PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

LEAP, a new laboratory test for evaluating auditory preference

Karolina SMEDS[1]; Martin DAHLQUIST; Josefina LARSSON; Petra HERRLIN; Florian WOLTERS

[1] ORCA Europe, Widex A/S

ABSTRACT

When hearing-device signal processing is evaluated in the laboratory, the testing usually incorporates speech testing or sound-quality evaluations with pre-recorded material. Neither real speech communication, nor more passive listening situations are usually included. A new test, the Live Evaluation of Auditory Preference (LEAP), was developed with the aim to broaden the scope of laboratory testing by focusing on a range of listening tasks. Real conversations between the test participant and one or two test leaders are central to the method. Furthermore, ecologically valid test scenarios with focused listening (such as watching TV) and scenarios with passive listening are also included. Experienced hearing-aid users evaluated the test using paired comparisons of preference for two hearing-aid settings. To validate the method, the results from the LEAP test were compared to results obtained in the field, where a smartphone prompted the assessors to make paired comparisons of preference using Ecological Momentary Assessments (EMA). The comparison of the results from the field and the laboratory was satisfactory. To make the results of a laboratory test indicative of real-life performance, a broadened test scope seems important, and the LEAP could be a step forward in that direction.

Keywords: Hearing aids, Paired comparisons, Ecological Momentary Assessments

1. INTRODUCTION

Evaluation of hearing-device benefit or preference can be done for various purposes. For example, a manufacturer might want to compare feature settings to find the most preferred alternative, or in the clinic, a dispenser might want to evaluate the benefit of a pair of hearing devices for an individual hearing-device user. There are several ways to subjectively assess the perceived hearing-device benefit or preference.

Laboratory tests such as measures of speech intelligibility and ratings of sound quality are often used, but most speech intelligibility tests are not particularly sensitive to differences in gain-frequency response shape (1), or might produce misleading results for various features, such as noise reduction (2). Despite the lack of benefit of noise reduction in terms of speech intelligibility, listeners may still prefer noise-reduction processing in terms of sound quality or comfort (3).

To capture this preference, individuals can rate preference or other perceptual attributes (such as speech clarity, listening comfort, and loudness). A signal-processing scheme can either be assessed in isolation or rated relative to a fixed reference. When two signal-processing schemes are directly compared, a paired-comparisons test paradigm is employed. Paired comparisons have proven to be sensitive for evaluation of small signal-processing differences (4).

A new laboratory test called the Live Evaluation of Auditory Preference (LEAP) was developed. The test includes preference judgments for a range of listening scenarios. These scenarios were selected based on the Common Sound Scenarios framework (5) (see below). The goal of developing the laboratory-based LEAP was to produce ecologically valid results that are comparable to real-life preferences. To evaluate this goal, the results of the LEAP test were compared to data obtained in the test participants' ordinary life using Ecological Momentary Assessments (6) (see below).

1.1 Common Sound Scenarios – CoSS

When designing a laboratory test that can be used to predict real-life performance, several factors need to be considered. The test scenarios need to be realistic in terms of stimuli selection, in reproduction of the stimuli, and in the listeners' tasks. The intentions and tasks that listeners have in a variety of listening situations were investigated by Wolters et al. (5). Based on a literature study, the authors used ten audiological articles that described listening situations people with hearing impairment

[1] karolina.smeds@orca-eu.info

encounter. The authors then looked at the intention the listeners had in the described situations and three intention categories were formed based on the data. These intention categories were further divided into seven task categories (Table 1). The created framework was called the Common Sounds Scenarios (CoSS).

Table 1. The CoSS framework. Based on data in 10 audiological articles, three intention categories were created and subdivided into seven task categories.

Intention Categories	**Task**
1. Speech Communication	1. Two people having a conversation 2. Several people having a shared conversation 3. Two or more people having a shared conversation through a communication device
2. Focused Listening (without participating in a conversation)	4. Live sounds – Focused listening to sound without being able to control the sound source 5. Media device – Focused listening to sounds while being able to control the sound source
3. Non-specific	6. Monitoring surroundings 7. Passive listening

1.2 Ecological Momentary Assessments – EMA

In clinical evaluations, interviews and questionnaires are often used to evaluate hearing-device benefit. However, one limitation of this method is recall bias, because it requires the hearing-device user to remember and generalize real-life experiences. A real-life test method that overcomes this limitation is Ecological Momentary Assessments (EMA). In EMA, recall bias is minimized because assessments are made in a participant's real life when a situation occurs. Furthermore, by making assessments several times a day, patterns in the data can be found.

