forum acusticum 2023

METHODS FOR EVALUATING PERFORMANCE WITH HEARING AIDS: WHAT ABOUT THE ECOLOGICAL VALIDITY?

Karolina Smeds* **Florian Wolters** **Petra Herrlin**
ORCA Europe, WS Audiology, Stockholm, Sweden

ABSTRACT

Keidser et al. (2020) published a definition of ecological validity: "In hearing science, ecological validity refers to the degree to which research findings reflect real-life hearing-related function, activity, or participation." At ORCA Europe, we have for a long time focused on learning more about people's auditory reality, i.e., the variety of listening demands and environments they experience in everyday life. We have strived to apply our understanding of auditory reality when developing methods for testing people's performance with hearing aids. Here, we will present examples of test methods used at ORCA Europe. Research using traditional laboratory tests, laboratory tests with more realistic listening tasks, and ecological momentary assessments will be included. Using our auditory reality data, the test methods will be informally evaluated in terms of how likely it is that they produce ecologically valid results. The evaluations will be based on knowledge about realistic signal-to-noise ratios and the type of activities people perform in everyday listening situations.

Keywords: *ecological validity, auditory reality, hearing aids, test methods*

***Corresponding author**: karolina.smeds@orca-eu.info.

1. INTRODUCTION

In recent years, many researchers have focused on introducing more realism into their laboratory testing. In 2020, a definition of ecological validity was published: "In hearing science, ecological validity refers to the degree to which research findings reflect real-life hearing-related function, activity, or participation." [1].
In our research group, we have investigated people's auditory reality, i.e., the variety of listening demands and environments people experience in everyday life. Based on the results of our auditory reality research, we have developed test methods for evaluation of performances with hearing aids. We have strived to develop test methods that have the potential to produce results that are indicative of real-life performance. In this paper, we present our research and describe our journey towards the design of such tests. In the Discussion section, we have included an informal evaluation of how likely our methods are of producing ecologically valid results.

2. SELECTED RESEARCH

In this section, selected research will be presented briefly to illustrate how our thinking about ecological validity has developed based on research findings.

2.1 Noise reduction in hearing aids

We started thinking about auditory reality and ecological validity when we investigated various ways to illustrate the effect of noise reduction (NR) on hearing-aid output [2]. For twelve hearing aids from various manufacturers, the long-term average gain reduction due to NR was determined. Real speech in speech-shaped noise was presented in an acoustic test chamber and coupler gain was measured with the NR algorithms turned on and off. The long-term average gain reduction varied substantially

among the hearing aids (Fig.1). It became obvious that the NR strategies used by hearing-aid manufacturers varied. For some of the hearing aids, like hearing aid G, the long-term average gain reduction was minimal, whereas the gain reduction for hearing aids B, C, and I was substantial. For hearing aids B and I, the gain reduction was substantial already for positive signal-to-noise rations, whereas for hearing aid C, the gain reduction was large only at negative signal-to-noise ratios.

