PROCEEDINGS of the
24th International Congress on Acoustics

October 24 to 28, 2022 in Gyeongju, Korea

ABS-0837

Using real group conversations in laboratory testing

Karolina SMEDS[1]; Petra HERRLIN[1]; Eline Borch PETERSEN[2]; Frédéric MARMEL[1];
Florian WOLTERS[1]; Anja Kofoed PEDERSEN[3]

[1] ORCA Europe, WS Audiology, Sweden

[2] Scientific Audiology, WS Audiology, Denmark

[3] Applied Audiological Research, WS Audiology, Denmark

ABSTRACT

There is currently scientific focus on improving the ecological validity of laboratory hearing assessments. Most work has highlighted the selection of listening scenarios and playback using multi-speaker setups. Our work has instead centered on investigating the listening tasks that can be used, specifically real conversations. Various outcomes can be obtained during a conversation. The actual conversation may be analyzed acoustically (e.g., turn-taking timing), linguistically (e.g., repetition behavior) or behaviorally (e.g., head and eye movements). Another option is to stage a conversation and use traditional psychoacoustical tests while conversing (e.g., paired comparisons of two hearing-aid settings). There are also various ways to elicit a conversation. Collaborative puzzle-solving, map-navigation, and spot-the-difference (Diapix) tasks have been used in conversations between two interlocutors. Here, we report on aspects of using real conversations in small groups of interlocutors. In one study, groups of three interlocutors subjectively evaluated four different ways to elicit group conversations, while acoustical analyses were used to explore how the elicitation method affected the conversations. In another study, groups of four interlocutors evaluated four staged scenarios in terms of perceived realism and conversation success. The experimental considerations for using group conversations in the laboratory will be discussed.

Keywords: Group conversation, Outcome measures, Conversation sparkers

1 INTRODUCTION

We want hearing-aid users to experience everyday benefit from their hearing aids. Therefore, evaluation methods that show a high degree of ecological validity, i.e., methods that are indicative of everyday performance, must be used when the aids are developed and fitted. To develop such evaluation methods, knowledge about peoples' auditory reality is important. Which listening situations are frequently experienced, and which are the listening demands in these situations?

In previous work, we investigated the signal-to-noise ratios (SNRs) experienced in the everyday life of 20 hearing-aid users. Generally, the SNRs were higher than in traditional laboratory testing, and many situations that were experienced to vary in listening difficulty, had similar SNRs (1). We further investigated common listening situations that people are encountering, and in the Common Sound Scenarios (CoSS) framework these situations were categorized into seven "task categories" (2). In subsequent studies, using Ecological Momentary Assessments (EMA), we have learned about how often these various types of situations occur in peoples' everyday life, how important they are to hear well in, and how difficult it is to hear in the situations (3).

We have also used EMA to evaluate people's preferred hearing-aid settings in everyday life (3). Two hearing-aid settings were implemented in two hearing-aid programs, and the preferred program was indicated when the participants were prompted to report. The advantage of using EMA is that evaluations are made in the participants self-selected everyday listening situations. However, there are also drawbacks with using EMA. Some difficult and important situations do not occur very often (3), which makes EMA inefficient. It may also be difficult or socially unacceptable to report in certain situations, especially in situations with social interaction (4).

Therefore, there is still a need for laboratory testing for these important, often difficult social interaction situations. When looking at the CoSS framework, it becomes apparent that traditional laboratory testing only targets "focused listening", i.e., listening intently and sometimes repeating back what was heard. Neither real conversations, nor more passive listening situations are normally used in laboratory testing. The current paper will explore ways to broadening laboratory testing to include real conversations.

Nicoras et al. (5) investigated the concept *conversation success* for people with normal and impaired hearing using group concept mapping. Seven factors important for conversation success were identified: 1. Being able to listen easily; 2. Being spoken to in a helpful way; 3. Being engaged and accepted; 5. Sharing information as desired; 5. Perceiving flowing and balanced interaction; 6. Feeling positive emotions; 7. Not having to engage coping mechanisms. These factors can potentially be used when developing outcome measures for hearing-aid satisfaction or benefit in conversation situations.