With the rise in the use of smartphones, EMA methodology has become increasingly popular. It is now possible to prompt test participants to make assessments either at specific times or when certain events happen, and to use conditional questionnaires, in which questions are appropriate for the specific situation. The construct validity of the EMA method for audiology research has been established by Wu et al. (7) and Timmer et al. (8).

However, the EMA test paradigm is not without limitations. It is a time-consuming way to learn about hearing-device signal-processing differences. In addition, for features that are supposed to help the hearing-device wearer only in particular situations, it might be ineffective to wait for these situations to occur in everyday life. There is, therefore, a need for laboratory tests that are time-efficient and controlled. Preferably, the results of these laboratory test should correspond with the outcome of more time-consuming field trials.

1.3 Development of the Live Evaluation of Auditory Preference – LEAP

When developing the LEAP test, one goal was to create a test that is indicative of real-life performance. Listening scenarios were selected based on the CoSS framework and included speech communication, focused listening, and more passive listening situations. In the current study, communication scenarios were emphasized. The test participants had conversations with the test leader, and for certain scenarios with a second conversation partner. With real conversations, the test provides an opportunity to include evaluation of the test participant's own voice, which might be central for certain research questions. The test setup also includes realistic visual cues and social pressure to hear, understand, and participate in the conversation, which increases the complexity of the test, especially since the evaluation of hearing-device preference is done while the conversation is ongoing. Conversations are inspired by choosing a picture from a stack of cards that display pictures of, for example, well-known geographical places, food and animals. If the chosen card does not elicit a conversation, a new card can be drawn.

Another goal was to make a test that is sensitive to small differences in hearing-device signal processing. Therefore, the current version of the test uses paired comparisons, which have been shown to be able to detect small differences. In the current study, the evaluated attribute was preference, but the test can also be used to evaluate, for instance, speech clarity or comfort.

The LEAP scenarios can take place in various types of pre-recorded background noise and various loudspeaker setups can be used. The current study focused on the feasibility of the main LEAP features, and a simple loudspeaker setup was used.

1.4 Aim and research questions

The aim of this study was to evaluate the feasibility of the LEAP method and to compare evaluations of rated preference using paired comparisons in the laboratory (LEAP methodology) and in the field (EMA methodology). The study answered the following research questions:

1. Is the LEAP method feasible to use in a paired-comparison paradigm?
2. Does the LEAP method produce results that are comparable to real-life preferences, evaluated using EMA?

2. METHODS

For two weeks, ten experienced hearing-aid users described their auditory reality and compared two hearing-aid programs using EMA in their everyday life. After the two-week period, the participants compared the same two hearing-aid programs using LEAP methodology in the laboratory.

2.1 Participants

The study included ten participants, five women and five men, each with binaural sensorineural hearing loss (gently downward sloping audiograms; PTA4: 31-53 dB HL, average 44 dB HL). Their ages ranged from 53 to 87 years (median 74 years). All were experienced users of binaural hearing aids (from various manufacturers). After completing the study, the participants received vouchers and hearing-aid batteries to a total value of SEK 1200 (approximately $120). Travel expenses were reimbursed.

2.2 Hearing aid fitting

Widex UNIQUE 440 receiver-in-the-ear hearing aids were fitted bilaterally with appropriate receivers and standard domes (7 pairs of tulip domes, 3 pairs of double domes). Two settings were compared. The two settings were used in hearing-aid programs 1 and 2 (half of the participants had setting A in program 1 and the other half had setting B in program 1). The test participant did not know which program contained which setting. A remote control (Widex RC-DEX) was used for program changes. The settings differed in their gain-frequency response shape; setting B provided on average 6-8 dB less gain than setting A in the 1-2 kHz range (Figure 1).

