frontiers | Research Topics

Trends in digital hearing health and computational audiology

Edited by
Faheema Mahomed-Asmail, Karina De Sousa and Laura Coco

Published in
Frontiers in Audiology and Otology
Frontiers in Neuroscience

FRONTIERS EBOOK COPYRIGHT STATEMENT

The copyright in the text of individual articles in this ebook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this ebook is the property of Frontiers.

Each article within this ebook, and the ebook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this ebook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or ebook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-8325-5771-6
DOI 10.3389/978-2-8325-5771-6

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Trends in digital hearing health and computational audiology

Topic editors

Faheema Mahomed-Asmail — University of Pretoria, South Africa
Karina De Sousa — University of Pretoria, South Africa
Laura Coco — San Diego State University, United States

Citation

Mahomed-Asmail, F., De Sousa, K., Coco, L., eds. (2024). *Trends in digital hearing health and computational audiology*. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-8325-5771-6

Table of contents

frontiers | Frontiers in Neuroscience

TYPE Editorial
PUBLISHED 26 November 2024
DOI 10.3389/fnins.2024.1522600

Check for updates

OPEN ACCESS

EDITED AND REVIEWED BY
Michela Chiappalone,
University of Genoa, Italy

*CORRESPONDENCE
Faheema Mahomed-Asmail
✉ faheema.mahomed@up.ac.za

RECEIVED 04 November 2024
ACCEPTED 12 November 2024
PUBLISHED 26 November 2024

CITATION
Mahomed-Asmail F, De Sousa K and Coco L (2024) Editorial: Trends in digital hearing health and computational audiology. *Front. Neurosci.* 18:1522600. doi: 10.3389/fnins.2024.1522600

COPYRIGHT
© 2024 Mahomed-Asmail, De Sousa and Coco. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Editorial: Trends in digital hearing health and computational audiology

Faheema Mahomed-Asmail[1]*, Karina De Sousa[1] and Laura Coco[2]

[1]Department of Speech-Language Pathology and Audiology, University of Pretoria, Pretoria, South Africa, [2]School of Speech, Language and Hearing Sciences, San Diego State University, San Diego, CA, United States

KEYWORDS

digital hearing health, computational audiology, telehealth, AI in hearing care, e-audiology, digital transformation

Editorial on the Research Topic
Trends in digital hearing health and computational audiology

Introduction

Traditional hearing health care (HHC) service delivery models focus on face-to-face, clinic-based testing, often requiring several patient visits (World Health Organization, 2013). However, access to these services remains limited globally, leaving millions with untreated hearing loss, which has pervasive and profound consequences (Olusanya et al., 2014; Shukla et al., 2020). The shift toward mHealth and modern machine learning present opportunities to increase access in HHC through scalable models of care. This can be facilitated by low-cost hearing devices, smartphone technologies, and equipping a larger number of specialists for medical and surgical management of ear and hearing diseases (Bernstein et al., 2018). Furthermore, computational auditory models, advanced algorithms, and the use of artificial intelligence offer promising avenues for developing new hearing solutions and optimizing existing ones (Boisvert et al., 2023).

This Research Topic aimed to collect the latest research in these areas to support the effective implementation of digital technologies and computational methods in order to improve accessibility to ear and hearing healthcare services. The special edition consists of 11 articles and spanned over two Frontiers journals, Frontiers in Neuroscience, and Frontiers in Audiology and Otology. The Research Topic was initiated in June 2023, and opened for submission from September 2023 to October 2024, with a total of 14 submissions being received. The papers included in this edition are broad in their scope, ranging from validation of automated audiometry to machine learning and artificial intelligence.

Advances in audiometric assessment and hearing conservation methods

Automated audiometry has been proposed as an alternative of diagnostic assessment to improve access to hearing care by reducing time and costs, especially in areas with limited specialist availability. Liu et al. examined the correlation of air-conduction thresholds between automated audiometry conducted in a non-isolated environment and manual audiometry performed in a soundproof setting on individuals with normal hearing and

varying degrees of hearing loss. Consistent with previous research (Corry et al., 2017; Mahomed et al., 2013), Liu et al. found comparable results between the two methods across hearing levels.

Hearing conservation programs rely on serial audiograms to monitor shifts in hearing over time. McMillan et al. identified limitations in traditional approaches to serial monitoring and proposed a new statistical modeling method using a Gaussian process. This approach enables individualized predictions and simplifies interpretation, providing a less biased, more accessible tool for early detection of hearing changes.

Speech-in-noise testing

Human communication often occurs under adverse acoustical conditions, where speech signals mix with interfering speech or noise. Speech-in-noise (SIN) audiometry is thus a valuable part of audiological diagnostics and clinical measurements. Génin et al. developed and normalized a French speech-in-noise (SIN) test, SoNoise, to use as both a screening and a clinical evaluation tool. Normative values for diotic and antiphasic presentations were established with findings accurately capturing SIN abilities across various populations. Whereas, Meyer et al. investigated the use of a humanoid NAO robot to present target sentences alongside competing masker speech in a speech-in-speech test framework. Functional similarity was found in speech intelligibility when the NAO robot was compared to a traditional computer setup, with participants generally positive toward robot interactions.

Hearing aid technology and user experience

Hearing aids (HA) are prescribed to enhance communication and improve the quality of life for those who have hearing loss, but many individuals do not wear them consistently due to discomfort, dissatisfaction, or perceived lack of benefit, especially in noisy environments (Heselton et al., 2022). Alishbayli et al. developed a fast, domain-free noise suppression method, Statistical Sound Filtering (SSF), which used sound textures' statistical properties to enhance speech clarity. The evaluation of SSF demonstrated improvements in sound quality and reduced background noise levels without compromising speech intelligibility suggesting that SSF could be effectively integrated into HAs. While Fourie et al. examined the positive experiences of HA users through ecological momentary assessment (EMA) and found significant benefits in various contexts, particularly in conversational settings and leisure activities. Similarly, Sheng et al. investigated the perceived benefits of over-the-counter (OTC) hearing aids, a recently launched category of hearing devices, revealing that users experienced satisfaction scores comparable to traditional hearing aid users, along with notable improvements in emotional health, relationships, and communication abilities.

Linked to intervention options, Madahana et al. developed and tested a monitoring system that integrates a smartwatch and smart hearing muff with sound sensors in a mock mine environment. The system effectively detected noise levels and successfully communicated alerts to miners; however, further refinements and testing are required. Similarly, Andersson et al. leveraged EMA on heart rate data to understand the factors influencing real-world listening experiences. Results from a preliminary study among individuals with no hearing loss indicated that momentary heart rate data helped improve the prediction of self-reported listening experiences (passive vs. active listening). This study underscores the potential of integrating physiologic EMA data to deepen our understanding of listening dynamics in everyday environments and suggests promising applications for improving hearing aid outcomes among individuals with hearing loss.

Auditory training and assessment for pediatric populations

Spatial hearing is crucial for communicating in noise and can improve with training. Parmar et al. present a novel virtual reality (VR) game for an intervention designed to enhance spatial hearing in children and young people with bilateral cochlear implants. The BEARS (Both Ears) approach leverages the engaging, interactive, and immersive format of VR to strengthen listening skills, with the aim of supporting communication skills in noisy environments.

Auditory processing disorder (APD) assessments present challenges due to the disorder's heterogeneous nature, necessitating significant experience and training for accurate diagnosis. Wimalarathna et al. used a Random Forest model to analyse data from APD clinical test batteries to categorize children with APD into specific clinical subgroups which achieved 90% accuracy.

Conclusion

This Research Topic of articles highlights innovative solutions that can significantly enhance the accessibility and effectiveness of ear and hearing healthcare services, addressing the critical need for more inclusive approaches to managing hearing health across diverse populations.

Author contributions

FM-A: Conceptualization, Project administration, Writing – original draft, Writing – review & editing. KD: Writing – original draft, Writing – review & editing. LC: Writing – original draft, Writing – review & editing.

Acknowledgments

We thank the authors of the manuscripts submitted to this Research Topic for their valuable contributions and the referees for their rigorous review.

Conflict of interest

KD reported receiving scientific consulting fees from the hearX Group outside the submitted work.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Bernstein, L. E., Besser, J., Maidment, D. W., and Swanepoel, D. W. (2018). Innovation in the context of audiology and in the context of the internet. *Am. J. Audiol.* 27(3S), 376–384. doi: 10.1044/2018_AJA-IMIA3-18-0018

Boisvert, I., Dunn, A. G., Lundmark, E., Smith-Merry, J., Lipworth, W., Willink, A., et al. (2023). Disruptions to the hearing health sector. *Nat. Med.* 29, 19–21. doi: 10.1038/s41591-022-02086-6

Corry, M., Sanders, M., and Searchfield, G. D. (2017). The accuracy and reliability of an app-based audiometer using consumer headphones: pure tone audiometry in a normal hearing group. *Int. J. Audiol.* 56, 706–710. doi: 10.1080/14992027.2017.1321791

Heselton, T., Bennett, R. J., Manchaiah, V., and Swanepoel, W. (2022). Online reviews of hearing aid acquisition and use: a qualitative thematic analysis. *Am. J. Audiol.* 31, 284–298. doi: 10.1044/2021_AJA-21-00172

Mahomed, F., Swanepoel De, W., Eikelboom, R. H., and Soer, M. (2013). Validity of automated threshold audiometry: a systematic review and meta-analysis. *Ear Hear.* 34, 745–752. doi: 10.1097/01.aud.0000436255.53747.a4

Olusanya, B. O., Neumann, K. J., and Saunders, J. E. (2014). The global burden of disabling hearing impairment: a call to action. *Bull. World Health Organ.* 92, 367–373. doi: 10.2471/BLT.13.128728

Shukla, A., Harper, M., Pedersen, E., Goman, A., Suen, J. J., Price, C., et al. (2020). Hearing loss, loneliness, and social isolation: a systematic review. *Otolaryngol. Head Neck Surg.* 162, 622–633. doi: 10.1177/0194599820910377

World Health Organization (2013). *Multi-country assessment of national capacity to provide hearing care*. Available at: https://www.who.int/publications/i/item/9789241506571 (accessed October 30, 2024).

frontiers | Frontiers in Audiology and Otology

TYPE Original Research
PUBLISHED 21 July 2023
DOI 10.3389/fauot.2023.1215965

Check for updates

OPEN ACCESS

EDITED BY
Laura Coco,
San Diego State University, United States

REVIEWED BY
Marc Lammers,
Antwerp University Hospital, Belgium
Stela Maris Aguiar Lemos,
Universidade Federal de Minas Gerais, Brazil

*CORRESPONDENCE
Hasitha Wimalarathna
✉ hwimalar@uwo.ca

RECEIVED 02 May 2023
ACCEPTED 05 July 2023
PUBLISHED 21 July 2023

CITATION
Wimalarathna H, Ankmnal-Veeranna S, Duong M, Allan C, Agrawal SK, Allen P, Samarabandu J and Ladak HM (2023) Using machine learning to assist auditory processing evaluation. *Front. Audiol. Otol.* 1:1215965. doi: 10.3389/fauot.2023.1215965

COPYRIGHT
© 2023 Wimalarathna, Ankmnal-Veeranna, Duong, Allan, Agrawal, Allen, Samarabandu and Ladak. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Using machine learning to assist auditory processing evaluation

Hasitha Wimalarathna[1,2]*, Sangamanatha Ankmnal-Veeranna[2,3], Minh Duong[2,4], Chris Allan[2,4], Sumit K. Agrawal[1,2,5,6,7], Prudence Allen[2,4], Jagath Samarabandu[1,5] and Hanif M. Ladak[1,2,5,6,7]

[1]Department of Electrical and Computer Engineering, Western University, London, ON, Canada, [2]National Centre for Audiology, Western University, London, ON, Canada, [3]College of Nursing and Health Professions, School of Speech and Hearing Sciences, The University of Southern Mississippi, Hattiesburg, MS, United States, [4]School of Communication Sciences and Disorders, Western University, London, ON, Canada, [5]School of Biomedical Engineering, Western University, London, ON, Canada, [6]Department of Medical Biophysics, Western University, London, ON, Canada, [7]Department of Otolaryngology-Head and Neck Surgery, Western University, London, ON, Canada

Introduction: Approximately 0.2–5% of school-age children complain of listening difficulties in the absence of hearing loss. These children are often referred to an audiologist for an auditory processing disorder (APD) assessment. Adequate experience and training is necessary to arrive at an accurate diagnosis due to the heterogeneity of the disorder.