Here, we will report on some of the activities we have been involved in that contribute to the understanding of conversations and which outcome measures we can develop. First, we will give an overview of ways to study conversations, with a focus on finding outcome measures for hearing-aid performance. Second, we will discuss different ways to spark a conversation in the laboratory, and we will present a study evaluating four conversation sparkers. Third, we will present a pilot study focusing on perceived realism and conversation success in group conversations. Last, some experimental considerations for using group conversations in the laboratory will be discussed.

In the following, the term "scenario" refers to listening situations implemented in the laboratory or clinic. Several aspects of these scenarios need to be designed carefully for laboratory testing to produce ecologically valid findings. A scenario includes an activity (for instance described using the CoSS task categories), an acoustic environment and its implementation. The scenarios should sound (and preferably also look) realistic, but it is also important that both the test participants and the hearing aids behave as they would in a corresponding real-life situation. The activities performed in the scenarios govern the outcome measures used.

2 OUTCOME MEASURES FOR CONVERSATIONS

Conversations are studied by researchers from many different fields, such as linguistics, social psychology, cognitive neuroscience, engineering, and audiology. The purposes of studying conversations differ and different research methods have developed in the various disciplines.

In a linguistic analysis of conversations, researchers may for instance investigate breakdown and repair in a conversation (6), lexical or syntactical alignment between interlocutors, or the use of backchanneling (verbal or nonverbal responses to a talker indicating for instance attention or agreement). The acoustic speech signals can be analyzed by looking for instance at speech levels, speech rate, and turn-taking timing (7). Body posture, gestures (8), head movements, eye gaze (9) and coordination of body movements can indicate difficulties or ease in a conversation. With the use of EEG or other electrophysiological correlates, selective attention (10), attention switching (11), and attention drifting (12) can be investigated.

Some of these investigation methods have indicated differences in conversation behavior between test participants with normal and impaired hearing. Researchers have shown that conversations between one young interlocutor with normal hearing and one older interlocutor with impaired hearing can work well, but that the communication effort was higher when the interlocutor with impaired hearing was unaided than when hearing-aids were used (7, 13).

In general, the investigation methods mentioned above can be difficult to use for instance during hearing-aid development, when a new hearing-aid feature will be compared to an old one. However, by tying some of the measures mentioned above to perceived conversation success (5), we will learn more about which measures could be useful as outcome measures.

Another outcome measure strategy is to stage a conversation situation and use traditional psychoacoustical methods to evaluate for instance hearing-aid preference. In our research group we did so when we developed the Live Evaluation of Auditory Preference (LEAP) laboratory test (14). The test used six mandatory test scenarios, selected based on the CoSS framework, and a few self-selected scenarios (derived from experienced listening situations during an EMA study). The scenarios were implemented using a basic loudspeaker setup. For the conversation situations, we used real conversations between the test participant and one or two test leaders. During the conversations, the test participants compared two hearing-aid programs using a paired-comparison paradigm and reported on their preferred setting.

3 CONVERSATION SPARKERS

Different methods have been used to initiate or spark conversations in the laboratory. For two talkers, Baker and Hazan (15) and Petersen et al. (7) used Diapix material, where the interlocutors have one version each of similar pictures and they together solve a "spot the difference task". Beechey et al. (16) elicited naturalistic conversation by providing the interlocutors with one version each of a puzzle with 10 by 10 items, where the task was to find a way from the upper left corner of the puzzle to the lower right corner by combining the information from the two sets of the puzzle. Doherty-Sneddon et al. (17) used two versions of a map and the task was to give and take instructions about a route to find a treasure. These tasks seem to have worked well in the mentioned studies. However, they are probably less suitable for group conversation. Further, with a strong visual focus on a picture, a puzzle or a map, the interlocutors might not pick up visual cues from the rest of the group or might exhibit an unnatural behavior regarding for instance eye contact.