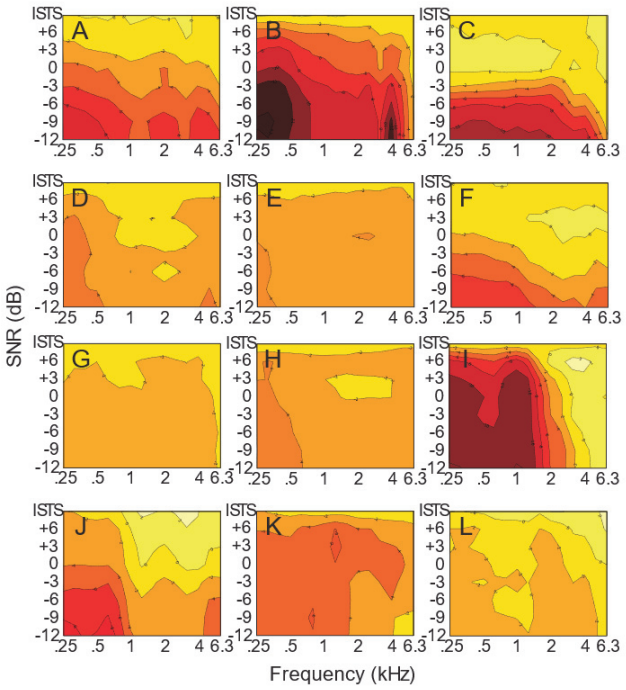

Figure 1. Each panel shows the measured long-term average gain reduction for one hearing aid for speech presented at 75 dB SPL. On the horizontal axis, frequency is represented from 250 to 6 000 Hz. The gain reduction measurements were made with one pure speech situation (no noise), represented at the top of each panel, and at seven signal-to-noise ratios (SNRs) ranging from +6 to -12 dB in 3-dB increments, represented on the vertical axis. The darker the color, the larger the gain reduction [2].

In a subsequent study, we investigated predictive measures of speech recognition after noise reduction processing [3]. Speech reception thresholds (SRTs) were determined as the signal-to-noise ratio (SNR) where the participants performed at 80% correct. A Swedish speech material with a fixed syntax (a "matrix test") was presented in babble noise. Testing was done for unprocessed materials and for material processed using three generic NR algorithms. A group of listeners with hearing impairment (HI) and a group with normal hearing (NH) participated. The results are displayed in Fig. 2 and reveal some potential issues. There is a large difference in results between the two groups of participants. The NH group performed at around 7 dB lower SNRs than the HI group. Further, there was a large difference in performance for individuals in the HI group. For the NR algorithm called WEDM, there was a difference of 10 dB between the best and the worst results. Although these issues did not create a large problem for the reported study, in general, this type of testing is problematic. If NR algorithms implemented in hearing aids had been used, test participants would have been tested using completely different NR settings, and it would be very difficult to understand the effect of the NR algorithms.

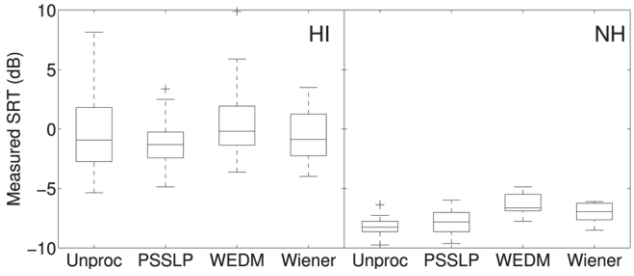

Figure 2. Speech reception thresholds (SRTs) for listeners with impaired hearing (HI, left) and normal hearing (NH, right) for three generic NR algorithms and one unprocessed version of the speech and noise. The lower the measured SRT, the better the result [3].

These two results in combination led to new thoughts: Comprehending speech in background noise is often described as one of the most difficult listening situations [e.g., 4]. What are the SNRs in these situations? Preferably, evaluation of NR algorithms should be made at realistic SNRs.

2.2 Realistic Signal-to-Noise Ratios

As a next step, we investigated realistic SNRs in hearing-aid users' everyday life [5]. Based on recordings made by twenty experienced and satisfied hearing-aid users [6], SNRs were estimated after identifying speech-

in-noise segments and noise-only segments of similar characteristics. Power calculations for these two types of segments were the basis of the SNR estimations. The recorded listening situations were grouped based on the type of background noise and the results for A-weighted levels for the better ear are seen in Fig. 3.

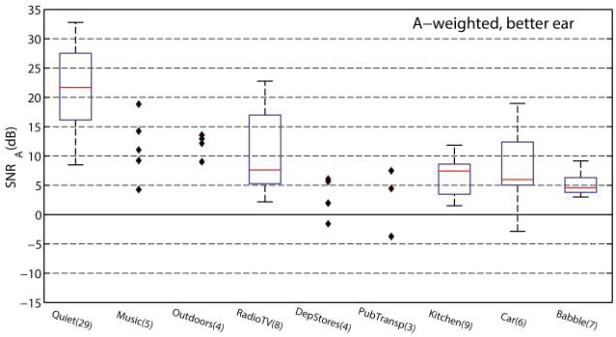

Figure 3. Estimated SNRs calculated based on A-weighted speech and noise levels for the better ear for nine different background noise types. The number in brackets after each category name gives the number of recordings in that category. For background noise types with fewer than six available recordings, raw data are plotted instead of box plots [5].