Objectives: The main goal of the study was to determine if machine learning (ML) can be used to analyze data from the APD clinical test battery to accurately categorize children with suspected APD into clinical sub-groups, similar to expert labels.

Methods: The study retrospectively collected data from 134 children referred for ADP assessment from 2015 to 2021. Labels were provided by expert audiologists for training ML models and derived features from clinical assessments. Two ensemble learning techniques, Random Forest (RF) and Xgboost, were employed, and Shapley Additive Explanations (SHAP) were used to understand the contribution of each derived feature on the model's prediction.

Results: The RF model was found to have higher accuracy (90%) than the Xgboost model for this dataset. The study found that features derived from behavioral tests performed better compared to physiological test features, as shown by the SHAP.

Conclusion: The study aimed to use machine learning (ML) algorithms to reduce subjectivity in audiological assessments used to diagnose APD in children and identify sub-groups in the clinical population for selective interventions.

Significance: The study suggests that this work may facilitate the future development of APD clinical diagnosis software.

KEYWORDS

auditory processing disorder, clinical data mining, audiology, hearing disorders, machine learning

1. Introduction

Auditory processing refers to how the brain interprets the sounds that one has heard. Normal auditory processing is important for understanding complex sounds, such as music or speech in difficult listening situations like classrooms, recreation, social gatherings, or restaurants. If the auditory system has weak processing skills, it can lead to listening problems [Cline, 2001; American Speech-Language-Hearing Association (ASHA), 2005]. Approximately 0.2–5% of normal-hearing children have difficulty understanding complex sounds, especially in difficult listening situations (Chermak et al., 1997; Nagao et al., 2016). These children are suspected of having Auditory Processing Disorder (APD). APD is usually identified by parents or teachers and requires an assessment by an audiologist for a formal diagnosis. APD assessments are typically carried out in specialized clinical centers. The audiologists who conduct these tests require extensive training and experience for proper assessment and diagnosis. However, there is a lack of consensus regarding which specific tests should be included in the APD assessment battery (Emanuel et al., 2011; Iliadou et al., 2017). Professional bodies including the American Speech-Language Hearing Association (ASHA) recommend using both behavioral and physiological measures (in a test battery approach) to assess auditory processing in children suspected of APD [American Speech-Language-Hearing Association (ASHA), 2005]. The behavioral component measures the child's ability to process acoustic stimuli (speech and non-speech) and respond verbally. The physiological component measures the overall integrity of the auditory system (Starr and Achor, 1975; Allen and Allan, 2014). A diagnosis of APD is made if the child's test scores are greater than two standard deviations from normative thresholds on two or more tests, or three standard deviations on one test [American Speech-Language-Hearing Association (ASHA), 2005].

There are typically very few referrals made to clinics for APD per year (Moore et al., 2018), making it difficult for training audiologists to gain sufficient practice assessing APD. The diagnosis of APD is also challenging due to its heterogeneity and associated comorbidities (Bamiou et al., 2001; Chermak, 2002; Sharma et al., 2009; Iliadou et al., 2017, 2018, 2019). As a result, there are very few studies on the management of APD children (Emanuel et al., 2011). Allen and Allan (2014) previously classified children with APD into clinical sub-groups based on how they performed on behavioral and physiological tests. Children who performed poorly on behavioral tests were considered behaviorally abnormal[1]; children who had atypical physiological findings were considered physiologically abnormal; children who performed poorly on both were considered abnormal across; and the children whose performance on both behavioral and physiological measures were within normal limits were categorized into a separate group. By identifying sub-groups of APD, an audiologist can better apply specific interventions the child may require. For example, children who have difficulty processing auditory information behaviorally may benefit from auditory training (Weihing et al., 2015), whereas children who show atypical physiologic processing may benefit from using frequency modulated (FM) systems (Hornickel et al., 2012; Rance et al., 2014). An FM system is a wireless device which reduces the background noise and improves sound clarity (Johnston et al., 2009). Children who have difficulty processing auditory information both behaviorally and physiologically may benefit from both auditory training and the use of an FM system [American Speech-Language-Hearing Association (ASHA), 1970; Sharma et al., 2012; Keith and Purdy, 2014; Smart et al., 2018]. Children whose performance is within normal limits on both behavioral and physiological measures may indicate to address non-auditory concerns, and a referral to another professional is required. Categorizing children into different subclinical groups is however complex, time consuming, and highly subjective.

Machine learning (ML) is becoming increasingly popular in the field of medicine to help clinicians make timely and accurate clinical diagnoses. ML techniques can be applied in designing software for clinical use by learning from the data (Davenport and Kalakota, 2019). Additionally, ML helps to reduce subjectivity in clinical judgment. Previously, ML models were considered "black box" models; however, with improvements in interpretability, models are now able to be better understood and applied in clinical settings (Ahmad et al., 2018). There is only one study in the literature that has used unsupervised ML techniques (hierarchical clustering) to identify sub-groups in APD data (Sharma et al., 2019). The study used data collected from 90 children aged 7–12.8 years old. Four sub-groups were found based on 10 variables, as follows:

- Group 1: Children with global deficits
- Group 2: Children with poor auditory processing, but good word reading and phonological awareness skills
- Group 3: Children with poor auditory processing, poor attention, and poor memory, but good language skills
- Group 4: Children with poor auditory processing and poor attention, but good memory skills

The assessments included in the analysis were behavioral tests and Cortical Evoked Auditory Responses (CEARs). However, APD is heterogenous, and it is therefore important to evaluate a variety of skills in the clinical assessment. There is a current lack of research using ML techniques to categorize APD data into sub-groups using both behavioral and physiological assessments.

The goal of our study was to determine if ML models can be used to learn and predict the diagnosis of APD with similar accuracy to clinical audiologists. Furthermore, we used interpretability techniques to identify how important each individual assessment within the APD battery is in arriving at an accurate label. The application of ML may show the diagnostic accuracy of APD, assist in centers where experts offer limited availability, and enable another tool together with clinical expertise to target individualized intervention of APD. To our best knowledge, this is the first study to:

1. Use supervised ML methods for APD data analysis from a comprehensive test battery that includes both behavioral and objective hearing assessments.

1 The word abnormal indicates the performance of the child in the auditory processing test battery fell at least two standard deviations below that of typically developing children. This is valid for any place the word "abnormal" is used in the paper.

2. Use interpretability techniques to identify which APD assessments contribute most to an accurate APD diagnosis based on expert labels.

2. Materials and methods

2.1. Dataset

Children with listening complaints (hearing in noise) and or poor academic performance were referred to the H.A. Leeper Speech & Hearing clinic at the University of Western Ontario, Canada for an assessment of APD. Data from 134 children between the ages of 5–17 years old (90 male; 44 female) were retrospectively collected from 2015 to 2021. The primary language of all the children was English. The Health Sciences Research Ethics Board of Western University, Canada, has approved the study (IRB 00000940).

2.2. Auditory processing audiological clinical test battery

The APD assessment is carried out in a test battery format following guidelines recommended by ASHA [American Speech-Language-Hearing Association (ASHA), 2005]. The test battery appraises the overall wellbeing of the auditory system, starting with how sound is processed and perceived by the auditory system. The Auditory Processing (AP) test battery consists of both behavioral and physiological measures. In behavioral tests, the processing and perception of auditory information are assessed. In physiological measures, the overall neuro-physiological wellbeing of the auditory system is assessed. Figure 1 shows a summary of the AP test battery.

First, patient demographics such as age, gender, birth history, middle ear history, family hearing issues, and additional health problems are completed, typically by a parent. Next is a detailed peripheral hearing assessment. In the peripheral hearing assessment, pure tone audiometry (the minimum intensity that a listener can detect for different test frequencies), tympanometry (an assessment of middle ear function), and otoacoustic emissions (a physiological measure that assesses the functioning of the outer hair cells) are completed to ensure that the child does not have any hearing loss. If a child fails any of the tests in the peripheral hearing assessment, auditory processing tests will not be administered [American Speech-Language-Hearing Association (ASHA), 2005]. Children with normal hearing as indicated by the peripheral hearing assessment will then undergo the auditory processing test battery. In this study, pure tone audiometry was conducted using the GSI-61 (Grason Stadler Inc, USA) Clinical Audiometer. The middle ear function was assessed using the GSI Audiostar (Grason Stadler Inc, USA) TympStar diagnostic middle ear analyzer. The otoacoustic emissions were measured through the Titan Suite.

The behavioral tests that are used are standardized and widely used in North America (Emanuel et al., 2011). Behavioral tests can be categorized into speech and non-speech tests. The Staggered Spondaic Word (SSW) test (Katz, 1998) is a dichotic listening

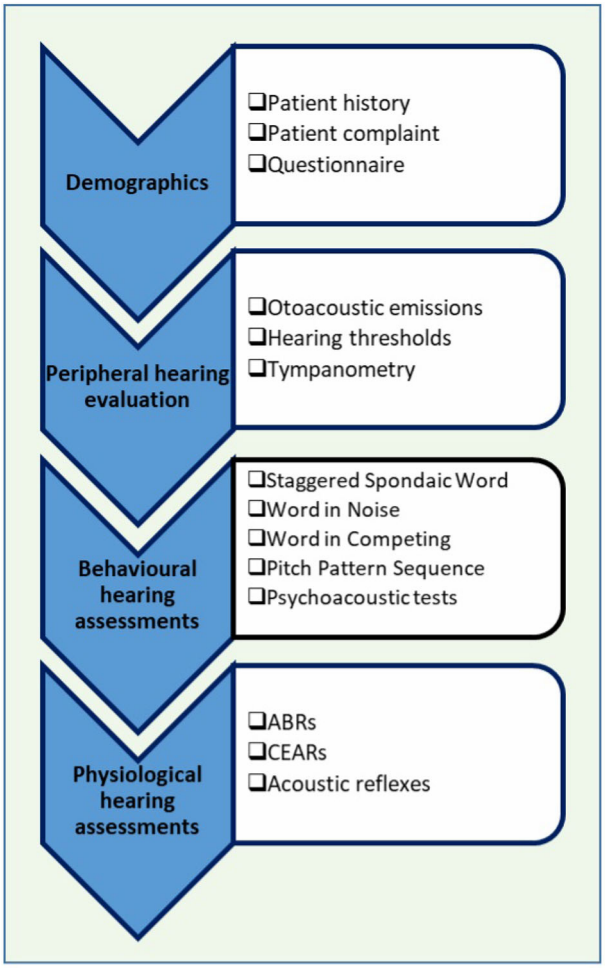


FIGURE 1
Auditory processing (AP) clinical test battery. ABR, auditory brainstem responses; CEARs, cortical evoked auditory responses.

test in which two spondees[2] are presented in a staggered fashion and the listener must repeat all four words. The Word in Noise (WIN) test (Wilson, 2003) assesses the individual's ability to listen to speech in noise. In WIN tests, words are presented in multi-talker babble at seven signal-to-noise ratios (SNR) (+24–0 dB). In the Word in Ipsilateral competing noise (WIC) test (Ivey, 1969), words are presented at +5 dB SNR. The Pitch Pattern Sequence (PPS; Pinheiro, 1977) test assesses the auditory system's ability to perceive and or process auditory stimuli in their order of occurrence. In this study, adaptive auditory discrimination tests (psychoacoustic tests), such as the ability to detect brief gaps in noise, amplitude modulation (20 and 200 Hz), and ability to discriminate frequency (1,000 Hz), were also performed for a portion of children. In the current study, speech behavioral assessments were conducted using the GSI-61 Clinical Audiometer and the psychoacoustic tests were carried out using the Tucker Davis System.

