OPEN ACCESS

Short Report

Evaluation of a mobile research platform for self-administered hearing aid benefit assessment in the field

Evaluation einer mobilen Forschungsplattform zur eigenständigen Ermittlung von Hörgerätenutzen im Feld

Abstract

Hearing aid benefit is typically assessed in highly controlled laboratory environments. While this approach leads to reliable and reproducible results, it lacks representation of real-world characteristics and daily-life challenges. In this study, a portable setup for systematic benefit assessment outside of the laboratory is introduced and evaluated. It consists of a software framework for self-administered hearing tests and an integrated hearing aid signal processing platform. Within this study, both audiological measurements and subjective assessments in real-world environments are made with this setup to evaluate its ability to capture a hearing aid benefit and differences between distinct hearing aid settings. Preliminary data show promising results as amplified settings on average perform better than a non-amplified setting with few exceptions.

Keywords: hearing aid benefit, field study, subjective assessments, self-adjustment

Zusammenfassung

Der Vorteil einer Hörgeräteversorgung wird typischerweise in kontrollierten Laborumgebungen ermittelt. Einerseits führt dieser Ansatz zu zuverlässigen und reproduzierbaren Ergebnissen, andererseits ist er nicht in der Lage, Probleme des echten Lebens widerzuspiegeln. In dieser Studie wird eine mobile Forschungsplattform vorgestellt, mit welcher ein Hörgerätevorteil systematisch außerhalb von Laborumgebungen erfasst werden kann. Diese besteht aus der Kombination einer Software für eigenständig durchführbare Hörtests und einer Hörgeräte-Signalverarbeitungsplattform. Es werden sowohl audiologische Messungen als auch subjektive Bewertungen im echten Leben durchgeführt, um die Plattform hinsichtlich ihrer Fähigkeit, einen Hörgerätevorteil zu erfassen, zu evaluieren. Vorläufige Daten sind vielversprechend, da die Ergebnisse für Messungen ohne Verstärkung im Mittel schlechter ausfallen als Messergebnisse mit aktiver Verstärkung mit lediglich vereinzelten Ausnahmen.

Kim M. Rullmann[1,2]
Theresa Jansen[2,3]
Volker Hohmann[1,2,3]
Hendrik Kayser[1,2,3]

1 Carl von Ossietzky University Oldenburg, Oldenburg, Germany

2 Cluster of Excellence "Hearing4all.connects", Oldenburg, Germany

3 Hörzentrum Oldenburg gGmbH, Oldenburg, Germany

Introduction

Hearing aids are the most common intervention for hearing loss and are deemed beneficial. Determining that benefit, however, is not a straightforward process and has to consider multiple aspects. Typically, hearing aid benefit is quantified based on highly controlled audiological tests in laboratory environments. While this leads to reproducible results, it lacks incorporation of subjective perceptions as experienced in everyday life. Within this project, a mobile research platform is introduced and evaluated in regard to its ability to assess hearing aid benefit outside of the laboratory. This platform consists of the combination of two pre-existing frameworks: the *Virtual Hearing Clinic* (VHC, [1]) and the *open Master Hearing Aid* (openMHA, [2]) software. The VHC is a framework for self-administered hearing tests and features modules for the assessment of hearing thresholds [3], speech intelligibility [4], and loudness perception [5]. So far, it has been missing a signal processing unit to allow for aided measurements. To that end, openMHA was integrated into the VHC as it is precisely such a framework for hearing aid signal processing. The scope of this study is to evaluate the suitability of this combination for hearing aid benefit assessment in general and in particular in a real-world setting. Hence, the following research questions are addressed:

1. Is the combination of VHC and openMHA able to capture a hearing aid benefit using behavioural tests available in the VHC?
2. Does this combination of VHC and openMHA enable the assessment of perceived differences between amplification settings in real-world environments?

The following sections describe the setup in greater detail and outline the accompanying evaluation study. Preliminary data of six test subjects is shown and first observations are made.