The range of SNRs was large and it was noticeable that there were very few recorded situations where the SNR was negative or even close to 0 dB. The lowest SNR was found for speech in babble where the median SNR was approximately 5 dB. These results have been replicated by Wu et al. [7].

The median SNRs were similar for many background noise types, even though people reported that the situations differed in difficulty [6]. There are certainly other factors than the SNR that affect the experience of a listening situation.

2.3 Common Sound Scenarios (CoSS)

With the aim to test hearing aids in ways that are representative of everyday listening, we set out to learn more about everyday listening. The work was inspired by a framework presented in a soundscape standard [8], where the *context* of a soundscape was central to its definition. Based on a literature review, a framework for Common Sound Scenarios (CoSS) was developed [9], focusing on hearing-related *intentions* and *tasks* in various situations (Fig. 4). For each task category, two example scenarios were presented, and each example was described in terms of the occurrence, the difficulty to hear, and the importance of hearing well. Both the examples and the descriptions were based on findings in the reviewed literature.

2.4 Investigating Auditory Reality using CoSS

We have used the CoSS framework in several studies where we have investigated people's auditory reality using Ecological Momentary Assessments (EMA). In one study [10], 19 experienced hearing-aid users were equipped with a smartphone-based EMA solution and prompted to respond to a survey seven times a day. When they started an EMA survey, they described their current listening situation and categorized it into one of the seven CoSS task categories. They also described background noise (if present) and any associated annoyance and then rated the difficulty to hear, the importance of hearing well, and how frequently the situation occurred in their everyday lives.

The collected data (Fig. 5) showed that speech communication situations amounted to roughly one-third of the reported situations, focused listening to almost one-quarter, and passive listening to almost half of the reported situations. Further, more than three-quarters of the situations were judged to occur in no noise or in noise that was not annoying at all.

When studying the situations that were judged to be very important to hear well in, the proportion of communication situations increased to more than half of the reported situations, but when looking at the situations that were judged to be both very important and occurring daily, focused listening to media increased to half of the reported situations. It was obvious that focused listening to TV (in particular), was a very common activity where it was important to hear well.

Our data show fewer communication situations and less difficult situations (especially due to noise) than people normally expect. When presenting this material, we have been asked if the somewhat "limited" or "easy" auditory reality, that we have found, might be explained by the fact that only hearing-aid users have participated. Perhaps these test participants have changed their lifestyle to avoid difficult situations. In an ongoing study, we are investigating the auditory reality of older and younger people with normal hearing in addition to older people with hearing impairment. Hopefully that study will shed light on possible avoidance patterns. Furthermore, we are planning auditory reality studies in other countries to investigate regional differences in people's auditory reality.

Intention	Speech communication						Focused listening				Non-specific			
Task	2 people		More than 2 people		Through device		Live sounds		Through media device		Monitoring surroundings		Passive listening	
	Two people having a conversation		Several people having a shared conversation		Two or more people having a shared conversation through a communication device		Focused listening to sound without being able to control the sound source		Focused listening to sound while being able to control the sound source		Conscious or unconscious screening of sound of relevance to current activity		Unconscious perception of environmental sounds, without relevance to current activity	
Scenario	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14
Occurrence														
Difficulty														
Importance														
Scenario	Conversation at home	Conversation on metro	Meeting in an office	Car ride with family	Phone call at home	Mobile call in the street	Lecture	At a concert	Watching TV	Listening to car radio	Vacuum cleaning	City walk	Relaxing with a book	Relaxing on train

Figure 4. The Common Sound Scenarios framework. Three intention categories (Speech communication, Focused listening, and Non-specific) were found. These were further subdivided into seven task categories. For each task category, two example scenarios are presented and occurrence, difficulty to hear and importance of hearing well are indicated (the darker the color, the higher the occurrence, difficulty and importance) .