Once the behavioral tests are completed, physiologic assessments are carried out. The auditory brainstem responses

2 Spondees are terms that accommodate two equally stressed syllables.

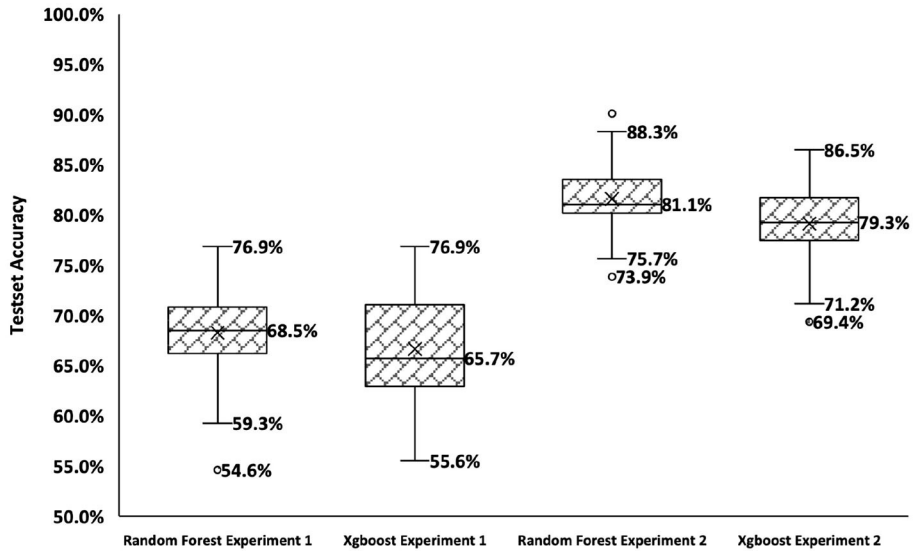


FIGURE 2
Summary of results for Experiments 1 and 2. The x-axis shows the four conditions, and the y-axis shows the accuracy over the test set.

(ABRs) and cortical evoked auditory responses (CEARs) are recorded from children. These evoked responses are recorded from both ears by placing a surface electrode on the scalp and presenting brief acoustic stimuli. The ABR was recorded by presenting a 100 μs rarefaction click stimulus at 80 dB nHL at a rate of 13.3 clicks/s. The CEARs were recording using a 60 ms tone stimuli at 1,000 Hz with an intensity of 70 dB nHL. The stimuli were presented monoaurally through ER-3A, Etymotic Research Inc insert earphones. The ABR occurs between 0 and 8 ms after stimulus onset, whereas the CEARs occur between 80 and 300 ms after stimulus onset. The recording windows were 10 and 750 ms for ABRs and CEARs, respectively. The responses were averaged and amplified with an amplification of 100 k for ABRs and 30 k for CEARs. Bandpass filters of 100–1,500 Hz were applied to ABRs and filters of 1–30 Hz were applied to CEARs. The artifact rejection was set to 23.8 μV for ABRs and 79.2 μV for CEARs. In the present study, the recording of CEARs took place at a separate appointment, and only limited data was available. To record these physiological signals, we used a Bio-logic Navigator Pro AEP system (Natus Medica, Inc).

The last test performed in the test battery is the acoustic reflex (middle ear muscle reflex) test. The acoustic reflexes are recorded by presenting loud tones to the ear. When loud tones are presented, the admittance of the tympanic membrane and middle ear system decreases due to stapedius muscle contraction. Presence of an acoustic reflex is an indication that the middle ear and the peripheral auditory system is intact. The GSI TympStar diagnostic middle ear analyzer was used to obtain the acoustic reflexes in the current study.

2.3. APD subgroups

A study conducted by Bellis and Ferre (1999) proposed the idea of determining different sub-groups of APD children. Previous studies conducted in the Child Hearing Research Laboratory at Western University (Allen and Allan, 2014) have also shown the importance of both physiological and behavioral assessments in the AP test battery thereby observing sub-groups in the APD data. A study by Sharma et al. (2019) used hierarchical cluster analysis to identify sub-groups in APD children. In the Sharma et al. (2019) study, data was collected from over 90 school-aged children (7–13 years old) who were suspected of having an APD. The collected data contained the outcomes of test results, which assessed the children's reading, language, cognition, and auditory processing. Initially, the dataset had 23 variables based on various auditory assessments, however, for the cluster analysis, only 10 variables were included, namely: phonological, irregular, TONI, Forward, Dichotic Digit Test, Language, Non-word, Attention, Backward DS, and Frequency Pattern Test. The cluster techniques used were hierarchical clustering, followed by k-means. Four clusters of children were identified: 35 children showed global deficits; 22 children showed poor auditory processing with good word reading and phonological awareness skills; 15 children had poor auditory processing with poor attention and memory, but good language skills; and 18 children had poor auditory processing and attention with good memory skills. However, the authors did not include any physiological data such as ABRs or otoacoustic emissions in the cluster analysis.

Cluster analysis techniques are unsupervised learning techniques, whereas in supervised learning, expert labels are used to train ML models. In supervised learning, after the model is trained with part of the labeled data (the "training" set), predictions are made on the other part of the data (the "test" set). The predicted results are compared to the labeled data to evaluate the accuracy of the model. The use of supervised or unsupervised techniques depends on whether human experts are available to provide the labels of the test set. Here, three expert audiologists with >10 years of experience assessing APD children labeled the dataset into four APD sub-groups (Allen and Allan, 2014) based on if children were

behaviorally and physiologically normal or abnormal. The four labels were presented as follows:

- "BnPn" = Behaviorally Normal and Physiologically Normal
- "BnPa" = Behaviorally Normal and Physiologically Abnormal
- "BaPn" = Behaviorally Abnormal and Physiologically Normal
- "BaPa" = Behaviorally Abnormal and Physiologically Abnormal

2.4. Feature engineering

In traditional ML algorithms, the data should be transformed to features that better represent the underlying problem to reach a satisfactory outcome. This process is called feature engineering. Deep Learning (DL), which is a sub-field in ML, does not require such manipulations; the model itself performs feature engineering. However, in medical applications, the use of DL techniques is limited due to the scarcity of data. Hence, traditional ML algorithms with effective feature engineering techniques may produce predictive models well-suited to the current problem. The feature engineering performed in our study was done with the advice of the domain experts. Since the AP assessments that are included in our study have standardized test scores, inserting the data as raw data into the ML pipeline seemed to interfere with the outcome of the tests. An additional problem was that some children did not finish all the assessments in the test battery for various reasons. Therefore, a better representation of data was needed to encode these clinical tests. Based on the expert agreement, the raw data encoding was carried out categorizing to "pass," "fail," "did not finish the assessment," and "missing data." One-hot encoding was conducted when feeding as features. The tests that were encoded in this manner were otoacoustic emissions, hearing thresholds, acoustic reflexes, and all the behavioral test results (both speech and non-speech tests).

The ABR and CEAR data are presented as clinical waveforms. To represent these data, we used the Continuous Wavelet Transform (CWT) as a feature extractor, as described in our previous work (Wimalarathna et al., 2021). The CWT is a time-frequency plot obtained by convolving a signal with a window function called a "mother wavelet." The mathematical equation for the wavelet transform is as follows (Torrence and Compo, 1998),

$$W(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} s(t)\psi^*\left(\frac{t-b}{a}\right)dt \tag{1}$$

In the equation, $s(t)$ represents the signal and the $\psi(t)$ represents the mother wavelet which is scaled by "a" and translated by "b". The CWT plot is obtained by convolving these scaled and translated versions of the mother wavelet. There are multiple mother wavelet types introduced in the literature, however since the ABRs and CEARs consist of peaks and valleys, intuitively the Gaussian mother wavelet was chosen throughout the study. If the ABR or the CEAR is windowed in the locations where peaks and valleys occur, it closely matches with the Gaussian wavelet. This was one major reason for selecting the Gaussian wavelet as it would best mimic how a clinician would select the peaks and valleys from a waveform.

In our previous study (Wimalarathna et al., 2021), the features extracted from the CWT were sent through a statistical feature selector and the models were trained to recognize abnormal versus normal ABR responses. In the previous study, 700 features were required to reach a 92% accuracy. For the present study, we wanted to further optimize the feature space and determine if a smaller number of features could represent the group differences between typically developing children and children suspected of having APD. The complex Gaussian mother-wavelet was used to compute the CWT plot for both ABRs and CEARs. The resulting CWT representation was a complex matrix. Therefore, to derive the features, we considered its magnitude and phase. From the magnitude plot of the CWT coefficient matrix, the coefficients were averaged across time and the coefficient of dispersion was calculated based on the following equation,

$$Coefficient\ of\ Dispersion = \frac{Q_3 - Q_1}{Q_3 + Q_1} \tag{2}$$

In the equation, Q_1 and Q_3 are the first and the third quartile of the average values, respectively. The angle of the CWT coefficients was first unwrapped across the time axis and the standard deviation was calculated as a feature. The ABRs and CEARs both contained four features in total, representing the magnitude and the phase. In a clinical setting, audiologists are interested in peaks, inter-peak intervals and their timing. However, there is additive subjectivity when an inexperienced audiologist analyzes the waveforms. Hence, by automatically calculating these features, subjectiveness in the analysis can be mitigated. The designed features represent similar characteristics of the waveform that clinicians derive manually. The ability of these features to explain the group differences were tested using ML models. Feature interaction was studied by adding and removing the features while observing the effects on accuracy.

2.5. Data augmentation techniques

APD is a rare disorder and clinics typically receive few APD referrals per year. It has also been reported that obtaining a referral for APD diagnosis is difficult (Moore et al., 2018; Agrawal et al., 2021). For these reasons, there was a limited amount of data available for this study and it took approximately six years to collect the data within the dataset. Data augmentation techniques may be used to overcome the difficulties associated with training ML models with small datasets. Several techniques have been identified in the literature, with resampling techniques being the most commonly used. Synthetic Minority Over-sampling TEchnique (SMOTE) is one such resampling technique where synthetic samples are generated for minority data instances (Chawla et al., 2002). The technique draws a new sample at a position (feature space) between samples. First, the algorithm selects a random instance from the minority class. Next, k nearest neighbors for that example are located. A synthetic example is then generated at a randomly chosen position in the feature space between the two instances and their randomly chosen neighbor (Brownlee, 2020a).

Recent advances in ML have led to the development of more sophisticated techniques for data augmentation, such as Generative Adversarial Networks (GANs). However, there are challenges in using this model for augmenting tabular data such as mixed

data types, continuous features having multimodal non-gaussian distributions, and highly imbalanced categorical columns. The Conditional Tabular GAN (CTGAN) model designed by Xu et al. (2019) has been able overcome these challenges and has been proven to perform better than the existing architectures. The model uses mode-specific normalizations to overcome the issue of non-Gaussian and multimodal distributions. Additionally, the training-by-sampling technique is included to solve the problem of imbalanced columns. In the present study, we utilized CTGAN and the SMOTE resampling technique separately to compare which data augmentation technique was best suited to our application.

2.6. Machine learning algorithms

Ensemble learning techniques are generally considered suitable to train with small amounts of data. These models aggregate the outcome of a large number of models to produce a single classifier (Breiman, 1996). Bagging (Breiman, 1996) and Boosting (Schapire, 1990; Freund and Schapire, 1996) are two popular techniques used in building accurate ensemble models. In Bagging, the ensemble classifier combines the output of various learned classifiers into a single classifier. Boosting technique iteratively invoke a weakly learned classifier producing multiple classifiers. These are finally combined to a single strong composite classifier similar to Bagging. There is theoretical and empirical evidence proving that ensemble learning techniques can reduce both the bias and variance components of errors made by ML models (Rokach, 2019).