Method

Setup

The main structure of the setup is based on the VHC. The VHC is implemented as a Python *Flask* application allowing access via any internet browser and easy handling via HTML pages. To allow for aided measurements, openMHA was interfaced with the VHC for processing of either prerecorded stimuli or live sound input via microphones. In this case, the combined framework of VHC and openMHA was installed on the *Portable Hearing Laboratory* (PHL) [6]. The PHL is a single-board computer extended with a multi-channel audio board running openMHA. It allows the attachment of *behind-the-ears* (BTE) hearing aid shells which include two microphones and a receiver-in-the-canal speaker. This setup allows the recording, processing, and playback of live sound input on portable hardware with low latency. Previous works on the VHC introduced modules for the assessment of the hearing threshold, speech intelligibility, and loudness perception. To allow for subjective in-situ evaluations of different hearing aid settings, an additional multiple comparison and rating tool that uses live sound input was implemented.
Within this study, a basic hearing aid configuration of openMHA is used to realise different amplification and compression settings. The main component is a six-channel dynamic range compressor with an attack time of 5 ms and a release time of 50 ms. Additionally, a frequency shifter is included to reduce acoustical feedback. Steps for equalisation and calibration of both input and output are included as well. In total, four compressor settings are used throughout this study. The first setting is the subjects' individual NAL-NL2 [7] prescription based on their audiogram with the goal to maximise speech understanding. Secondly, the loudness-based procedure *trueLOUDNESS* [8] is used which aims at normalising binaural broadband loudness perception. In addition to these prescription-based settings, a self-adjusted setting is used. To that end, another module was implemented in the VHC based on a self-adjustment procedure developed by Gößwein et al. [9]. During that procedure, the subject can manipulate the compressor settings in terms of overall gain and spectral tilt by moving a point within a two-dimensional user interface. The starting point of the procedure is the subject's individual NAL-NL2 prescription and they are instructed to find the setting they personally like best. A stimulus consisting of a spatially diffuse cafeteria background noise recording and sentences from the German Matrix Sentence Test [10] is played back during the adjustment. The SNR is kept constant at +5 dB and the procedure is executed three times at different levels: 65 dB SPL, 50 dB SPL, and 80 dB SPL. Afterwards, the three resulting gain curves are merged into a single one. Thus, there are three amplified settings (NAL-NL2, trueLOUDNESS, self-adjusted) and the fourth setting is a non-amplified setting for which openMHA is set to an acoustical transparency mode.

Procedure

Overall, the study is structured across three appointments for each participant. The first appointment mainly consists of laboratory reference measurements which serve as a gold standard reference for the VHC-based measurements. At first, an audiogram is measured alongside basic anamnesis to check if inclusion criteria have been fulfilled. The included participants should have a mild to moderate symmetric hearing loss that could be classified as N2 or N3 according to Bisgaard's standard audiograms [11], be experienced hearing aid users, have no cognitive deficits, and should be able to handle a hand-held tablet. Included participants' (2 male, 4 female) mean age is 65.3 years (38–76 years) and each gave informed consent to participate. The study has been approved by the ethics committee of the University of Oldenburg (Drs. EK/2021/031-10).
Using the *Oldenburg Measurements Applications*, speech intelligibility and loudness perception are measured using established laboratory hardware without any amplification. Speech intelligibility is measured adaptively using the German Matrix Test (Oldenburg sentence test, OLSA; [10]) in a diotic condition via headphones (Sennheiser HDA200). In total, three lists of 20 sentences are presented with the first two serving as training lists. Loudness perception is captured using the *Adaptive Categorical Loudness Scaling* (ACALOS, [12]) procedure and the same hardware as before. Additionally, the required loudness scalings for the trueLOUDNESS fitting procedure (see [8]) are performed. After completing the laboratory measurements, the audiogram is measured anew using the VHC-based setup in a calm office room without special acoustic treatment. The stimuli are played back via consumer headphones (Audio-Technica ATH-M50x). During this appointment, scans of the subjects' pinna are made for the production of individual ear moulds in-between appointments.
The second appointment exclusively takes place in an office room using the VHC-based setup. This time, the BTEs and individual ear moulds are used instead of headphones. At first, the self-adjustment procedure is performed. The individual NAL-NL2 and trueLOUDNESS