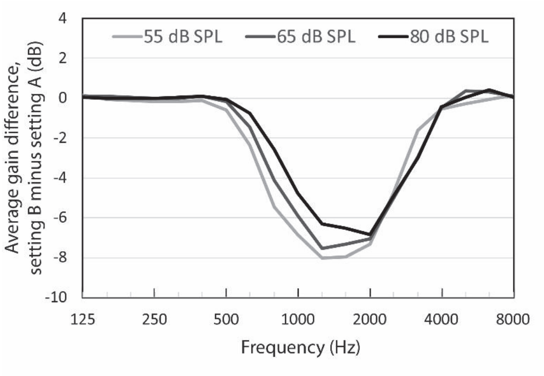

Figure 1. Long-term average gain difference between settings B and A, averaged across all participants. Gain measurements made using a speech signal at three input levels, 55, 65, and 80 dB SPL.

2.3 Smartphone

The participants were provided with Motorola G4 Play smartphones. The phones were equipped with a reminder application for prompting, and a questionnaire implemented using Google forms. Individual, anonymized user accounts were created so that the collected data could not be connected to a specific person without a decoder key.

2.4 Field Trial Procedure – EMA

In the field, the participants provided data on auditory reality and made paired comparisons of the two hearing-aid programs for two weeks. They were prompted to fill out the questionnaire every 90 minutes during the day. The participants could also make self-initiated, un-prompted entries.

When the smartphone alarm went off, the participants were instructed to make paired comparisons of the two hearing-aid programs using the remote control. Then they filled out a two-section questionnaire.

The first section focused on auditory reality. Here, the participants indicated if the report was prompted or self-initiated. They described where they were (e.g., on a bus, in the kitchen) and the situation/activity (e.g., reading a book, talking to my daughter while cooking). They then classified the situation into one of the seven task categories used in the CoSS framework.

In the second section, the test participants indicated their preferred program (1 or 2). They could also answer that they could not hear a difference between the programs, or that they could hear a difference but had no preference. When the participants returned to the laboratory after the field-trial period, they were asked about their overall preference, by answering the question "If you could just keep one of the settings, which one would you keep?".

2.5 Laboratory Trial Procedure – LEAP

To evaluate the basic LEAP principles, paired comparisons of preference were performed using a simple loudspeaker setup in an ordinary office room (without special acoustical treatment, Figure 2). In the current study, the following five mandatory tests scenarios were included:

1. Conversation in quiet, 2 people (CoSS task category 1)
2. Conversation in canteen noise, 2 people (CoSS task category 1)
3. Conversation in quiet, 3 people (CoSS task category 2)
4. Watching TV (CoSS task category 5, focused listening through media)
5. Vacuum cleaning with small macaroni on the floor (CoSS task category 6, monitoring)