Several ML algorithms are available in the literature that use Bagging and Boosting techniques (Odegua, 2019). In our study, we selected Random Forest (RF) as an ensemble algorithm from the Bagging techniques and Xgboost (Xgb) from the Boosting techniques. The RF algorithm combines bagging with bootstrap sampling. Xgb uses a highly scalable tree ensemble boosting algorithm. Even though there are many algorithms available in the literature that can train a model with small datasets, it is best to consider minimizing bias and variance to not only fit the test data but also generalize well on test/validation data (Maheswari, 2019). Certain algorithms are prone to overfitting if not carefully chosen. The traditional learning algorithms such as ensemble algorithms, perform better compared to deep learning architectures which utilize neural networks (Alom et al., 2019). This was observed when we trained a Neural Network model.

Additional problems encountered with some ML algorithms include class imbalance (Brownlee, 2020a), non-representative data (Menon, 2020), and the curse of dimensionality (Karanam, 2021). However, ensemble methods such as RF and Xgb are less likely to be associated with such challenges when using small datasets. We applied hyperparameter tuning (tuned hyperparameters are included in Table A1), cross-validation, stratified sampling, and resampling techniques (SMOTE Chawla et al., 2002) to overcome the challenges of a small dataset.

2.7. Interpretability techniques

Machine Learning models have long been considered black-box models until recently, when the research community discovered techniques to disentangle the internal mechanisms of the models. This has helped build trust in the use of ML models for sensitive applications, such as in the field of biomedicine (Rudin, 2019; Auslander et al., 2021; Papastefanopoulos et al., 2021). There are two scopes of interpretability in ML models, per sample interpretation (local) and overall interpretation (global). There are several software libraries available to interpret an ML model both locally and globally. Shaply Additive Values by Lundberg and Lee (2016) is an interpretability technique that uses coalition game theoretical approaches to explain a model's predictions. It has been implemented as a Python library named "SHAP," which stands for SHaply Additive exPlanations (Mazzanti, 2020). In SHAP, the feature values of a data instance act as players in a coalition. The computed SHAP values represent how to fairly distribute the prediction among the features. The explained SHAP model can be represented by the following equation (Bagheri, 2022),

$$g(x') = \phi_0 + \sum_{j=1}^{K} \phi_j x'_j \tag{3}$$

The $g(x')$ in the equation is the explanation model. Coalition vector is represented as $x' \in \{0,1\}^K$, where K is the maximum coalition size. The Shapley value is $\phi_j \in \mathbb{R}$, which is the feature attribution for a feature j. The Shapley value reveals how to fairly distribute a prediction among the features assuming that each feature value of the instance is a "player" in a game where prediction is the payout. In this study, the SHAP Python library (Lundberg and Lee, 2017) was used to interpret the models.

TABLE 1 Details of the experiments conducted.

Experiment number	Features	Sample size
1	SSW, PPS, WIC/WIN, ABR magnitude and phase, acoustic reflexes	134
2	SSW, PPS, WIC/WIN, ABR magnitude and phase, CEARs magnitude and phase, acoustic reflexes, frequency discrimination, gap detection, and amplitude modulation	46

SSW, staggered spondaic word; PPS, pitch pattern sequence; WIC, word in competing; WIN, word in noise; ABR, auditory brainstem responses; CEARs, cortical evoked auditory responses.

2.8. Experiments

In Experiment 1, only ABRs were considered as the number of CEARs was only available for 46 children. There were four ABR signals (two from both ears) considered from each of the 134 children, resulting in 536 data instances. For Experiments 2, each child had eight CEAR signals resulting in a total number of 368 instances for the dataset, including the ABRs. The train/test split was chosen to be 70:30 across all the experiments since a balance for both training and testing data was required due to the small dataset sizes. In all experiments, to find the confidence bounds of the model, iteratively 100 shuffled random splits of train/test (train/test

TABLE 2 Equations used to calculate the evaluation metrics. The k in the equations indicates a class (either BaPa, BaPn, BnPa, or BnPn).

Evaluation metric	Equation
Accuracy	$\frac{\sum_K TP+TN}{\sum_K TP+FP+FN+TN}$
Sensitivity/Recall	$\frac{TP_K}{TP_K+FN_K}$
Specificity	$\frac{TN_K}{TN_K+FP_K}$
Precision	$\frac{TP_K}{TP_K+FP_K}$
F1-score	$2 * \frac{Precision_K * Recall_K}{Precision_K + Recall_K}$
Informedness	$Sensitivity_K + Specificity_K - 1$
Markedness	$\frac{TP_k}{(TP_k+FP_k)} + \frac{TN_k}{(TN_k+FN_k)} - 1$

Accuracy is calculated by summating values from each class. Sensitivity, specificity, precision, F1-score, informedness, and markedness are shown for an individual class.
TP, true positive; TN, true negative; FP, false positive; FN, false negative; BnPa, behaviorally normal and physiologically abnormal; BnPn, behaviorally normal and physiologically normal; BaPa, behaviorally abnormal and physiologically abnormal; BaPn, behaviorally abnormal and physiologically normal.

split was not fixed) were considered from each ML algorithm. It results in 100 models trained on different train/test splits of the data. All the training iterations included both hyperparameter tuning using Random Search (Bergstra and Bengio, 2012) and stratified cross-validation (Brownlee, 2020b). This revealed how confident each model was in predicting the labels of the dataset.

2.9. Evaluation metrics and statistical tests

The performance of the ML models was evaluated using the metrics listed in Table 2. A true positive (TP) or a true negative (TN) indicates cases where the model and the label provided by the clinician agree. When the model and the labels disagree, false negatives (FN) and false positives (FP) are encountered. Calculating TN, TP, FN, and FP from the confusion matrix in the case of a multi-class classification problem is different compared to a binary classification problem. Further details on these calculations can be found in Grandini et al. (2020) and Shmueli (2019). The informedness and markedness were calculated from the sensitivity, specificity, and precision. Informedness combines both sensitivity and specificity to measure the consistency of predictions from the ML model, whereas markedness measures the trustworthiness of predictions made by the ML model (Powers, 2020).

Statistical significance tests were utilized to arrive at conclusions based on the evaluated metrics. The Friedman test was conducted to evaluate the significance of the results. The Friedman test is a non-parametric test used to compare group differences (Scheff, 2016).

2.10. Programming packages

All algorithms used were written in the Python programming language. Several software libraries were employed. The Pandas library (McKinney et al., 2011) was first used to pre-process the data. CWT analysis was carried out using the PyWavelets library (Lee et al., 2019). The Scikit-learn package (Pedregosa et al., 2011) contained all the ML algorithms that were used in the study. Finally, the SHAP library (Lundberg and Lee, 2017) was used to interpret the models.

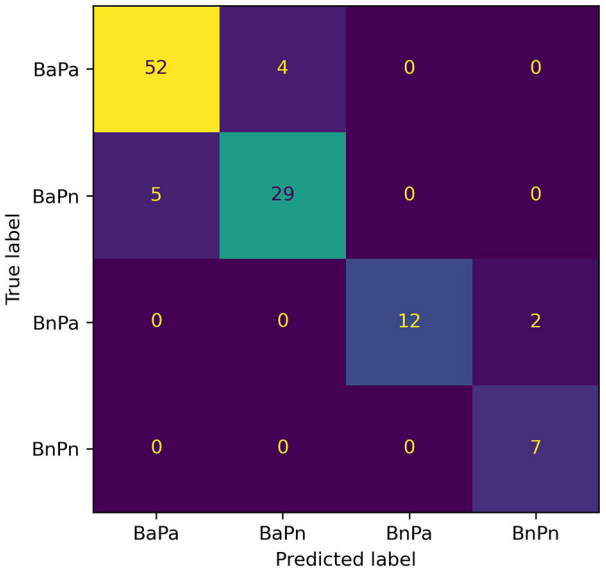


FIGURE 3
Confusion matrix for the RF model for Experiment 2. The diagonal shows the correctly classified data instances while the off-diagonal shows the incorrectly classified instances. BaPa, behaviorally abnormal and physiologically abnormal; BaPn- behaviorally abnormal and physiologically normal; BnPa, behaviorally normal and physiologically abnormal; BnPn, behaviorally normal and physiologically normal.

3. Results

The results obtained for the experiments as mentioned in Table 1 are shown in Figure 2. These results were obtained from training 100 different train/test splits from the data. Each data point shows the accuracy for the test set after the model was trained with hyperparameter tunning and cross-validation. In Experiment 1, the RF model has a negatively skewed distribution (mean = 68.3%, median = 68.5%) with a standard deviation of 4%, while the Xgb model shows a positively skewed distribution (mean = 66.6%, median = 65.7%) with a standard deviation of 5%. The median accuracy of the RF model is greater than the Xgb model (2.8%). In Experiment 2, the RF shows a positively skewed distribution (mean = 81.6%, median = 81.1%), while Xgb shows a negatively skewed distribution (mean =79.1%, median = 79.3%). The RF model shows slightly better median accuracy than Xgb (1.8% difference).

A Friedman test revealed a significant difference in the results of experiments $[X_r^2(3, N = 100) = 239.823, p < 0.05]$. The Bonferroni multiple comparison test was next used to compare pairwise performances for each experiment. The test revealed that each pair of experiments has significant differences in performance. Experiment 2 showed significantly better results ($p < 0.01$) for both RF and Xgb models compared to Experiment 1. Experiment 2 contained features derived from all the tests from both the behavioral and physiological test battery, whereas Experiment 1

10.3389/fauot.2023.1215965

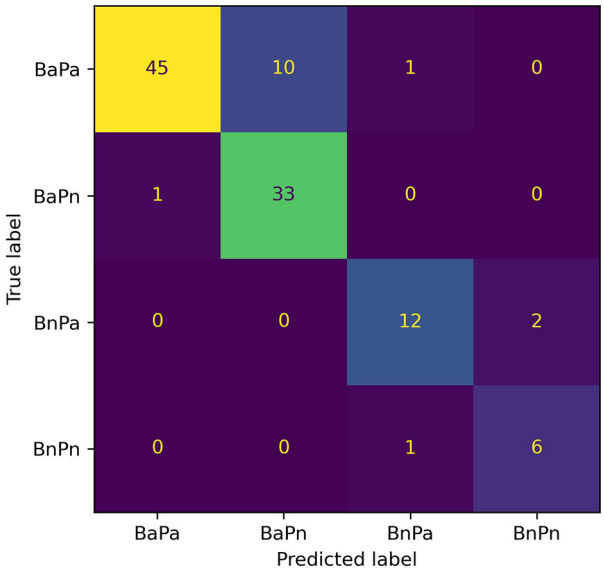


FIGURE 4
Confusion matrix for the Xgb model for Experiment 2. The diagonal shows the correctly classified data instances while the off-diagonal shows the incorrectly classified instances. BaPa, behaviorally abnormal and physiologically abnormal; BaPn, behaviorally abnormal and physiologically normal; BnPa, behaviorally normal and physiologically abnormal; BnPn, behaviorally normal and physiologically normal.

contained only features from ABRs due to the lack of data for CEARs.

The model with the greatest accuracy out of the 100 models generated from each different train/test splits of data for Experiment 2 is RF with a 90.1% overall accuracy and Xgb with an 86.5% overall accuracy. Figures 3, 4 show the confusion matrices for the two models. The diagonals in the two matrices show the correctly classified instances and the other indices show the incorrectly classified instances. Based on the confusion matrices, the performance metrics listed in Table 2 were calculated for each class. Table 3 shows the calculated performance metrics. It can be observed that the performance metrics of the RF model outperformed the Xgb model in most performance metrics for each class. Hence the RF model was selected as the best model.