We therefore suggested and evaluated four other group conversation sparkers: 1. *Photographs* with keywords (example: a photograph of a huge pan of paella with the key words "food, restaurant, vacation"), 2. *Consensus* tasks (example: "Agree on a dinner menu exclusively with dishes none of you like"), 3. *Timeline* tasks (order four historical events chronologically), and 4. *Free conversation*. Ten examples of each sparker were evaluated by four groups with three conversation partners in each group. Test participants were recruited among unacquainted colleagues from different departments and without insights into the project. The four types of sparkers were evaluated subjectively using questionnaires and ranking of the sparkers. In addition, outcome measures such as speaking time, utterance length, turn-taking information, and speech levels were used.

Overall, the participants judged all sparkers to be natural, engaging and with a good flow (7-8 on a 10-point scale). The results showed that the Timeline task was least suitable. It had significantly lower ratings on the question "Did the task make you feel that you could contribute to the conversation" ($p<0.05$). It also created the most unbalanced conversation in terms of test participants' speaking time (Figure 1). The task created more silence, more very short utterances, longer gaps in turn-taking, and fewer turn-takings than the other sparkers. The fact that there is a "right answer" to the task probably contributed to it being less successful as a sparker. Although the outcomes of the Free speech task were not found to differ from the remaining tasks, the participants ranked the Free speech as the least suitable conversational sparker (although not statistically significant). Participants commented that not providing any conversational sparker would make it difficult for some participants/groups to hold a conversation. Both the Photos with key words and the Consensus task seemed appropriate as group conversation sparkers.

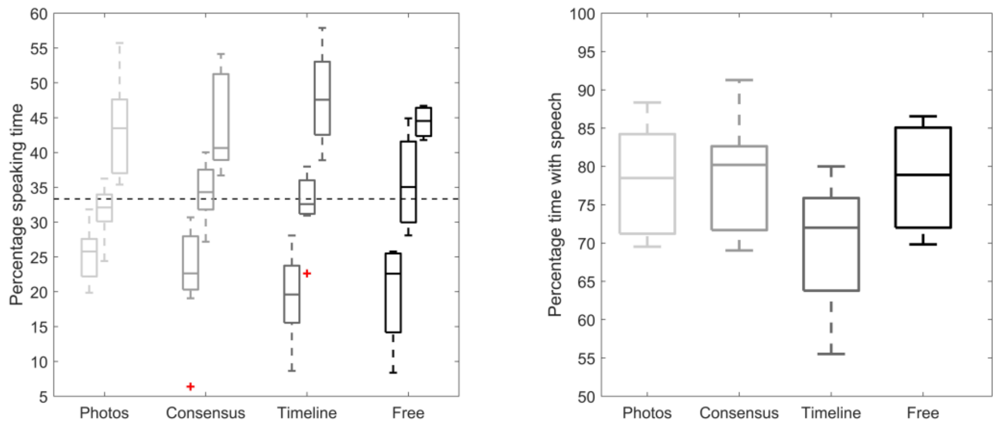

Figure 1 – The left panel shows the speaking time for the three participants in each group. For each conversation sparker, the data for the participants who spoke least are presented first and the data for the participants who spoke most are presented last. The right panel shows the overall proportion of speaking time for each sparker. Medians are represented by the horizontal line within each box. Outliers are defined as values outside 1.5 times the box length, and the whiskers extend to the highest and lowest values when outliers are excluded.

4 LEAP Group Conversation

In a pilot study, a group conversation version of the LEAP test was investigated. The study aimed at producing a set of group conversation scenarios in an office setup and evaluate perceived realism and conversation success. In each conversation group four people participated, two with hearing impairment and two with normal hearing.