settings are determined in-between appointments and stored for each individual participant. Within this appointment, the speech intelligibility and loudness perception measurements are repeated using the VHC-based setup this time. Test configurations are kept equivalent to the laboratory measurements during the first appointment. This time, however, measurements with different hearing aid settings are made in a randomised order. For these tests, the prescription-based settings NAL-NL2 and trueLOUDNESS are used as well as the non-amplified setting twice. The VHC dashboards are implemented to guide the participants through the tests on their own and for the most part the participants are left on their own with help available next door on request.

The third and last appointment takes place in a shopping mall to obtain real-world assessments. Using the multiple comparison module, four different hearing aid settings (NAL-NL2, trueLOUDNESS, self-adjusted, non-amplified) are rated against each other at a bus stop with adjoining traffic, a relatively quiet shoe store, and a busy café in the middle of the mall. The multiple comparison module allows switching between the different settings in real-time in a double-blind manner. All settings are compared directly against each other while sliders can be used to rate the respective settings on a verbally annotated scale from zero to a hundred. Subjects are asked to rate the settings according to subjective speech understanding, listening effort, perceived loudness, and overall preference. Afterwards, the appointment is concluded with a short informal interview during which the participants have the opportunity to remark on aspects such as usability.

Results

Here, only preliminary results, that is 6 complete data sets, can be shown. Therefore, a formal statistical analysis is omitted and assessments are kept to qualitative observations. A publication with the full 20 data sets and appropriate analysis is in the works and will be submitted for publication at a later stage [13].

Audiological tests

Figure 1 shows the speech reception thresholds (SRT_{50}) for the five measurement conditions of the OLSA measurement, that is the laboratory reference condition and the four VHC conditions. While there is some spread in the data, it can be seen that the two amplified conditions (NAL-NL2, trueLOUDNESS) have lower median values of SRT_{50} than the three conditions without amplification (Laboratory, non-amplified, non-amplified repeated). This indicates a hearing aid benefit. So far, no clear difference between the two amplified settings can be seen although there is noticeably less spread in the trueLOUDNESS data. Additional data and analysis are needed to show if meaningful differences between the amplified and non-amplified settings can be found.

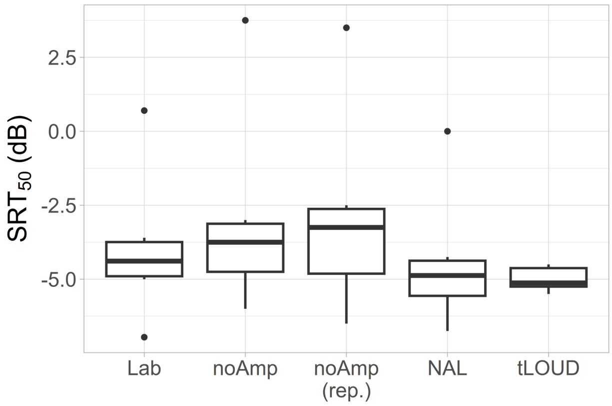

Figure 1: SRT_{50} values of OLSA speech intelligibility measurements for the five measurement conditions: laboratory unaided (Lab), VHC with non-amplified setting (noAmp), repetition of VHC with non-amplified setting (noAmp (rep.)), VHC with NAL-NL2 setting (NAL), VHC with trueLOUDNESS setting (tLOUD)

In Figure 2 the mean loudness functions for the five different measurement conditions are shown. These mean functions were derived by first fitting loudness functions to the individual data and then calculating the median level for each CU category across participants. These median values are then used to fit the average functions shown here. The mean functions corresponding to the two amplified conditions (NAL-NL2, trueLOUDNESS) are shifted noticeably to lower levels compared to the three non-amplified conditions. Thus, a hearing aid benefit is indicated but no clear difference between the two settings can be seen here.