The SHAP interpretations for the selected best-performing model, RF, are shown in Figure 5. The x-axis in the plot shows the mean SHAP values for each feature on the y-axis. The features on the y-axis are ordered from highest to lowest impact, from top to bottom. Each bar shows the contribution from each of the four APD sub-groups. A higher mean SHAP scores that the feature is largely contributing to the outcome of the model. The features contributing to the outcome the most were the SSW scores, and the features contributing the least were the right and left contralateral acoustic reflexes. From the physiological hearing test battery, the features derived from cortical responses were ranked higher compared to ABR features. From the acoustic reflexes, the ipsilateral recordings of both right and left ears were ranked higher compared to the contralateral recordings. The amplitude modulation at 20 Hz and gap detection from

TABLE 3 Evaluation metrics of Random Forest and Xgboost models of Experiment 2. All the values in the table are in percentages.

Class	Random forest						Xgboost					
	Sensitivity	Specificity	Precision	F1-score	Informedness	Markedness	Sensitivity	Specificity	Precision	F1-score	Informedness	Markedness
BaPa	92.9	90.9	91.2	92.0	83.8	83.8	80.4	98.2	97.8	88.2	78.5	80.9
BaPn	85.3	94.8	87.9	86.6	80.1	81.5	97.1	87.0	76.7	85.7	84.1	75.3
BnPa	85.7	100.0	100.0	92.3	85.7	97.9	85.7	97.9	85.7	85.7	83.7	83.7
BnPn	100.0	98.1	77.8	87.5	98.1	77.8	85.7	98.1	75.0	80.0	83.8	74.0

BnPa, behaviorally normal and physiologically abnormal; BnPn, behaviorally normal and physiologically normal; BaPa, behaviorally abnormal and physiologically abnormal; BaPn, behaviorally abnormal and physiologically normal.

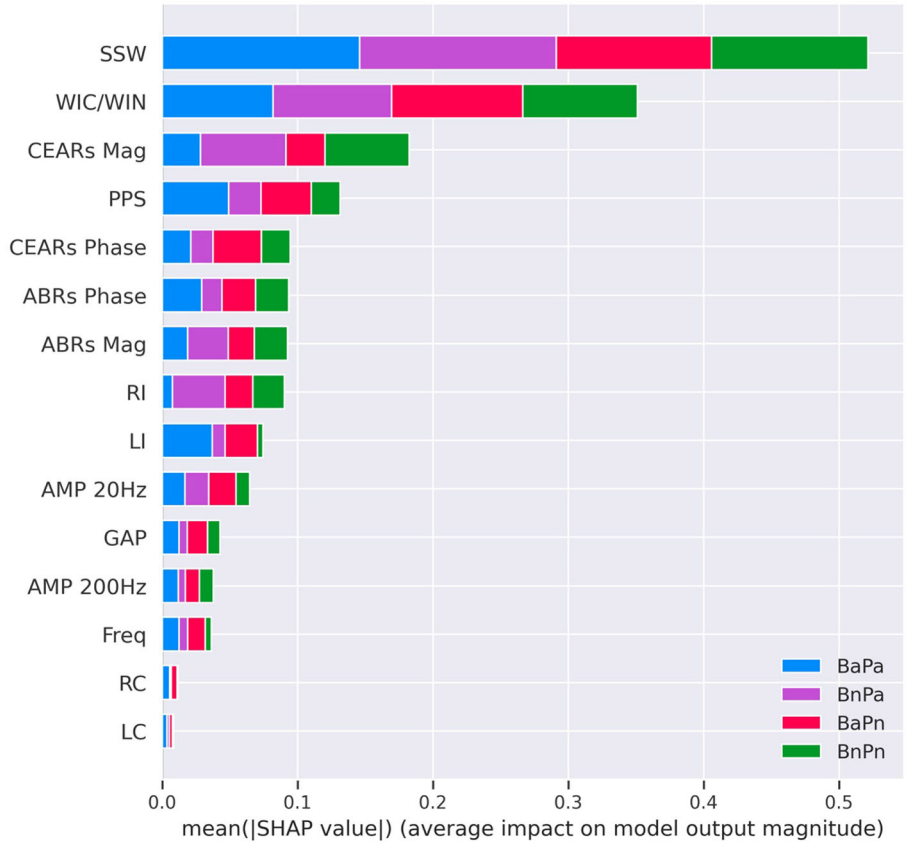


FIGURE 5
SHAP interpretations for the Random Forest model for Experiment 2. This was the best-performing model with an accuracy value of 90.1%. The X-axis of the diagram shows the mean SHAP values, and the y-axis shows the features contained in the model ordered from highest (top) to lowest (bottom) mean SHAP value. Each bar represents a combination of average values of the contribution from each subgroup. SSW, staggered spondaic word; WIC, word in competing; WIN, word in noise; CEARs, cortical evoked auditory responses; ABR, auditory brainstem responses; Mag, magnitude; PPS, pitch pattern sequence; BnPa, behaviorally normal and physiologically abnormal; BnPn, behaviorally normal and physiologically normal; BaPa, behaviorally abnormal and physiologically abnormal; BaPn, behaviorally abnormal and physiologically normal; LI, left Ipsi; RI, right Ipsi; LC, left contra; RC, right contra; GAP, GAP detection; AMP, AMPlitude modulation; Freq, frequency detection.

psychoacoustics showed higher mean SHAP values compared to frequency discrimination.

4. Discussion

The present work explored the use of supervised ML techniques to analyze data collected from children suspected of APD from a period of approximately six years. We determined RF and Xgb models to be the best suited for this study as they are both ensemble learning models that can perform well with small datasets. Data augmentation techniques can be used to improve the performance of ML models trained with small datasets. Here, we used the CTGAN augmentation technique (Xu et al., 2019). We found no significant difference in using CTGAN in conjunction with either the RF or Xgb models. CTGAN did significantly improve the results when used with a Neural Network. The accuracy of the neural network model without CTGAN for Experiments 1 and 2 were 51(±0.02) and 0.50(±0.04)%, respectively. The accuracies were improved to 69(±0.02)% for Experiment 1 and 70(±0.02)% for Experiment 2 with CTGAN. However, the accuracy obtained for Neural Network models was lower overall compared to RF and Xgb models.

The assessment of hearing thresholds is conducted early in the AP testing battery, and if a child passes each threshold, they are tested on the remainder of the battery. In this study, all children showed hearing thresholds (at conventional frequencies 250–8,000 Hz) within normal limits, and there were no considerable differences in hearing thresholds across the population. The ML results indicated that hearing thresholds showed the least impact on the outcome of the models. The features that were shown to impact the outcome of the models the most based on SHAP interpretations were derived from the behavioral tests. The SSW ranked first in both experiments, indicating that it impacted the outcome of the models the most. This is consistent with the literature, in which the SSW test is identified as a standard test for the auditory processing assessment (Emanuel et al., 2011). This consistency further indicates that ML models can comply with expert knowledge.

The inclusion of cortical responses in the physiological hearing assessments provided additional information about the neurophysiology of the auditory system. However, certain test

batteries do not include an assessment of cortical responses. Through our experiments, we noted that features derived from cortical responses largely contributed to the output of the models, as indicated by the SHAP interpretations. The features derived from CWT represent a summary of peak amplitudes and latencies. Even though the dataset with cortical features was smaller, the accuracy of both the RF and Xgb models was higher compared to the dataset without cortical features (Experiment 1). The extracted features suggested that cortical evoked responses provided significant information about auditory processing in these children. The emerging literature also suggests that cortical evoked responses are atypical in children referred for an auditory processing evaluation (Barker et al., 2017; Hussain et al., 2022). It is therefore recommended to evaluate cortical responses in the APD assessment and include these features in future studies that aim to use ML for automating APD diagnosis. Currently, it is not clear about the effect of maturation, morphology, and inter-subject variability in cortical evoked responses on these features. Hence, a thorough study of the features with a larger dataset is required in the future. It will help clinical understanding and the Machine Learning model reach higher accuracies.

The ipsilateral acoustic reflexes from the physiological hearing test battery also contributed largely to the model outcome compared to the contralateral reflexes based on the SHAP interpretations. In exploring the data, only a few children had elevated thresholds reflected through ipsilateral reflexes, and most children showed reflex thresholds within the normal limits overall.

Psychoacoustic tests use non-speech stimuli and can be used to validate the results of behavioral tests. The data from these tests were included in Experiment 2. However, the contribution from psychoacoustic tests was lesser compared to the behavioral assessments that used speech stimuli and the physiological tests. In the final model, amplitude modulation detection at 20 Hz and GAP detection were the tests that provided the most impact to the model outcome compared to the frequency detection and amplitude modulation detection at 200 Hz. After discussing with clinicians, it was found that sometimes children have difficulties with performing the frequency detection test compared to GAP in noise. In addition, detecting amplitude modulation at 20 Hz is easier compared to 200 Hz. Hence, this was further evidence that the model can be used to output accurate predictions in accordance with current clinical knowledge.

There are only a few studies in the literature that use ML techniques to analyze APD data in children (Strauss et al., 2004; Sharma et al., 2019; Cassandro et al., 2021). Sharma et al. (2019) used behavioral assessments data to cluster APD children into four sub-groups using hierarchical clustering techniques. The auditory processing assessments used by Sharma et al. (2019) were different than those used in the present study; we used both behavioral and physiological data, as recommended by [American Speech-Language-Hearing Association (ASHA), 2005]. The sub-groups identified by Sharma et al. (2019) have very few similarities to those identified in the present study. However, the group identified as "global deficit" is similar to the group we identify as "behaviorally abnormal and physiologically abnormal (BaPa)," where all assessments are outside of the normal thresholds. The study conducted by Strauss et al. (2004) used the β-waveform of the binaural interaction component in auditory brainstem responses along with a support vector machine model to detect APD in children. The study did not identify subgroups in the data, but rather aimed to identify children at risk for APD from those not at risk. Cassandro et al. (2021) have used cluster analysis to identify issues in students tested for dyslexia accompanied by poor auditory skills. Out of the four participants in the cluster who had poor audiometric profiles and were suspected of APD, only one subject was identified as APD.

The clinical workflow used here can be adopted in future work aiming to study APD data as we followed a comprehensive test battery based on ASHA guidelines. The use of ML techniques discussed in this paper may also be applied to future studies aiming to develop automated platforms to assess other clinical test batteries. Since this study focused on the technical aspects of ML, we did not discuss the clinical management of the identified sub-groups in detail. This would require further work by clinicians and researchers. However, we believe our study may aid such discussions as we have presented an objective tool to categorize children with APD into clinical sub-groups. We have further revealed the contribution of each assessment contained in the AP test battery on the model outcomes. It should be noted that the use of ML tools are meant to complement rather than replace clinical decision making.

There are relatively few referrals made for APD assessments in children and there is no definitive way to determine how much data is needed for an ML experiment before collecting the data. Our experiments were done based on the limited data we had available. Hence, the sample size of the clinical population is the main limitation of our study. A larger dataset will ensure improved generalization, model performance, stability, and validity in machine learning models. Future studies should be conducted with a larger data set in children referred for an auditory processing evaluation. A wide variety of complex algorithms, such as deep neural networks, could be explored with larger datasets which will help to derive a strong understanding of the clinical problem and reduce the number of tests used in the diagnosis. Hence, this study can be viewed as exploratory, where future studies may adopt our methods from both clinical and ML workflows. Future studies may explore solutions to the difficulties associated with collecting APD datasets such as forming larger, multi-center collaborations. One such solution may be the use of a federated ML system, in which researchers for different centers may contribute training data to the same model without exposing personal information (Yang et al., 2019).

5. Conclusion

The purpose of this study was to explore the use of ML techniques as a potential tool to aid in the analysis of the AP test battery. Data from children suspected of APD were classified into clinical sub-groups based on their performances on both behavioral and physiological hearing assessments. The RF model was shown to perform the best, with an average accuracy of 90%, an average sensitivity of 91%, and an average specificity of 96% for all sub-groups. The model was able to identify the critical subgroup BaPa, in which children performed poorly in both behavioral and physiological assessments, with a sensitivity and specificity

of 93 and 91%, respectively. The group that performed within normal limits in the test set (BnPn) were correctly identified with a sensitivity and specificity of 100 and 98%, respectively. This study further highlighted the utility of each individual test contained within the AP test battery in making predictions that agree with clinical understanding.

Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: the data used in this study are not publicly available due to ethical concerns regarding participant privacy and confidentiality. Requests to access these datasets should be directed to hwimalar@uwo.ca.

Ethics statement

The studies involving human participants were reviewed and approved by Western University Research Ethics Board, University of Western, London, Ontario, Canada. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

HW: conceptualization, methodology, software, data curation, writing—original draft, validation, writing—review and editing, and visualization. SA-V: conceptualization, validation, formal analysis, investigation, data curation, and writing—review and editing. MD: conceptualization, data curation, formal analysis, investigation, and writing—review and editing. CA: conceptualization, validation, investigation, resources, data curation, writing—review and editing, supervision, and project administration. SA: conceptualization, writing—review and editing, supervision, and funding acquisition. PA: conceptualization, validation, formal analysis, resources, data curation, writing—review and editing, supervision, project administration, and funding acquisition. JS and HL: conceptualization, validation, resources, writing—review and editing, supervision, and funding acquisition. All authors contributed to the article and approved the submitted version.

Funding

Funding was provided by the Natural Sciences and Engineering Council of Canada (NSERC-RGPIN-2017-04755) and Ontario Research Fund (ORF-RE08-072).

Acknowledgments

The authors greatly appreciate the support provided by Lauren Siegel for manuscript editing and review.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Agrawal, D., Dritsakis, G., Mahon, M., Mountjoy, A., and Bamiou, D. E. (2021). Experiences of patients with auditory processing disorder in getting support in health, education, and work settings: findings from an online survey. *Front. Neurol.* 12:607907. doi: 10.3389/fneur.2021.607907

Ahmad, M. A., Eckert, C., and Teredesai, A. (2018). "Interpretable machine learning in healthcare," in *Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, 559–560. doi: 10.1145/3233547.3233667

Allen, P., and Allan, C. (2014). Auditory processing disorders: relationship to cognitive processes and underlying auditory neural integrity. *Int. J. Pediatr. Otorhinolaryngol.* 78, 198–208. doi: 10.1016/j.ijporl.2013.10.048

Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M. S., et al. (2019). A state-of-the-art survey on deep learning theory and architectures. *Electronics* 8:292. doi: 10.3390/electronics8030292

American Speech-Language-Hearing Association (ASHA) (2005). *Central auditory processing disorders*, Technical report. Available online at: www.asha.org/policy (accessed December 20, 2021).

American Speech-Language-Hearing Association (ASHA) (1970). *Guidelines for Fitting and Monitoring FM Systems*. Available online at: https://www.asha.org/policy/GL2002-00010/ (retrieved February 18, 2022).

Auslander, N., Gussow, A. B., and Koonin, E. V. (2021). Incorporating machine learning into established bioinformatics frameworks. *Int. J. Mol. Sci.* 22:2903. doi: 10.3390/ijms22062903

Bagheri, R. (2022). *Introduction to Shap Values and Their Application in Machine Learning*. Available online at: https://towardsdatascience.com/introduction-to-shap-values-and-their-application-in-machine-learning-8003718e6827 (retrieved April 05, 2022).

Bamiou, D.-E., Musiek, F., and Luxon, L. (2001). Aetiology and clinical presentations of auditory processing disorders?a review. *Arch. Dis. Childh.* 85, 361–365. doi: 10.1136/adc.85.5.361

Barker, M. D., Kuruvilla-Mathew, A., and Purdy, S. C. (2017). Cortical auditory-evoked potential and behavioral evidence for differences in auditory processing between good and poor readers. *J. Am. Acad. Audiol.* 28, 534–545. doi: 10.3766/jaaa.16054

Bellis, T. J., and Ferre, J. M. (1999). Multidimensional approach to the differential diagnosis of central auditory processing disorders in children. *J. Am. Acad. Audiol.* 10, 319–328. doi: 10.1055/s-0042-1748503

Bergstra, J., and Bengio, Y. (2012). Random search for hyper-parameter optimization. *J. Mach. Learn. Res.* 13, 281–305.

Breiman, L. (1996). Bagging predictors. *Mach. Learn.* 24, 123–140. doi: 10.1007/BF00058655

Brownlee, J. (2020a). *8 Tactics to Combat Imbalanced Classes in Your Machine Learning Dataset*. Available online at: https://machinelearningmastery.com/tactics-to-combat-imbalanced-classes-in-your-machine-learning-dataset/ (retrieved June 9, 2022).

Brownlee, J. (2020b). *How to Fix k-Fold Cross-Validation for Imbalanced Classification*. Available online at: https://machinelearningmastery.com/cross-validation-for-imbalanced-classification/ (retrieved November 15, 2021).

Cassandro, C., Manassero, A., Landi, V., Aschero, G., Lovallo, S., Albera, A., et al. (2021). Auditory processing disorders: diagnostic and therapeutic challenge. *Otorhinolaryngology* 71, 120–124. doi: 10.23736/S2724-6302.21.02387-2

Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). Smote: synthetic minority over-sampling technique. *J. Artif. Intell. Res.* 16, 321–357. doi: 10.1613/jair.953

Chermak, G. D. (2002). Deciphering auditory processing disorders in children. *Otolaryngol. Clin. N. Am.* 35, 733–749. doi: 10.1016/S0030-6665(02)00056-7

Chermak, G. D., Musiek, F. E., and Craig, C. H. (1997). Central Auditory Processing Disorders: New Perspectives. *Singular*. 370–376.

Cline, L. (2001). *CAPD Technical Assistance Paper*. Technical report. Florida Department of Education. Available online at: https://www.aitinstitute.org/CAPD_technical_assistance_paper.pdf (retrived 4 January 2022).

Davenport, T., and Kalakota, R. (2019). The potential for artificial intelligence in healthcare. *Fut. Healthc. J.* 6, 94–98. doi: 10.7861/futurehosp.6-2-94

Emanuel, D. C., Ficca, K. N., and Korczak, P. (2011). Survey of the diagnosis and management of auditory processing disorder. *Am. J. Audiol.* 20, 48–60. doi: 10.1044/1059-0889(2011/10-0019)

Freund, Y., and Schapire, R. E. (1996). "Experiments with a new boosting algorithm," in *ICML, Vol. 96*, 148–156.

Grandini, M., Bagli, E., and Visani, G. (2020). Metrics for multi-class classification: an overview. *arXiv:2008.05756*. doi: 10.48550/arXiv.2008.05756

Hornickel, J., Zecker, S. G., Bradlow, A. R., and Kraus, N. (2012). Assistive listening devices drive neuroplasticity in children with dyslexia. *Proc. Natl. Acad. Sci. U.S.A.* 109, 16731–16736. doi: 10.1073/pnas.1206628109

Hussain, R. O., Kumar, P., and Singh, N. K. (2022). Subcortical and cortical electrophysiological measures in children with speech-in-noise deficits associated with auditory processing disorders. *J. Speech Lang. Hear. Res.* 65, 4454–4468. doi: 10.1044/2022_JSLHR-22-00094

Iliadou, V. V., Chermak, G. D., Bamiou, D.-E., and Musiek, F. E. (2019). Gold standard, evidence-based approach to diagnosing APD. *Hear. J.* 72, 42–45. doi: 10.1097/01.HJ.0000553582.69724.78

Iliadou, V. V., Ptok, M., Grech, H., Pedersen, E. R., Brechmann, A., Deggouj, N., et al. (2017). A European perspective on auditory processing disorder-current knowledge and future research focus. *Front. Neurol.* 2017:622. doi: 10.3389/fneur.2017.00622

Iliadou, V. V., Ptok, M., Grech, H., Pedersen, E. R., Brechmann, A., Deggouj, N., et al. (2018). European 17 countries consensus endorses more approaches to APD than reported in Wilson 2018. *Int. J. Audiol.* 57, 395–396. doi: 10.1080/14992027.2018.1442937

Ivey, R. (1969). *Words in ipsilateral competition (WIC)* (Unpublished Master Thesis). Colorado State University, Fort Collins, CO, United States.

Johnston, K. N., John, A. B., Kreisman, N. V., Hall, J. W. III, Crandell, C. C., Johnston, K. N., et al. (2009). Multiple benefits of personal FM system use by children with auditory processing disorder (APD). *Int. J. Audiol.* 48, 371–383. doi: 10.1080/14992020802687516

Karanam (2021). Available online at: https://towardsdatascience.com/curse-of-dimensionality-a-curse-to-machine-learningc122ee33bfeb (accessed December 5, 2021).

Katz, J. (1998). *The Staggered Spondaic Word Test (SSW)*. Vancouver, WA.

Keith, W. J., and Purdy, S. C. (2014). "Assistive and therapeutic effects of amplification for auditory processing disorder," in *Seminars in Hearing, Vol. 35* (Thieme Medical Publishers), 27–38. doi: 10.1055/s-0033-1363522

Lee, G., Gommers, R., Waselewski, F., Wohlfahrt, K., and O'Leary, A. (2019). Pywavelets: A python package for wavelet analysis. *J. Open Source Softw.* 4:1237. doi: 10.21105/joss.01237

Lundberg, S. M., and Lee, S.-I. (2017). "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems*, eds I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Curran Associates), 30.

Maheswari, J. P. (2019). *Breaking the Curse of Small Datasets in Machine Learning: Part 1*. Available online at: https://towardsdatascience.com/breaking-the-curse-of-small-datasets-in-machine-learning-part-1-36f28b0c044d (retrieved December 10, 2021).

Mazzanti, S. (2020). Shap values explained exactly how you wished someone explained to you. *Towards Data Sci.* 3:2020. Available online at: https://towardsdatascience.com/shap-explained-the-way-i-wish-someone-explained-it-to-meab81cc69ef30 (retrieved December 11, 2021).

Mckinney, W. (2011). *Pandas: a Foundational Python Library for Data Analysis and Statistics*. Available online at: https://www.dlr.de/sc/en/Portaldata/15/Resources/dokumente/pyhpc2011/submissions/pyhpc2011_submission_9.pdf

Menon, S. (2020). Stratified sampling in machine learning. Available online at: https://medium.com/analytics-vidhya/stratified-sampling-in-machine-learning-f5112b5b9cfe-

Moore, D. R., Sieswerda, S. L., Grainger, M. M., Bowling, A., Smith, N., Perdew, A., et al. (2018). Referral and diagnosis of developmental auditory processing disorder in a large, United States hospital-based audiology service. *J. Am. Acad. Audiol.* 29, 364–377. doi: 10.3766/jaaa.16130

Nagao, K., Riegner, T., Padilla, J., Greenwood, L. A., Loson, J., Zavala, S., and Morlet, T. (2016). Prevalence of auditory processing disorder in school-aged children in the mid-Atlantic region. *J. Am. Acad. Audiol.* 27, 691–700. doi: 10.3766/jaaa.15020

Odegua, R. (2019). "An empirical study of ensemble techniques (bagging, boosting and stacking)," in *Proc. Conf. Deep Learn* (IndabaXAt).

Papastefanopoulos, V., Kotsiantis, S., and Linardatos, P. (2021). Explainable AI: a review of machine learning interpretability methods. *Entropy* 23, 1–45. doi: 10.3390/e23010018

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* 12, 2825–2830. doi: 10.5555/1953048.2078195

Pinheiro, M. L. (1977). Tests of central auditory function in children with learning disabilities. *Central Audit. Dysfunct.* 223–256.