4.1 Method

4.1.1 Research participants

20 test participants (5 groups) were recruited for the experiment. The inclusion criteria were age between 65 and 79 years, fluent Swedish speakers with good eyesight (after correction). The inclusion criteria for participants with impaired hearing were moderate hearing loss in both ears and hearing-aid users for more than 6 months. The normal-hearing participants had self-reported normal hearing and had passed an online speech-in-noise hearing test (horseltestaren.se). In one of the five group conversations, one of the participants with normal hearing was unable to attend and was replaced by a colleague. Her data were excluded from the analyses. Audiograms for the two participant groups are shown in Figure 2.

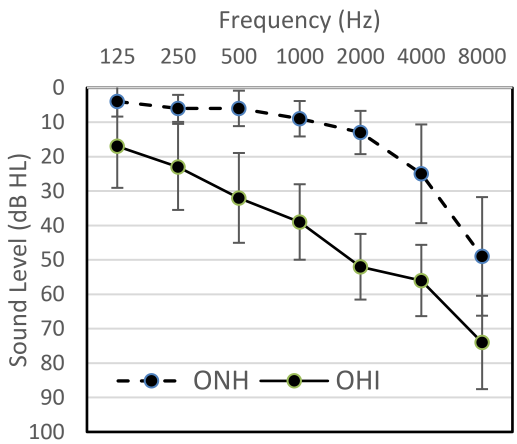

Figure 2 – Average audiogram (best ear) and standard deviation for older normal-hearing (ONH) and older hearing-impaired (OHI) test participants.

4.1.2 Setup and Sound Files

Four listening scenarios were implemented. Sound files were selected from the Ambisonic Recordings of Typical Environments (ARTE) database (18). Background sound levels were set based on the levels reported when the sound files were recorded or sound levels reported in our SNR study (1), and informal listening by a group of colleagues with normal hearing.

1. Dinner at home (ARTE: 04_Living_Room_MOA_31ch, 52-55 dBA)
2. Dinner in restaurant (ARTE: 12_Food_Court_1_MOA_31ch, 57-60 dBA)
3. Business meeting (only video projector noise, 44-46 dBA)
4. Party (ARTE: 09_Dinner_party_MOA_31ch, 70 dBA)

Three of the scenarios were implemented in a meeting room (345×693 cm). The room was acoustically treated with a suspended mineral fiber ceiling (pending down 55 cm from the 379 cm inner ceiling) and a wall-hanging treatment (238×116 cm). Thin curtains (floor to ceiling) covered a 3 m long window section on one wall. Sound files were played from a laptop equipped with a RME Fireface 802 sound card. Four loudspeakers (Genelec 8030 CP), placed in and oriented towards the corners, were used. The test participants were placed at a table, two persons (one with impaired hearing and one with normal hearing) at each side of the table.

The Party scenario was implemented in a kitchenette area with an adjacent room and corridors. The main area was 354×336 cm with a ceiling height (suspended mineral fiber) of 324 cm. In one section (80 cm closest to the kitchen part) the ceiling was lower (220 cm). Two loudspeakers placed at the two ends of the kitchenette area were used. In this scenario, the participants were standing and could place themselves freely.

4.1.3 Conversation Sparkers

Conversation sparkers were selected based on the study described above. For scenarios 1, 2, and 4 Photos with key words were used. For scenario 3, a Consensus task was used, and all test participants were asked to act as "secretaries" for the business meeting and note down what the group agreed on.

4.1.4 Subjective Evaluations

After each conversation, the participants used a tablet (Samsung Galaxy Tab S6 Lite) to rate realism and conversation success using a 5- or 6-step Likert scale (if not otherwise indicated below).

Realism:

1. How realistic did you perceive the situation?

Conversation success:

2. How easy was it to hear in the situation?
3. How engaged did you feel in the conversation?
4. How well were you able to say what you wanted to say?
5. How well did you think that the conversation was flowing?
6. How successful did you think that the conversation was?

After the last test scenario, the group discussed how realism could be increased in the scenarios and what constitutes conversation success.

