*, 1970.

[10] J. Starkey Duncan, "On the Structure of Speaker-Auditor Interaction during Speaking Turns," *Lang. Soc.*, vol. 3, no. 2, pp. 161–180, 1974.

[11] E. B. Petersen, E. Walravens, and A. K. Pedersen, "Real-life Listening in the Lab : Does Wearing Hearing Aids Affect the Dynamics of a

Group Conversation ?," in *Conference: Proceedings of the 26th Workshop on the Semantics and Pragmatics of Dialogue*, 2022, pp. 221–224.

[12] R. Baker and V. Hazan, "DiapixUK: Task materials for the elicitation of multiple spontaneous speech dialogs," *Behav. Res. Methods*, vol. 43, no. 3, pp. 761–770, 2011.

[13] A. Sørensen, "PDF and PSD files of DiapixDK picture materials - adapted Danish version [Data set]," p. http://doi.org/10.5281/zenodo.4638693, 2021.

[14] N. H. de Jong and T. Wempe, "Praat script to detect syllable nuclei and measure speech rate automatically," *Behav. Res. Methods*, vol. 41, no. 2, pp. 385–390, 2009.

[15] R. Fusaroli, K. Tylén, K. Garly, J. Steensig, M. H. Christiansen, and M. Dingemanse, "Measures and mechanisms of common ground: backchannels, conversational repair, and interactive alignment in free and task-oriented social interactions," in *CogSci 2017 - Proceedings of the 39th Annual Meeting of the Cognitive Science Society: Computational Foundations of Cognition*, 2017, pp. 2055–2060.

[16] M. Heldner, A. Hjalmarsson, and J. Edlund, "Backchannel relevance spaces," in *Nordic Prosody: Proceedings of the XIth Conference, Tartu 2012*, 2013, pp. 137–146.