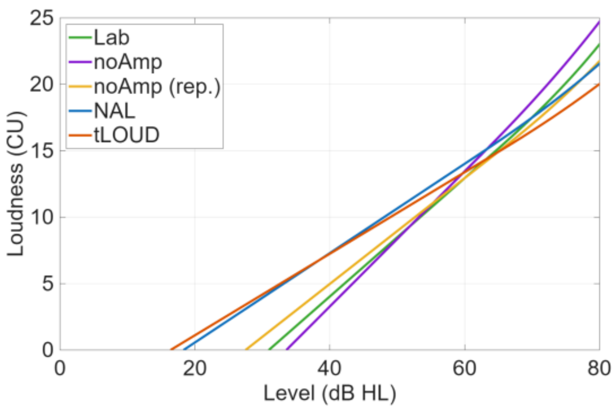

Figure 2: Average loudness functions derived from the ACALOS measurement for the five measurement conditions: laboratory unaided (Lab), VHC with non-amplified setting (noAmp), repetition of VHC with non-amplified setting (noAmp (rep.)), VHC with NAL-NL2 setting (NAL), VHC with trueLOUDNESS setting (tLOUD)

Subjective field data

In Figure 3 the subjective ratings for all four target questions (speech intelligibility, listening effort, loudness, preference) are shown separately for each environment (bus stop, café, shoe store). Since the meaning of the respective scale changes based on the question, red and green lines are plotted in addition to the data to indicate