4.2 Results

The focus of this pilot study was to evaluate perceived realism and conversation success in the four scenarios.

4.2.1 Realism

The Party was judged to be most realistic and the Dinner at home least realistic (Figure 3). In the discussions after the testing, the participants commented that standing up in the kitchen area and being able to move freely in relation to the other interlocutors contributed to the perceived realism in the Party situation. The background noise type and level also contributed to the realism, but a few participants with hearing impairment commented that these situations can be even louder in real life. The Business meeting was judged to be realistic because the testing took place in an actual meeting room.

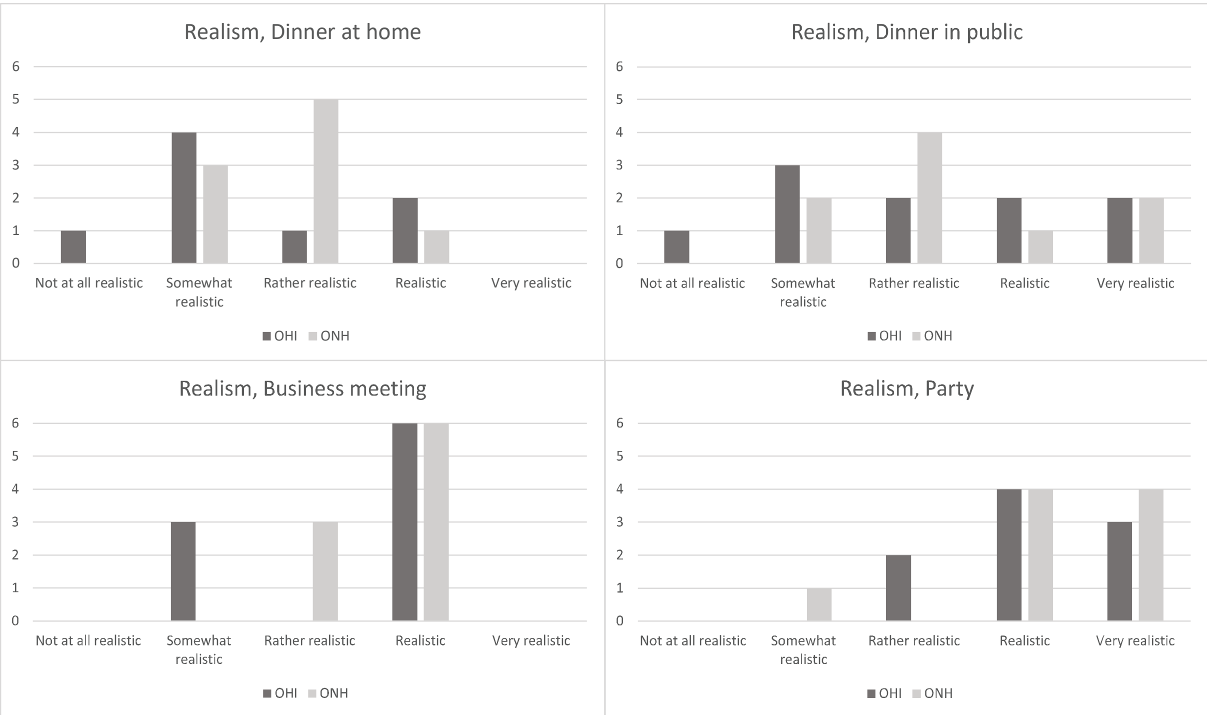

Figure 3. Perceived realism in the four scenarios used. OHI = Older participants with hearing impairment. ONH = Older participants with normal hearing.

The two dinner situations were judged to be less realistic, and in the subsequent discussions the participants said this mainly was related to the type and level of the background noise. Test participants, especially those with impaired hearing, commented that these situations are usually more difficult in real life. They thought that the background noise was too static and that it should include transient sounds from silverware, porcelain, and glasses. It could also include someone coughing or yelling, a chair being pushed, a telephone ringing, a baby crying or similar. The background sound levels were also judged to be too low.