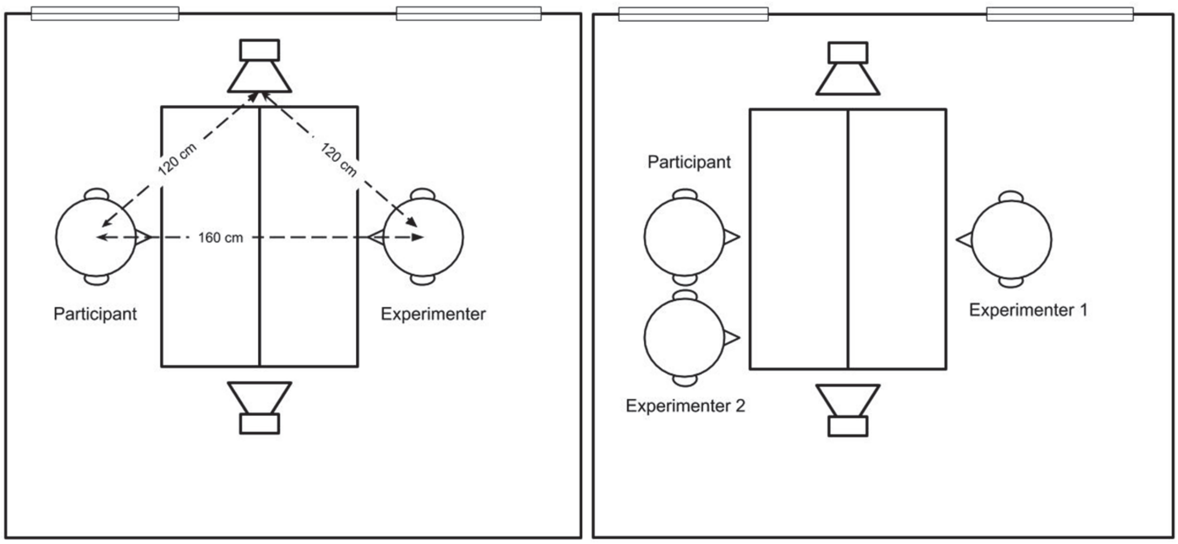

Figure 2. LEAP test setups. The left panel shows the setup for test scenario 1 and 2 (Conversation 2 people). The right panel shows the setup for test scenario 3 (Conversation 3 people).

In addition to these mandatory test situations, the participants studied a list of their own listening situations, encountered during the EMA phase, and selected the two most common, the two most important, and the two most difficult situations that occurred during the field trial. If these self-selected situations were not covered by the mandatory test scenarios, the test leader tried to include them in the LEAP test. However, the room, the simple loudspeaker setup, and the availability of conversations partners posed limitations to what could be simulated.

The LEAP test used the same hearing-aid programs, hearing-aid remote control and smartphone that was used in the EMA test. The smartphone questionnaire was modified, but the paired-comparison section was the same as in the EMA part of the study.

The LEAP test round (with mandatory and individually selected test scenarios) was presented twice, directly after each other. Results from the two rounds were pooled for the general analysis, but the test design also allowed for a test-retest analysis.

In a final interview, the test participants gave their opinion about the LEAP test methodology and indicated if they thought the LEAP scenarios resembled their commonly experienced listening situations.

3. RESULTS

For the LEAP test, on average two individually selected test scenarios were added to the five mandatory scenarios. Scenarios from CoSS categories 2 (Speech communication with more than 2 people) and 7 (Passive listening) were most commonly added to the mandatory scenarios in order to cover participants' selection of their most important, difficult and common listening situations. On average, the test participants spent 4 minutes (range 3-5 minutes) per test scenario. On average, the whole test took 56 minutes.

In total, the participants provided 1044 EMA entries during the 2-week field-trial period, i.e., on average 7 entries per day. The person who reported the least, had on average 4 entries per day. For 64% of the entries, the participants had made paired comparisons. For 27% of the cases, they did (for various reasons) not make paired comparisons, and in 9% of the cases they had not used their hearing aids (but still described their auditory reality). 89% of the entries were prompted and 11% were self-initiated.

3.1 Auditory Reality

The auditory reality data are summarized in Figure 3, where the proportion of entries is shown in the three CoSS intention categories (left panel) and in the seven CoSS task categories (right panel). 30% of the entries were in the speech communication intention category and most of these were communication with one other person. 23% of the entries were in the focused listening intention category, and most of these were focused listening through a media device, most commonly a TV. Almost half of the entries were in the non-specific category, mainly in that of passive listening.

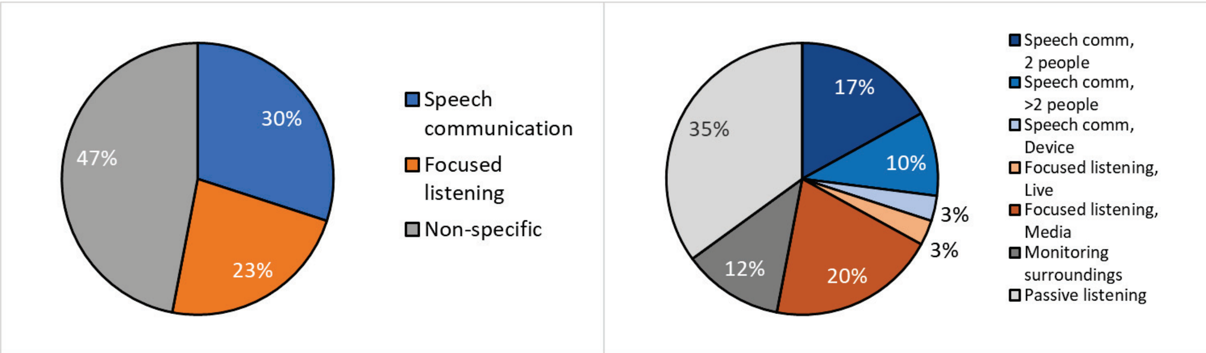

Figure 3. The test participants' auditory reality described using the three CoSS intention categories (left panel) and the seven CoSS task categories (right panel).