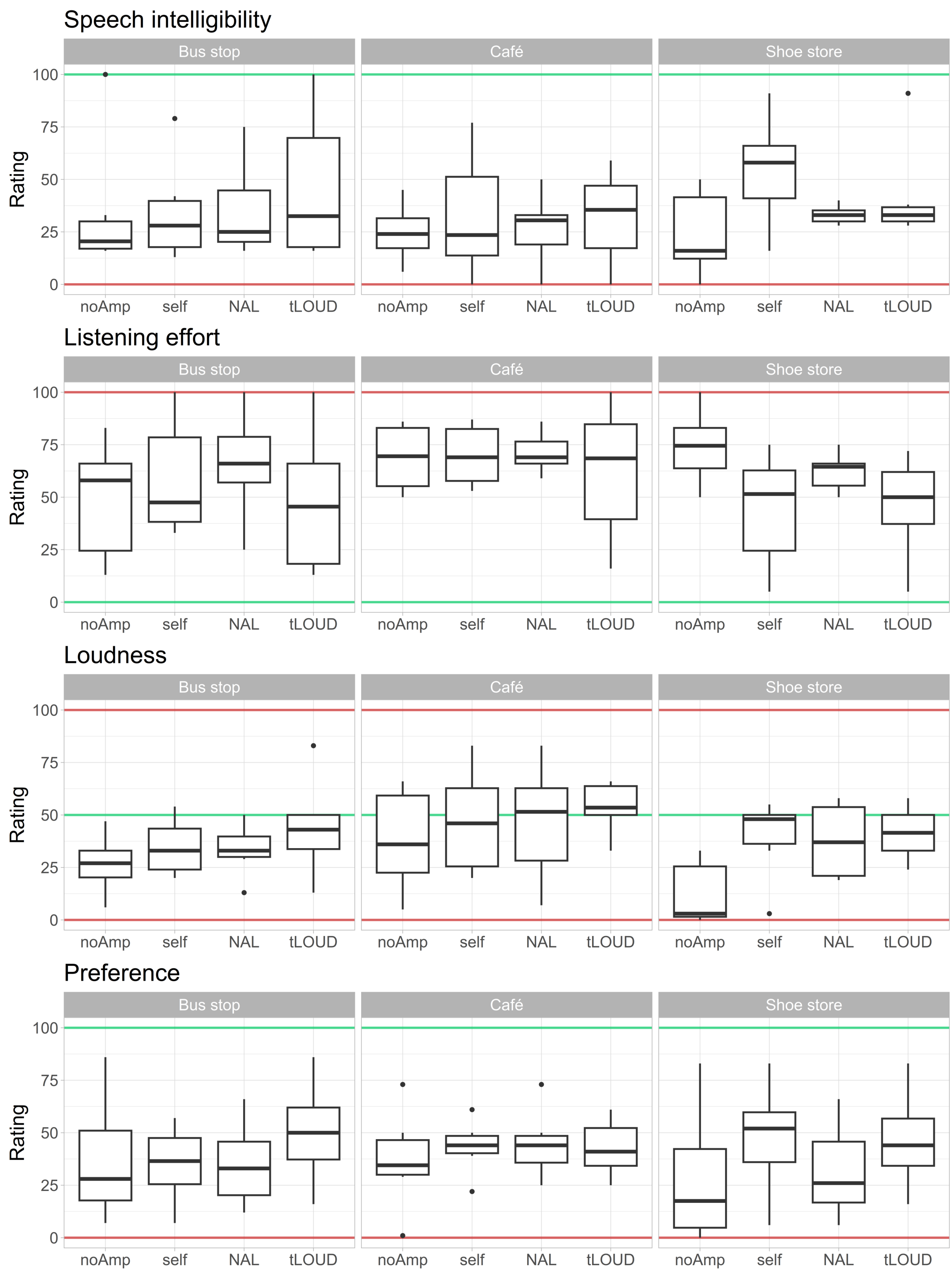

Figure 3: Subjective ratings for the four different target questions shown separately for each environment. The red and green lines represent the worst and best possible rating, respectively. The abscissa indicates the four different hearing aid settings under test: NAL-NL2 (NAL), non-amplified (noAmp), self-adjusted (self), and trueLOUDNESS (tLOUD).

the worst and best possible rating, respectively. Once more data is available, analysis regarding the effects of both hearing aid setting and environment is apposite. For now, it can be seen that for all questions the ratings in the café environment appear very similar across hearing aid settings. The clearest differences between settings can be seen in the shoe store environment. Out of the three environments, the shoe store is the least acoustically challenging which seemingly makes it easier for the subject to perceive differences between hearing aid settings. Nonetheless, in most environments the non-amplified setting is rated the worst, indicating a hearing aid benefit. Current exceptions include the listening effort rating at the bus stop where NAL-NL2 performs worst and the café environment where the ratings are overall too similar.

Conclusion

The current data suggests that the non-amplified setting on average performs worst, indicating a hearing aid benefit with only few exceptions. Additionally, some differences between the different hearing aid settings for both the audiological tests and the subjective assessments are observed. To make final statements, the complete data needs to be evaluated and analysed for statistical significance. The resulting data will be published at a later stage [13]. Nonetheless, the data indicate that the combination of the VHC with a research hearing aid holds the potential to be a suitable tool for investigating hearing aid benefit in the field.

Notes

Conference presentation

This contribution was presented at the 28th Annual Conference of the German Society of Audiology and published as an abstract [14].

Funding

This project was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2177/2 – Project ID 390895286.

Acknowledgements

The authors thank Chiara Haf and Kerstin Sommer for their assistance with data collection, and the participants for their valuable time.

Competing interests

The authors declare that they have no competing interests.