4.2.2 Conversation success

Here, the results of all questions related to conversation success are presented. Questions 2-5 are related to four of the factors found to be important for conversation success in the study by Nicoras et al. (5), whereas question 6 asks for conversation success directly.

The results of the question "How difficult/easy was it to hear in the situation?" (question 2) showed the expected difference between the two participant groups (Figure 4). The participants with normal hearing generally rated the question higher than the participants with impaired hearing. The situations that were perceived easiest to hear in were Dinner at home and Business meeting, where the noise levels were low. None of the situations were described as "very difficult" to hear in. The participants, in particular those with impaired hearing, commented that listening is more difficult in real life. It was also slightly easier for the participants with normal hearing to say what they wanted (question 4) and they perceived slightly higher conversation flow (question 5). The conversation flow was similar for all scenarios.

The test participants showed the highest engagement (question 3) in the Business meeting (where the consensus task was used as a sparker) and Dinner in public (where picture cards with keywords were used). Both participant groups showed similar degree of engagement.

In terms of conversation success (question 6), dinner at home was judged to be the least successful conversation. This scenario was presented first for all groups, which might explain the result. There was no difference in judged conversation success between the two participant groups. In the subsequent discussion about the conversations, the test participants agreed that the most important factor for conversation success was that everybody joins, but it was not judged important that everybody spoke equally much. They also mentioned that it is important that all interlocutors "give and take", i.e., both listen and speak and that everybody should be included in the conversation in a natural manner.