3.2 Paired comparisons of preference

The results of the paired comparisons for the LEAP method are shown in Figure 4, left panel. The right panel shows the corresponding results obtained with EMA. In both cases, the distribution is based on all replies (two test rounds for LEAP and all available EMA data). The general preference pattern using the two methods is quite similar, even if the preference for Setting A is clearer in the LEAP results than in the EMA results. Most of the participants could hear a difference between the programs, and the proportion of "hear no difference" and "hear difference but have no preference" is similar for the two methods. The test-retest data for LEAP showed the same preference for the two presentation rounds (mandatory test cases) in 80 % of all comparisons.

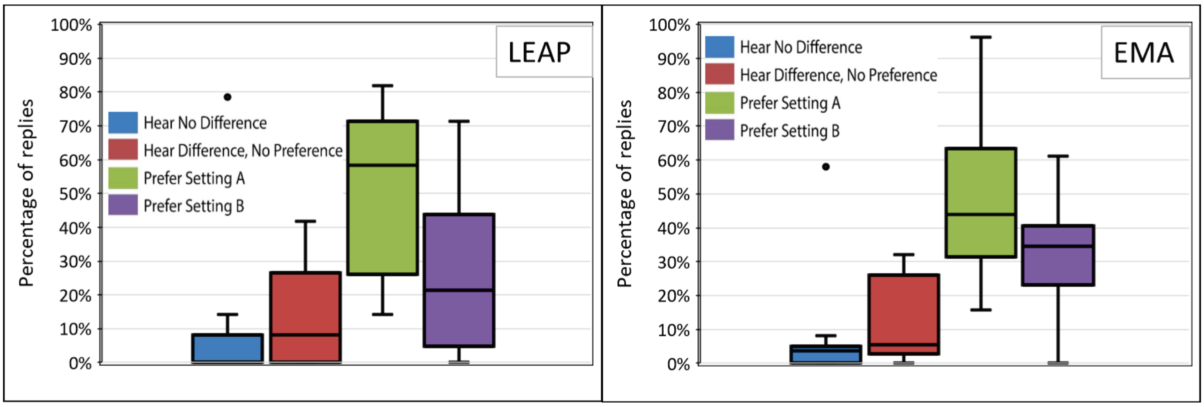

Figure 4. The distribution of preference for the two hearing-aid settings for LEAP data (left panel) and EMA data (right panel). For each participant, the percentage of entries for the four response alternatives were calculated across all test scenarios (LEAP) and entries (EMA). The boxes show inter-quartile values across all participants. Medians are represented by the horizontal line within the 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.