Powers, D. M. (2020). Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation. *arXiv preprint arXiv:2010.16061*. doi: 10.48550/arXiv.2010.16061

Rance, G., Saunders, K., Carew, P., Johansson, M., and Tan, J. (2014). The use of listening devices to ameliorate auditory deficit in children with autism. *J. Pediatr.* 164, 352–357. doi: 10.1016/j.jpeds.2013.09.041

Rokach, L. (2019). *Ensemble Learning: Pattern Classification Using Ensemble Methods*. World Scientific. doi: 10.1142/11325

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Mach. Intell.* 1, 206–215. doi: 10.1038/s42256-019-0048-x

Schapire, R. E. (1990). The strength of weak learnability. *Mach. Learn.* 5, 197–227. doi: 10.1007/BF00116037

Scheff, S. W. (2016). *Fundamental Statistical Principles for the Neurobiologist: A Survival Guide*. Academic Press.

Sharma, M., Purdy, S. C., and Humburg, P. (2019). Cluster analyses reveals subgroups of children with suspected auditory processing disorders. *Front. Psychol.* 10:2481. doi: 10.3389/fpsyg.2019.02481

Sharma, M., Purdy, S. C., and Kelly, A. S. (2009). Comorbidity of auditory processing, language, and reading disorders. *J. Speech Lang. Hear. Res.* 52, 706–722. doi: 10.1044/1092-4388(2008/07-0226)

Sharma, M., Purdy, S. C., and Kelly, A. S. (2012). A randomized control trial of interventions in school-aged children with auditory processing disorders. *Int. J. Audiol.* 51, 506–518. doi: 10.3109/14992027.2012.670272

Shmueli, B. (2019). *Multi-Class Metrics Made Simple, Part 1: Precision and Recall*. Available online at: https://towardsdatascience.com/multi-class-metrics-made-simple-part-i-precision-and-recall-9250280bddc2

Smart, J. L., Purdy, S. C., and Kelly, A. S. (2018). Impact of personal frequency modulation systems on behavioral and cortical auditory evoked potential measures of auditory processing and classroom listening in school-aged children with auditory processing disorder. *J. Am. Acad. Audiol.* 29, 568–586. doi: 10.3766/jaaa.16074

Starr, A., and Achor, L. J. (1975). Auditory brain stem responses in neurological disease. *Arch. Neurol.* 32, 761–768. doi: 10.1001/archneur.1975.00490530083009

Strauss, D., Delb, W., and Plinkert, P. (2004). Objective detection of the central auditory processing disorder: a new machine learning approach. *IEEE Trans. Biomed. Eng.* 51, 1147–1155. doi: 10.1109/TBME.2004.827948

Torrence, C., and Compo, G. P. (1998). A practical guide to wavelet analysis. *Bull. Am. Meteorol. Soc.* 79, 61–78. doi: 10.1175/1520-0477(1998)079<0061:APGTWA>2.0.CO;2

Weihing, J., Chermak, G. D., and Musiek, F. E. (2015). Auditory training for central auditory processing disorder. *Semin. Hear*. 36, 199–215. doi: 10.1055/s-0035-1564458

Wilson, R. H. (2003). Development of a speech-in-multitalker-babble paradigm to assess word-recognition performance. *J. Am. Acad. Audiol*. 14, 453–470. doi: 10.1055/s-0040-1715938

Wimalarathna, H., Ankmnal-Veeranna, S., Allan, C., Agrawal, S. K., Allen, P., Samarabandu, J., et al. (2021). Comparison of machine learning models to classify auditory brainstem responses recorded from children with auditory processing disorder. *Comput. Methods Prog. Biomed*. 200:105942. doi: 10.1016/j.cmpb.2021.105942

Xu, L., Skoularidou, M., Cuesta-Infante, A., and Veeramachaneni, K. (2019). "Modeling tabular data using conditional GAN," in *Advances in Neural Information Processing Systems, 32*.

Yang, Q., Liu, Y., Chen, T., and Tong, Y. (2019). Federated machine learning: concept and applications. *ACM Trans. Intell. Syst. Technol*. 10, 1–19. doi: 10.1145/3298981

10.3389/fauot.2023.1215965

Appendix

TABLE A1 Tuned hyperparameters of Machine Learning models used in the study.

ML model	Tuned hyperparameters
Random forest	Number of estimators, size of the random subsets of features, maximum depth of individual trees, minimum samples to split on at an internal node of the trees, minimum leaf nodes after splitting a node
Xgboost	Column sample by tree, gamma, learning rate, maximum depth, number of estimators, subsample, regularization parameter alpha
Neural network	Hidden Layers, Activation function, Optimization function, Learning Rate, Iterations

frontiers | Frontiers in Audiology and Otology

TYPE Original Research
PUBLISHED 19 September 2023
DOI 10.3389/fauot.2023.1226946

Check for updates

OPEN ACCESS

EDITED BY
Karina De Sousa,
University of Pretoria, South Africa

REVIEWED BY
Sridhar Krishnamurti,
Auburn University, United States
Richard Charles Dowell,
The University of Melbourne, Australia

*CORRESPONDENCE
Bernhard Englitz
✉ bernhard.englitz@donders.ru.nl

†PRESENT ADDRESS
Noah J. Schlegel,
Neurotechnology, Faculty IV Electrical Engineering and Computer Science, Technische Universität Berlin, Berlin, Germany

‡These authors have contributed equally to this work

RECEIVED 22 May 2023
ACCEPTED 22 August 2023
PUBLISHED 19 September 2023

CITATION
Alishbayli A, Schlegel NJ and Englitz B (2023) Using auditory texture statistics for domain-neutral removal of background sounds. *Front. Audiol. Otol.* 1:1226946. doi: 10.3389/fauot.2023.1226946

COPYRIGHT
© 2023 Alishbayli, Schlegel and Englitz. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Using auditory texture statistics for domain-neutral removal of background sounds

Artoghrul Alishbayli‡, Noah J. Schlegel†‡ and Bernhard Englitz*

Computational Neuroscience Lab, Donders Institute, Radboud University, Nijmegen, Netherlands

Introduction: Human communication often occurs under adverse acoustical conditions, where speech signals mix with interfering background noise. A substantial fraction of interfering noise can be characterized by a limited set of statistics and has been referred to as auditory textures. Recent research in neuroscience has demonstrated that humans and animals utilize these statistics for recognizing, classifying, and suppressing textural sounds.

Methods: Here, we propose a fast, domain-free noise suppression method exploiting the stationarity and spectral similarity of sound sources that make up sound textures, termed Statistical Sound Filtering (SSF). SSF represents a library of spectrotemporal features of the background noise and then compares this against instants in speech-noise-mixtures to subtract contributions that are statistically consistent with the interfering noise.

Results: We evaluated the performance of SSF using multiple quality measures and human listeners on the standard TIMIT corpus of speech utterances. SSF improved the sound quality across all performance metrics, capturing different aspects of the sound. Additionally, human participants reported reduced background noise levels as a result of filtering, without any significant damage to speech quality. SSF executes rapidly (~100× real-time) and can be retrained rapidly and continuously in changing acoustic contexts.

Discussion: SSF is able to exploit unique aspects of textural noise and therefore, can be integrated into hearing aids where power-efficient, fast, and adaptive training and execution are critical.

KEYWORDS

sound textures, noise reduction, speech enhancement, hearing aids, statistical learning

Highlights

- Acoustic textures are defined by time-independent statistics and occur frequently.
- Learning a library of spectrotemporal features rapidly filters out acoustic textures.
- Filtering suppresses background noises across different auditory textures.
- Human and automatic performance evaluation demonstrate suppression.
- Filtering is fast and can thus be integrated into mobile devices and hearing aids.

Introduction

Auditory signals rarely arrive at the ear in pure and unambiguous form but are usually mixed with other competing sounds. Masking of relevant information by irrelevant noise is not unique to the auditory system: occlusion of surfaces in a complex visual scene poses an analogous signal processing problem that requires disambiguation and segregation of sources (Handel, 2006; Minaee et al., 2022). However, unlike in the visual domain, in the auditory domain, the noise is superimposed onto the signal which creates an

ill-posed source separation problem for the auditory system (McDermott, 2009). During the course of evolution, the auditory system evolved an impressive ability to extract relevant information from complex scenes with multiple interfering sources, an effect known as the cocktail party effect (Middlebrooks et al., 2017). Although the specific neural mechanisms responsible for this ability remain poorly understood, extensive research has documented the processes through which the auditory system of an organism responds to the noise in complex auditory scenes. These processes include segregation by fundamental frequency, dip listening, better-ear listening, binaural unmasking, etc. (see Culling and Stone, 2017 for an overview).

However, what is achieved seamlessly by a normally functioning system, becomes a challenge with hearing loss (Koole et al., 2016). To address the issue, various noise reduction approaches have been developed over the past few decades (Loizou, 2013b; Henry et al., 2021). They vary in multiple dimensions: some of the methods use real-time data (Braun et al., 2021) collected using a single channel microphone (Huang and Benesty, 2012; Lee and Theunissen, 2015), while others are used in post-processing and utilize multiple recording channels (Tzirakis et al., 2021), which can provide extra spatial cues that can aid in solving the problem. While noise reduction approaches typically do not improve speech intelligibility itself, the subjective listening experience does improve with indications of less cognitive load for normal hearing persons (Sarampalis et al., 2009) and reduced listening effort for less distorted speech in people with hearing loss (Fiedler et al., 2021). Classically, noise reduction algorithms use signal processing methods, but recent developments in the field have led to increased use of machine learning techniques that allow more flexibility in terms of target selection and enhancement in more complex, non-stationary background noise conditions because they make fewer assumptions about the nature of noise.

Sounds with relatively constant statistical features over time have been categorized as acoustic textures, for example, the sound of rain, fire, or flocks of birds (McDermott and Simoncelli, 2011). Most auditory textures are physically generated by the superposition of a limited range of constituent sounds, which occur independently or with limited statistical dependencies between the constituent sounds. Previous research has shown that humans can recognize and differentiate acoustic textures on the basis of their statistics (McDermott and Simoncelli, 2011; McDermott et al., 2013). However, previous approaches in noise reduction have not made use of this inherent structure of acoustic textures, despite their frequent role as background sounds during everyday audition.

In this study, we propose a noise reduction method that utilizes these inherent statistical regularities to attenuate background sounds and thus improve the signal-to-noise ratio of embedded speech sounds. Specifically, we represent the ensemble of constituent sounds using samples of background sounds, identified around or between speech samples. Assuming an additive mixture, we then clean the speech-in-noise sample by identifying exemplars that provide the best match to the instantaneous spectrogram. This approach extends previous approaches of spectral noise subtraction (Boll, 1979) by relating it to the statistics of natural background sounds. Importantly, we do not create an explicit statistical model of the background noise, as (i) this usually requires more data to be well-constrained and (ii) the internal, statistical predictability would be too limited to remove specific instantaneous sounds randomly occurring inside the auditory textures (see Discussion for details).

Applied to the TIMIT database in the context of artificial and natural acoustic textures, the filtered result exhibited an improved representation of the speech as measured by a standard deep neural network (DNN) based speech recognition system, spectrogram correlations, and automated estimation of speech quality. Similarly, online psychoacoustic experiments on human participants also indicated an improvement in the quality of the sounds. In comparison with other machine learning approaches, our system does not require extensive training but rapidly adapts to the recent history of background noise, and runs faster-than-real-time on computational resources currently available in mobile phones. If translated to specialized processors in hearing aids, it may be feasible to run on preprocessors for hearing aids and cochlear implants.

Methods

Sound material

Generation of artificial textures

Auditory textures used as background noise were generated using a slight modification of the "Sound Texture Synthesis Toolbox" developed by McDermott and Simoncelli (2011). The changes allowed the mixing of statistical features from different sound sources while sampling the statistical space of natural sounds in a controlled fashion. In total, we generated six different textures with different combinations of marginal (mean, variance) and correlation statistics taken from real textures (Table 1). The algorithm calculates marginal moments and/or correlations from the example sound which are then taken as target statistics for synthesis. The synthesis starts from a Gaussian white noise which is then iteratively shaped to match predetermined target statistics using the conjugate gradient method. Those statistics were transformed per frequency bin, which makes the resulting sound rather similar to the original if that sound is well-defined by the used statistics. In the case of sound textures, it has been shown that this procedure is able to produce compelling sounds that are indistinguishable from original sources in many cases (McDermott and Simoncelli, 2011). Multiple 50 s texture files were generated for each set of statistics, which was long enough for the combination with a few speech samples while still ensuring convergence of the synthesis. Figure 1A shows spectrograms of generated textures.