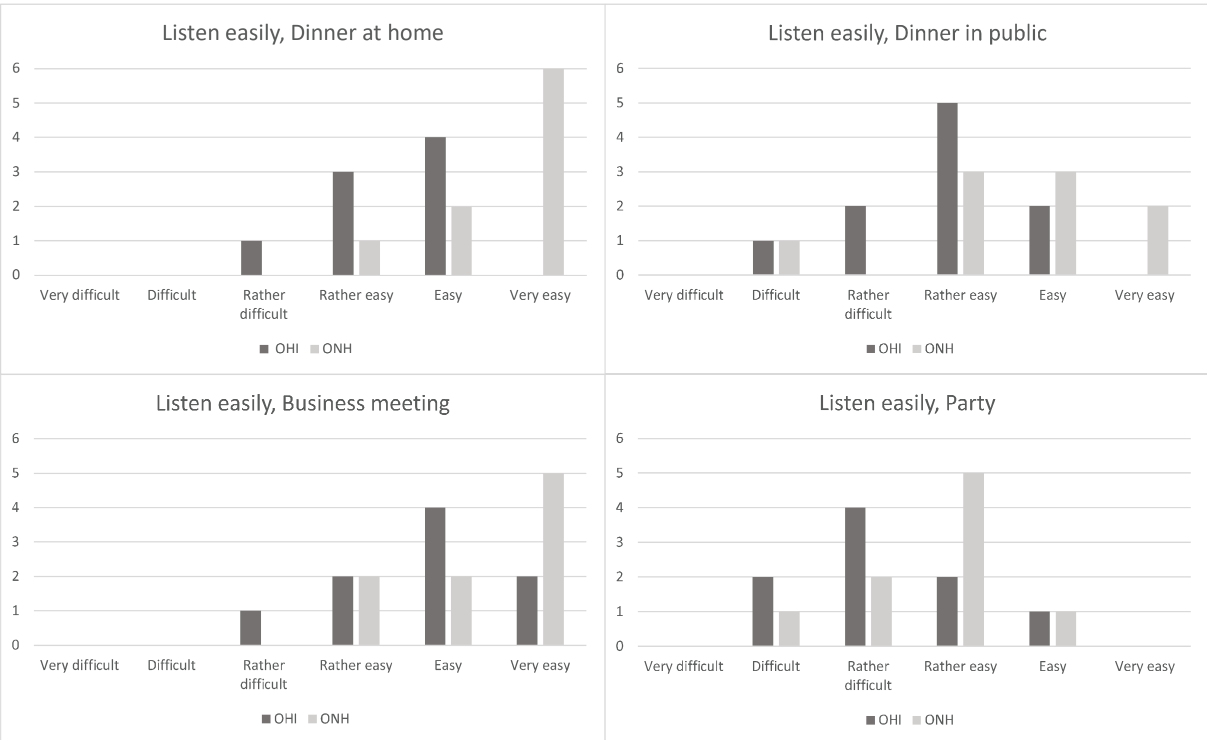

Figure 4. Perceived ability to listen easily in the four scenarios. OHI = Older participants with hearing impairment. ONH = Older participants with normal hearing.

5 EXPERIMENTAL CONSIDERATIONS

In the two experiments described above, office rooms and a limited number of loudspeakers were used. In the LEAP Group Conversation study, perceived realism was investigated, but detailed physical measurements were not made to ensure that the sound fields in the four scenarios resembled the sound fields in corresponding everyday listening situations.

The obvious way to improve the acoustical representation of the background sounds would be to use an anechoic test room with many loudspeakers, using for instance higher order ambisonics to playback multi-microphone recordings faithfully (19). There are two problems with this approach for group conversations. First, it will be difficult to place a group of interlocutors in the sound field in such a way that the acoustic representation is accurate for all participants. The so-called “sweet spot”, where the reproduction error is small, is usually too small to host all participants. It might be worse to sit far outside the sweet spot in an ambisonics playback situation than in a more simplistic setup in a room with some reverberation, which smooths the sound field. Second, the voices of the interlocutors will sound as if they do not belong to the intended scenario if the background sounds were recorded in a reverberant environment. If the scenario is one that could take place in an environment similar to, for instance, the office and kitchenette rooms used in the Group LEAP study, it might be better to test where the representation of the interlocutors’ voices is more realistic.

Another option is to use so called Variable Room Acoustics Systems (VRAS), where microphones are picking up the interlocutors’ voices, which are then processed so that the voices match the scenario (20). Researchers need to ensure not only that the scenario sounds realistic, but also that the physical sound field is accurate, since the interlocutors will hear their communication partners as a combination of direct sound from the communication partners’ mouth and the processed version from the loudspeakers.

Another question is related to the visual realism. In the LEAP Group Conversation study described above, the Party and the Business meeting were described as the most realistic. For these situations, the visuals of the scenario corresponded roughly to what it would be in real life. Standing with a group of people in a kitchen area at a party or sitting down at a business meeting in a room that is used for business meetings is realistic. For the dinner situations, further visual input could potentially have increased the perceived realism.

6 DISCUSSION AND CONCLUSIONS

We have presented some of our ongoing work where we include real conversations in laboratory testing and strive to develop outcome measures indicating how hearing aids function in real group conversations. Many ways to study conversations have been presented. Some of these can potentially be used as outcome measures during hearing-aid development if they can be tied to conversation success. Alternatively, traditional psychoacoustical test methods, such as ratings or paired comparisons of certain attributes, can be used in staged group conversations. Different ways to elicit or spark group conversations have been evaluated. Photos with keywords and Consensus questions seemed to work well. We further staged four group conversation scenarios, with four interlocutors in each scenario. The scenarios were evaluated in terms of perceived realism and conversation success. The results will be used to refine our conversation scenarios. Finally, we discussed the experimental considerations when creating group conversations in the laboratory.

ACKNOWLEDGEMENTS

We want to thank the participants who contributed to this research.