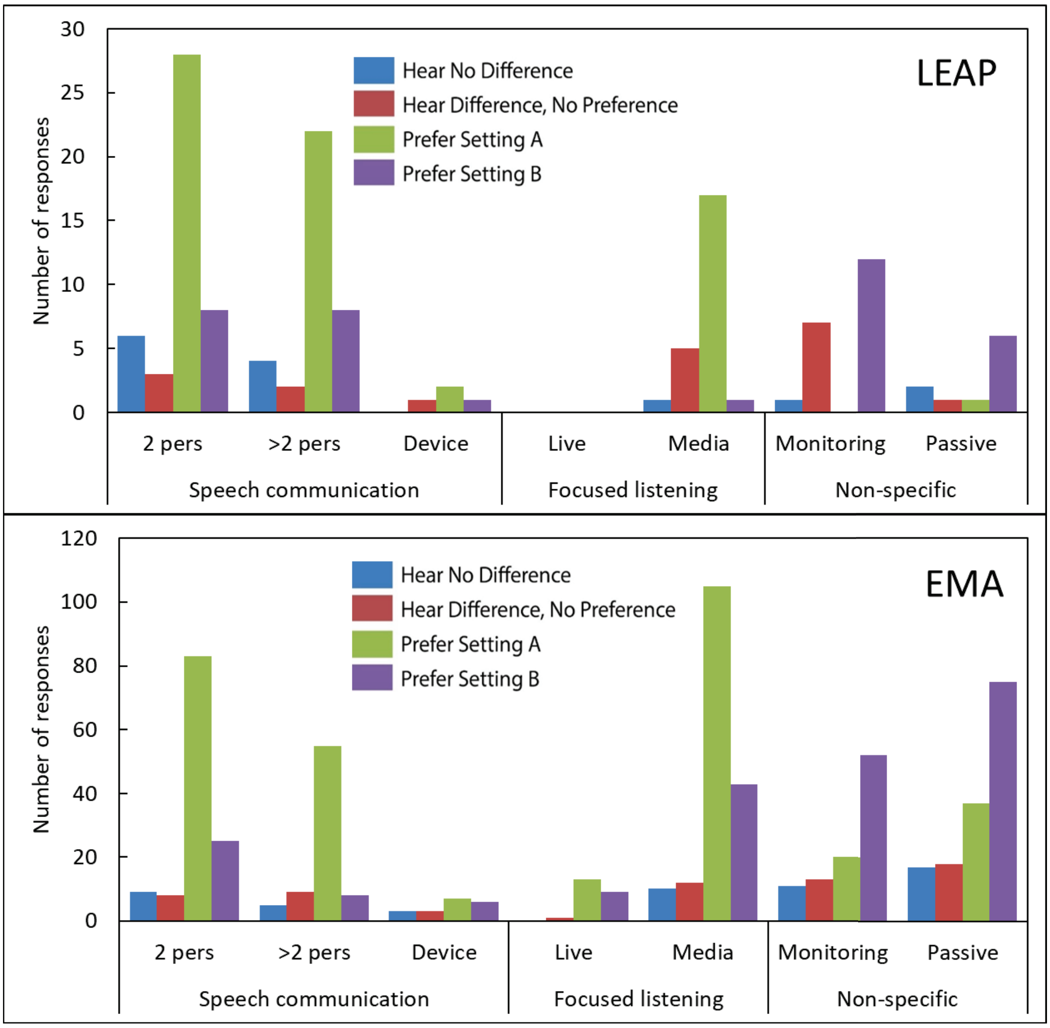

Figure 5. Preference distribution for LEAP data (upper panel) and EMA data (lower panel). The number of preferences for the response alternatives is shown for the seven CoSS task categories.

The preference picture becomes more nuanced when the entries are divided into the 7 CoSS task categories (Figure 5). Setting A was generally preferred in speech communication situations and in situations with focused listening, whereas Setting B was generally preferred in the non-specific situations, including “monitoring surroundings” and “passive listening”. This pattern is similar for the LEAP data (Figure 5, upper panel) and the field EMA data (Figure 5, lower panel). When the distributions over CoSS task categories are compared between the panels, it can be seen that Speech communication scenarios are over-represented in LEAP compared to the field, whereas Monitoring, and Passive listening are under-represented.

8 of the 10 participants had the same overall preferred hearing-aid setting for both LEAP and EMA, when the proportion of preferences were calculated individually. After the field-trial period, the participants were asked which setting they would keep if only allowed to just keep one. Eight participants selected Setting A, and two participants selected Setting B.

In the final interview, all test participants showed a predominantly positive attitude towards the LEAP methodology. They described the test in positive terms, like “exciting”, “interesting” and “fun”, but some of the participants also used terms as “demanding”. All the participants thought that the implemented LEAP scenarios resembled their experienced everyday life situations, but half of them also criticized some aspects of some scenarios. Most of the criticism was related to the observation that some of the LEAP test scenarios were not as difficult as the corresponding real-life situations.