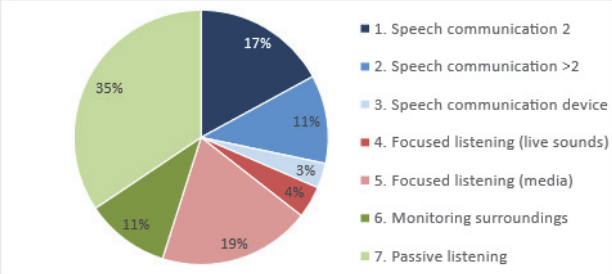

Figure 5. The distribution of responses over the seven CoSS task categories.

2.5 Live Evaluation of Auditory Preference

Based on what we had learned about people's auditory reality, specifically listening intentions and tasks, we realized that most laboratory testing belong to the Focused Listening CoSS category. There is usually a combination of listening to recorded material (Focused listening to media) but without being able to control the sound source (typical for Focused listening to live sounds). To broaden laboratory testing, we decided to create a laboratory test in which test scenarios were selected and designed based on the CoSS framework and which would cover also Speech communication situations and Passive listening. The LEAP test (Live Evaluation of Auditory Preference) was developed.

In one LEAP study [11], 19 older hearing-aid users participated. Six mandatory test scenarios, selected based on the CoSS framework, were implemented in an ordinary office room. The test scenarios included 1) communication with two people in quiet (the participant and a test leader were seated across each other at a table, Fig. 6), 2) communication with two people in car noise (the test leader moved to sit next to the participant and car noise was played from two loudspeakers), 3) communication between three people in cafeteria noise (another test leader joined; one test leader was seated across the table from the participant and the other was seated next to the participant), 4) TV (computer screen and loudspeakers placed about 2.3 m from the participant; documentary with a narrator and occasional background music), 5) Radio (jazz trio with female vocals), and 6) passive listening (participant and test leader sorting papers together). Conversations (in test scenarios 1-3) were sparked using photos. While the participants were in the six scenarios, their task was to

switch between two hearing-aid settings (using a paired-comparison paradigm) and decide which one they preferred.

Figure 6. An example of the LEAP setup for scenario 1.

Our main question was if it would be possible to include more realistic tasks in laboratory testing. We were particularly satisfied to see how well it worked with real conversations (while doing paired comparisons). The SNRs in the experiment were naturally tailored to the background noise levels (which were selected based on our previous work) and ended up close to the SNRs found in our previous study. The most difficult scenario to implement was the passive listening.

The preferred hearing-aid setting was also investigated in an EMA study, where the same participants compared the same hearing-aid settings in their everyday life for about a week. The results of the laboratory test and the field test were qualitatively very similar.

3. DISCUSSION

Below, we will use our reported auditory reality studies (the investigation of realistic SNRs, the development of the CoSS framework, and the subsequent use of the CoSS framework when studying auditory reality using EMA) for an informal evaluation of how likely our reported tests are to produce results that are indicative of performance in everyday life. It is important to remember that the test methods always aim at evaluating performance with hearing aids.

When evaluating generic NR algorithms (Fig. 2), we used a traditional speech test with a closed-set vocabulary and with a fixed sentence structure. For the participants with normal hearing, the SRTs needed for 80% correctly identified words were low, around -7 dB. Our data on realistic SNRs (Fig. 3) indicate that these SNRs are not common in everyday life. Although the SRTs for participants with impaired hearing were higher, the SNRs were still lower than for a typical speech-in-babble situation. Further, participants with normal and impaired hearing did not really evaluate the same systems. The same is true for the individuals in the group of participants with impaired hearing, where the spread in the resulting SRTs can be expected to be as large as in Fig. 2. We can quite safely say that this type of adaptive speech testing is not appropriate when evaluating for instance hearing-aid features whose performance depends on the SNR. The alternative would be to use a fixed realistic SNR. However, the drawback with using a fixed SNR is that it might be difficult to find a speech material that does not lead to floor and/or ceiling effects in the results, especially if both listeners with impaired and normal hearing are participating.

Further, the task performed in a traditional speech test is only relevant for focused listening, since central aspects of communications are missing. For instance, in real conversations there is a need to plan what to say while listening and turn-taking timing is important. Further, there is social pressure to respond and contribute adequately. But, in real conversations it is also possible to ask questions and ask for repetitions, making it less central to understand every word.