References

1. Schell-Majoor L, Rullmann KM, Bruns T, Jansen T, Hohmann V, Kayser H, Kollmeier B. The Virtual Hearing Clinic (VHC)- a modular online platform for hearing research and hearing health care [Preprint]. arXiv; 2026. DOI: 10.48550/arXiv.2608.05665
2. Kayser H, Herzke T, Maanen P, Zimmermann M, Grimm G, Hohmann V. Open community platform for hearing aid algorithm research: open Master Hearing Aid (openMHA). SoftwareX. 2022 Jan;17:100953. DOI: 10.1016/j.softx.2021.100953
3. Xu C, Hülsmeier D, Buhl M, Kollmeier B. How Does Inattention Influence the Robustness and Efficiency of Adaptive Procedures in the Context of Psychoacoustic Assessments via Smartphone? Trends Hear. 2024 Jan-Dec;28:23312165241288051. DOI: 10.1177/23312165241288051
4. Saak S, Kothe A, Buhl M, Kollmeier B. Comparison of user interfaces for measuring the matrix sentence test on a smartphone. Int J Audiol. 2025 Jul;64(7):745-57. DOI: 10.1080/14992027.2024.2385551
5. Xu C, Schell-Majoor L, Kollmeier B. Development and verification of non-supervised smartphone-based methods for assessing pure-tone thresholds and loudness perception. Int J Audiol. 2026 Jul;65(7):793-803. DOI: 10.1080/14992027.2024.2424876
6. Pavlovic C, Hohmann V, Kayser H, Wong L, Herzke T, Prakash SR, Maanen P. Open portable platform for hearing aid research. J Acoust Soc Am. 2018;143(3_Suppl):1738. DOI: 10.1121/1.5035670
7. Keidser G, Dillon H, Flax M, Ching T, Brewer S. The NAL-NL2 Prescription Procedure. Audiol Res. 2011 Mar 23;1(1):e24. DOI: 10.4081/audiores.2011.e24
8. Oetting D, Hohmann V, Appell JE, Kollmeier B, Ewert SD. Restoring Perceived Loudness for Listeners With Hearing Loss. Ear Hear. 2018 Jul/Aug;39(4):664-78. DOI: 10.1097/AUD.0000000000000521
9. Gößwein JA, Rennies J, Huber R, Bruns T, Hildebrandt A, Kollmeier B. Evaluation of a semi-supervised self-adjustment fine-tuning procedure for hearing aids. Int J Audiol. 2023 Feb;62(2):159-71. DOI: 10.1080/14992027.2022.2028022
10. Wagener K, Kühnel V, Kollmeier B. Entwicklung und Evaluation eines Satztests für die deutsche Sprache I: Design des Oldenburger Satztests. Zeitschrift für Audiologie. 1999;38(1):4-15.
11. Bisgaard N, Vlaming MS, Dahlquist M. Standard audiograms for the IEC 60118-15 measurement procedure. Trends Amplif. 2010 Jun;14(2):113-20. DOI: 10.1177/1084713810379609
12. Brand T, Hohmann V. An adaptive procedure for categorical loudness scaling. J Acoust Soc Am. 2002 Oct;112(4):1597-604. DOI: 10.1121/1.1502902
13. Rullmann KM, Jansen T, Hohmann V, Kayser H. Self-Administered Hearing Aid Fitting And Benefit Assessment In The Field Using A Research Hearing Aid. Forthcoming 2026.
14. Rullmann KM, Jansen T, Kayser H, Hohmann V. Evaluation einer mobilen Forschungsplattform zur eigenständigen Ermittlung von Hörgerätenutzen im Feld. In: Deutsche Gesellschaft für Audiologie e. V., editor. 28. Jahrestagung der Deutschen Gesellschaft für Audiologie. Oldenburg, 04.-06.03.2026. Düsseldorf: German Medical Science GMS Publishing House; 2026. Doc241. DOI: 10.3205/26dga241

Corresponding author:
Kim M. Rullmann
Carl von Ossietzky University Oldenburg, Ammerländer Heerstraße 114–118, 26129 Oldenburg, Germany
kim.rullmann@uni-oldenburg.de

Please cite as
Rullmann KM, Jansen T, Hohmann V, Kayser H. Evaluation of a mobile research platform for self- administered hearing aid benefit assessment in the field. GMS Z Audiol (Audiol Acoust). 2026;8:Doc19.
DOI: 10.3205/zaud000096, URN: urn:nbn:de:0183-zaud0000961

This article is freely available from
https://doi.org/10.3205/zaud000096

Published: *2026-09-01*