REFERENCES

1. Smeds K, Wolters F, Rung M. Estimation of signal-to-noise ratios in realistic sound scenarios. J Am Acad Audiol. 2015;26(2):183-96.
2. Wolters F, Smeds K, Schmidt E, Christensen EK, Norup C. Common Sound Scenarios: A context-driven categorization of everyday sound environments for application in hearing-device research. J Am Acad Audiol. 2016;27(7):527-40.
3. Smeds K, Gotowiec S, Wolters F, Herrlin P, Larsson J, Dahlquist M. Selecting scenarios for hearing-related laboratory testing. Ear Hear. 2020;41 Suppl 1:20S-30S.
4. Schinkel-Bielefeld N, Kunz P, Zutz A, Buder B. Evaluation of Hearing Aids in Everyday Life Using

Ecological Momentary Assessment: What Situations Are We Missing? Am J Audiol. 2020;29(3S):591-609.
5. Nicoras R, Gotowiec S, Hadley LV, Smeds K, Naylor G. Conversation success in one-to-one and group conversation: a group concept mapping study of adults with normal and impaired hearing. International Journal of Audiology. 2022:1-9.
6. Lind C, Hickson L, Erber NP. Conversation repair and adult cochlear implantation: A qualitative case study. Cochlear Implants International. 2006;7(1):33-48.
7. Petersen EB, MacDonald EN, Josefine Munch Sorensen A. The Effects of Hearing-Aid Amplification and Noise on Conversational Dynamics Between Normal-Hearing and Hearing-Impaired Talkers. Trends Hear. 2022;26:23312165221103340.
8. Ter Bekke M, Drijvers L, Holler J. The predictive potential of hand gestures during conversation: An investigation of the timing of gestures in relation to speech. 2020.
9. Hadley LV, Brimijoin WO, Whitmer WM. Speech, movement, and gaze behaviours during dyadic conversation in noise. Scientific reports. 2019;9(1):1-8.
10. Ding N, Simon JZ. Emergence of neural encoding of auditory objects while listening to competing speakers. Proceedings of the National Academy of Sciences. 2012;109(29):11854-9.
11. Miran S, Akram S, Sheikhattar A, Simon JZ, Zhang T, Babadi B. Real-time tracking of selective auditory attention from M/EEG: A bayesian filtering approach. Frontiers in neuroscience. 2018;12:262.
12. Lesenfants D, Francart T. The interplay of top-down focal attention and the cortical tracking of speech. Scientific reports. 2020;10(1):1-10.
13. Beechey T, Buchholz JM, Keidser G. Hearing Aid Amplification Reduces Communication Effort of People With Hearing Impairment and Their Conversation Partners. J Speech Lang Hear Res. 2020;63(4):1299-311.
14. Smeds K, Larsson J, Dahlquist M, Wolters F, Herrlin P. Live Evaluation of Auditory Preference, a Laboratory Test for Evaluating Auditory Preference. J Am Acad Audiol. 2021;32(8):487-500.
15. Baker R, Hazan V. DiapixUK: task materials for the elicitation of multiple spontaneous speech dialogs. Behavior research methods. 2011;43(3):761-70.
16. Beechey T, Buchholz JM, Keidser G. Eliciting naturalistic conversations: A method for assessing communication ability, subjective experience, and the impacts of noise and hearing impairment. J Speech Lang Hear Res. 2019;62(2):470-84.
17. Doherty-Sneddon G, Anderson A, O'malley C, Langton S, Garrod S, Bruce V. Face-to-face and video-mediated communication: A comparison of dialogue structure and task performance. Journal of experimental psychology: applied. 1997;3(2):105.
18. Weisser A, Buchholz JM, Oreinos C, Badajoz-Davila J, Galloway J, Beechey T, et al. The Ambisonic Recordings of Typical Environments (ARTE) Database. Acta Acustica united with Acustica. 2019;105(4):695-713.
19. Hohmann V, Paluch R, Krueger M, Meis M, Grimm G. The Virtual Reality Lab: Realization and Application of Virtual Sound Environments. Ear Hear. 2020;41 Suppl 1:31S-8S.
20. Klatte M, Lachmann T, Meis M. Effects of noise and reverberation on speech perception and listening comprehension of children and adults in a classroom-like setting. Noise Health. 2010;12(49):270-82.