The LEAP test incorporates many of the aspects of everyday listening. Communication situations were included, and these worked well with the conversation sparker used. Scenarios with focused listening to TV and radio were easy to implement, whereas passive listening scenarios were more difficult. It is an oxymoron to evaluate a passive listening scenario. As soon as you start to listen carefully for the evaluation, it is no longer passive listening. We have previously tried other tasks for the non-specific CoSS intention category. A monitoring task in noise was created by letting the test participants vacuum clean the floor where some small pasta had been sprinkled. A purely passive listening task with reading has also been tried. Here, more work is needed to decide on a suitable task and a relevant outcome measure.

Although the LEAP test was generally successful in broadening the test tasks (especially by incorporating real conversations) and in using realistic SNRs (tailored both to the background noise and potentially also to a test participant's hearing ability), a very basic

loudspeaker setup was used (Fig. 6). For the hearing-aid characteristics investigated in the reported study, this did not constitute a problem, but such a loudspeaker setup could not be used for instance if evaluating a directional microphone or other features that require a more realistic sound field.

With EMA it is possible to move testing to participants' everyday life. Using EMA, the sound field and the activities are realistic. However, by asking participants to for instance perform paired comparisons of preference for two hearing-aid settings, the testing is interrupting the activity the participant was already doing. Despite the limitations in our experiments, we were encouraged by the fact that the LEAP test and the EMA study gave very similar results regarding the preferred hearing-aid settings. Although we cannot say that we have confirmed high ecological validity, the two tests validated each other to some degree.

4. CONCLUSION

This discussion shows the difficulties of developing test methods and designing studies that provide results that are indicative of everyday performance. Also, the difficulties of evaluating ecological validity are illustrated. Another paper in these proceedings (Akeroyd et al., 2023) suggests a table that can potentially be used for evaluation of ecological validity.

5. ACKNOWLEDGEMENTS

The authors want to thank Josefina Larsson and other co-authors of previous papers for their work to understand auditory reality and how this knowledge can be incorporated in testing the performance with hearing aids.

6. 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*, 41 Suppl 1: p. 5S-19S, 2020.

[2] K. Smeds, et al. "Noise reduction in modern hearing aids – Long-term average gain measurements using speech". in *The International Symposium on Auditory and Audiological Research (ISAAR). Binaural processing and spatial hearing*. Helsingør, Denmark, 2009.

[3] K. Smeds, et al., "Comparison of predictive measures of speech recognition after noise reduction processing". *J Acoust Soc Am*, 136(3): p. 1363-1374, 2014.

[4] S. Kochkin, "MarkeTrak VIII: Consumer satisfaction with hearing aids is slowly increasing." *Hear J,* 63(1): p. 19-32, 2010.

[5] K. Smeds, F. Wolters, and M. Rung, "Estimation of signal-to-noise ratios in realistic sound scenarios." *J Am Acad Audiol*, 26(2): p. 183-96, 2015.

[6] K.C. Wagener, M. Hansen, and C. Ludvigsen, "Recording and classification of the acoustic environment of hearing aid users". *J Am Acad Audiol*, 19(4): p. 348-70, 2008.

[7] Y.H. Wu, et al., "Characteristics of real-world signal to noise ratios and speech listening situations of older adults with mild to moderate hearing loss." *Ear Hear*, 39(2): p. 293-304, 2018.

[8] ISO/DIS 12913-1. "Acoustics. Soundscape–part 1: Definition and conceptual framework." International Standard Organization Geneva, Switzerland, 2014.

[9] F. Wolters, et al., "Common Sound Scenarios: A context-driven categorization of everyday sound environments for application in hearing-device research". *J Am Acad Audiol*, 2016. **27**(7): p. 527-40.

[10] K. Smeds, et al., "Selecting scenarios for hearing-related laboratory testing". *Ear Hear,* 41 Suppl 1: p. 20S-30S, 2020.

[11] K. Smeds, et al., "Live Evaluation of Auditory Preference, a Laboratory Test for Evaluating Auditory Preference." *J Am Acad Audiol*, 32(8): p. 487-500, 2021.