4. DISCUSSION

Overall, the LEAP method worked well on a group level where it produced results that were comparable to the results obtained in the field using EMA (Figure 4 and Figure 5). On an individual level, the two tests also produced similar results. The LEAP method has a potential to predict real-life hearing-aid preference better than “traditional” laboratory tests by:

- addressing common mandatory listening situations with a possibility to add individually selected common, important and challenging test scenarios
- focusing on intention, especially realistic communication aspects (own voice, visual cues, social pressure)
- using realistic speech levels and signal-to-noise ratios (tailored both to the background noise and to the listener)

In the current study, only a subset of CoSS scenarios was implemented, and speech communication scenarios were over-represented (compared to the auditory reality data presented in Figure 3). These conversation scenarios worked well, and the test participants could handle the slightly complex test situation where they kept a conversation going while making paired comparisons using the remote control and reporting the results on the smartphone. The clearer preference for Setting A in the LEAP data, compared to the EMA data, is congruent with the over-representation of speech communication situations in LEAP. For communication situations, Setting A was preferred both in the lab and in the field.

One difficulty with the test is to evaluate preference for situations with passive listening. When test participants are asked to compare hearing-device settings, the intention in a passive scenario automatically changes to active listening. The monitoring task used in the current study seemed to work well as a “non-specific” scenario.

The simple loudspeaker setup in an ordinary office room used in the current study worked well for the evaluated signal-processing difference. It would, however, not be adequate for the evaluation of other types of signal-processing schemes, such as directional microphones, where a more refined loudspeaker setup would be needed. An alternative to a more advanced laboratory setup could be to take test participants to various locations and perform the LEAP test in realistic environments. This has been tried in our research group and proved to work well.

With the simple loudspeaker setup, LEAP could potentially be used in a clinical setting. One round of five mandatory test scenarios lasted 20 minutes, which might be possible to incorporate in a clinic if a specific feature were to be evaluated.

In the current study, the test participants evaluated the prescription differences in the field first and then in the laboratory. The two hearing-aid settings were allocated to the two hearing-aid programs in a balanced way across test participants but, for practical reasons, they were individually fixed during the EMA and LEAP trials. This meant that the test participants had time to build up a preference during the field trial that they might have kept during the LEAP test. In future studies, it would be beneficial to randomize the hearing-aid settings between test rounds. It would also be interesting to see if the LEAP results can *predict* EMA results, not only replicate them.

In the current study, LEAP was used to evaluate preference using paired comparisons, but the method can also be used to evaluate other attributes (such as speech clarity or comfort) and using other test paradigms (such as ratings of sound quality).

5. CONCLUSIONS

Paired comparisons of preference were made in the laboratory using a new method called LEAP, and in the field using EMA.

1. The LEAP method was feasible to use. Test participants generally found the LEAP scenarios to correspond to situations experienced in the field. When individually chosen listening situations could not be implemented satisfactorily, it was mainly because of the simple loudspeaker setup or because a larger group of conversation partners would have been needed to set the stage.
2. The LEAP results corresponded reasonably well to the results obtained using EMA. The general pattern of preference was the same for the two methods, and 8 out of 10 test participants had the same overall preference using LEAP and EMA.

ACKNOWLEDGEMENTS

The authors want to thank Sarah Gotowiec for helpful comments on an earlier version of the manuscript, and the test participants for their efforts.

REFERENCES

1. van Buuren RA, Festen JM, Plomp R. Evaluation of a wide range of amplitude-frequency responses for the hearing impaired. Journal of Speech and Hearing Research. 1995;38:211-21.
2. Naylor G. Theoretical Issues of Validity in the Measurement of Aided Speech Reception Threshold in Noise for Comparing Nonlinear Hearing Aid Systems. Journal of the American Academy of Audiology. 2016;27(7):504-14.
3. Dahlquist M, Lutman ME, Wood S, Leijon A. Methodology for quantifying perceptual effects from noise suppression systems. Int J Audiol. 2005;44(12):721-32.
4. Amlani AM, Schafer EC. Application of paired-comparison methods to hearing aids. Trends Amplif. 2009;13(4):241-59.
5. 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. Journal of the American Academy of Audiology. 2016;27(7):527-40.
6. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol. 2008;4:1-32.
7. Wu YH, Stangl E, Zhang X, Bentler RA. Construct Validity of the Ecological Momentary Assessment in Audiology Research. Journal of the American Academy of Audiology. 2015;26(10):872-84.
8. Timmer BHB, Hickson L, Launer S. Ecological Momentary Assessment: Feasibility, Construct Validity, and Future Applications. Am J Audiol. 2017;26(3S):436-42.