Source of real texture

To also test the process with a real texture, we needed a natural texture example with constant statistics and a duration of at least 90 min (the test part of TIMIT is roughly 87 min long). We chose a 3-h continuous rain recording (The Relaxed Guy, 2014) with subjectively little change over time. The first 30 s of the file were discarded to reduce the potential statistical effects of fading in, it was then downsampled to 16 kHz and saved as a WAV file.

Speech samples

Human speech samples used in this study were obtained from the TIMIT corpus (Garofolo et al., 1993), which contains broadband recordings of 630 speakers of eight major dialects of American English, each reading ten phonetically rich sentences. For objective testing of the algorithm, the entire test set of TIMIT was used, comprising 1,680 files. For human evaluation of the algorithm, due to overall time limitation (1h), we selected a subset of 36 unique speech files where variables such as gender ($n = 2$), dialect ($n = 8$), speaker ID ($n = 33$) and sentence type ($n = 3$) were made as diverse as possible (see below for other details on the human experiment).

TABLE 1 Source sounds for artificial texture generation.

TextureID	Envelope mean	Envelope variance	Correlation
$Base_1$ $\rho\uparrow$	Bee swarm	Pile driver	Fast running river
$Base_1$ $\rho\downarrow$	Bee swarm	Pile driver	Jogging on gravel
$Base_2$ $\rho\uparrow$	Bubbling water	Drumroll	Fast running river
$Base_2$ $\rho\downarrow$	Bubbling water	Drumroll	Jogging on gravel
$Base_3$ $\rho\uparrow$	Shaking coins	Crowd noise	Fast running river
$Base_3$ $\rho\downarrow$	Shaking coins	Crowd noise	Jogging on gravel

Mixing of speech and noise

For each texture type, a speech-texture mix was created with the TIMIT test set. Every speech file was mixed with an individual texture sequence after both signals were normalized to a standard deviation of 1. For testing of the algorithm, a constant SNR of 0 dB was set for the mixture, except when SNR was varied during parameter exploration. After combining, the signal was normalized to a fixed standard deviation of 0.05 to avoid clipping in the WAV files which occurs at absolute values above 1. Texture samples were drawn uniquely and continuously without overlap from the created or real texture files.

Filtering algorithm

Briefly, the filtering process was a matched subtraction on the spectrogram level using a fast k-nearest neighbors (KNN) search

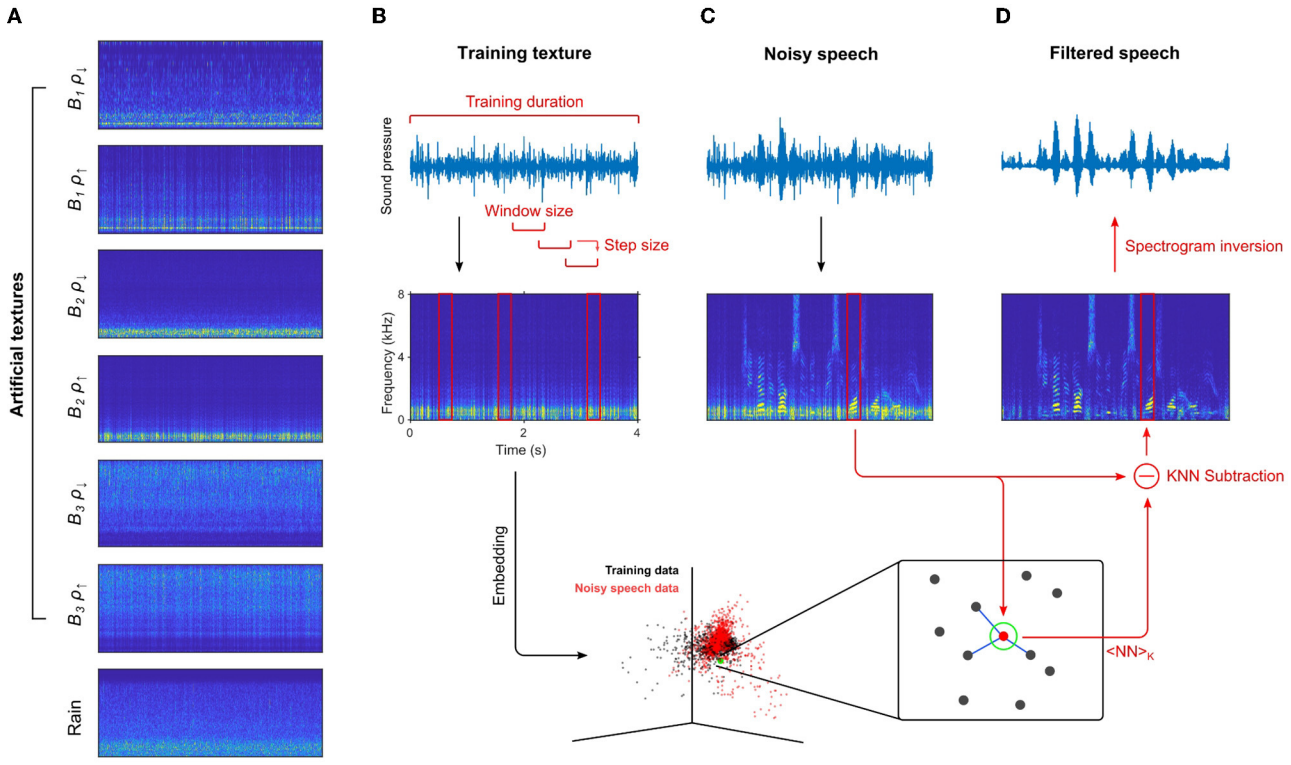


FIGURE 1
Speech and noise sources used and filtering principle architecture. **(A)** Spectrograms of artificial and natural textures used as background noise in testing of the algorithm. **(B)** Filter training and evaluation on noisy data. The training data **(top)** is converted to a spectrogram, governed by the window length, step size and window function. In the training step, short sections of the spectrogram are high-pass filtered and then embedded (bottom, black) in high-dimensional vector space (dimension: $N_{FFT} \times N_{WindowSteps}$) to represent the short-term spectrotemporal statistics. **(C)** Samples of speech with added noise are transformed to spectrograms **(top)** and embedded into the same space as the training data (bottom, red). **(D)** Filtering of the noisy speech sample is performed by selecting a number of nearest neighbors inside the embedding for each time-instant, averaging them and subtracting them from the noisy data. A playback-capable sound pressure representation of the filtered spectrogram is then generated via spectrogram inversion using the original phases from the noisy speech sample **(top)**.

over the training data as an estimator of the texture contribution to the sound (subtrahend), with appropriate transformations between sound pressure and spectrogram representations.

Training of the filter

The internal representation of the filter was spectrotemporal snippets from the training data represented as vectors. For this purpose, the training data was first transformed from sound pressure to a spectrogram using the short-term Fourier transform, represented as $S_T(t,f)$ below, where T indicates the training data. As usual, this transformation is parameterized by the window size and the step size. Typical values were 16 and 2 ms, respectively, but we explore the effect of these and other parameters in Figures 2E–K. The spectrogram was high-pass filtered, by subtracting the local context, i.e., the temporal average over a bidirectional window of length $T_{Multistep} = 50$ ms per frequency bin, which deemphasizes speech contributions to the instantaneous spectrum. The training data was then represented as points in a high-dimensional space, by linearizing short segments of dimension $N_{FFT} \times N_{WindowSteps}$, where $N_{WindowSteps}$ is number of subsequent time-steps embedded, i.e.,

$$E(S_T(t,f)) \rightarrow \Re^{N_{FFT} x N_{WindowSteps}}(t).$$

The resulting representation discretely approximates the distribution of the texture in the coordinates of the spectrotemporal snippets by sampling it. This representation captures the joint occurrence of different frequencies over adjacent time points in the texture. We also tried directly representing products of frequency channels, which, however, did not improve performance, while strongly increasing the runtime.

The training data was provided in two different ways: either, a single textural sound of length L_{Train} was provided (which we refer to as "supervised"); or a speech-in-noise sample with a total amount of texture L_{Train} was provided (which we refer to as "unsupervised"). In the first case, the algorithm knew the training data, in the second case, we used an unsupervised method of training data extraction based on voice activity detection (VAD), similar to an earlier study (Xu et al., 2020). In this approach, we used a method called robust voice activity detection (rVAD), described in detail elsewhere (Tan et al., 2020), to detect speech-free regions of noised sound clips and use the extracted sound fragments to train the filter as described above.

Application of filter

After the filter had been trained, it was applied continuously to speech-in-texture mixtures. The latter were short-term Fourier-transformed using the same parameters as the training data, including the referencing to the local temporal average over the window $T_{Multistep}$. For each time step, the distance of all training data samples to the current, brief spectrogram was then computed (Matlab function: KDTreeSearcher). The average of the $N_{Neighbor}$ closest training points was then computed as an approximation to the current noise. The resulting texture spectrogram was then subtracted from the sound mixture in the dB scale, after which the sound was reverted back to a linear scale. More specifically, the spectrogram of the current noisy speech sample $S_N(t,f)$ was embedded:

$$E(S_N(t,f)) \rightarrow \Re^{N_{FFT} x N_{WindowSteps}}(t).$$

then for each time point t find

$$S_{Neighbors} = \{\tau_i\} = \min_\tau \left| E(S_N(t,f)) - E(S_T(\tau,f)) \right|.$$

In the latter, the set of closest points of size $N_{Neighbor}$ was chosen, and then subtracted from the current spectrogram, i.e.,

$$E(S_F(t,f)) = E(S_N(t,f)) - \langle E(S_T(\tau,f)) \rangle_{\{\tau_i\}}.$$

After subtraction, the linear magnitude was transformed back into a sound pressure wave using the original phases for all frequencies (using the *idgtreal* function, Pruša et al., 2014). Further exploiting the stationarity, the estimated texture was limited to the 95th percentile of the marginal amplitudes of the training data. This approach reflects the temporally invariant composition of auditory textures by estimating the noise component using the known "repertoire" of sounds. Naturally, longer training data will improve this estimate, however, a near-plateau was already reached after only a few seconds of training data (see Figure 2J).

Performance evaluation

Pointwise correlation

This metric is computed as the correlation of the spectrograms of the filtered or noised sounds with that of the clean speech sound: the higher the correlation to the clean speech, the better the filtering at the spectrogram level. The improvement as a result of filtering is reported as a percent increase in correlation with the clean speech [i.e., $(\rho_{filtered} - \rho_{noised})/\rho_{noised}$].

DNN label error rate

Another test for speech intelligibility is the performance of automatic speech recognition software. Since this is also a potential field of usage for the filtering method, this kind of performance measurement was a sensible choice. We employed a bi-directional LSTM (Graves et al., 2006) as a speech recognition neural network, which was trained on the TIMIT dataset (see Supplementary material for details). For each timestep in the spectrotemporal domain of the sound, a phoneme probability result was calculated with 62 softmax values for the 62 possible phoneme labels (including one empty/repeat label). Using the phoneme prediction logits (input of the softmax function) of a sound file's time steps, the phoneme label was predicted using a beam-search decoder. This predicted label was used to compute the LER as its mean edit distance to the true label. Improvement in DNN label error rate was computed as a percent decrease of LER in filtered sound in comparison to the LER under noisy conditions.

Perceptual evaluation of speech quality

PESQ is a standard method for objectively measuring listening quality based on the comparison between clean reference

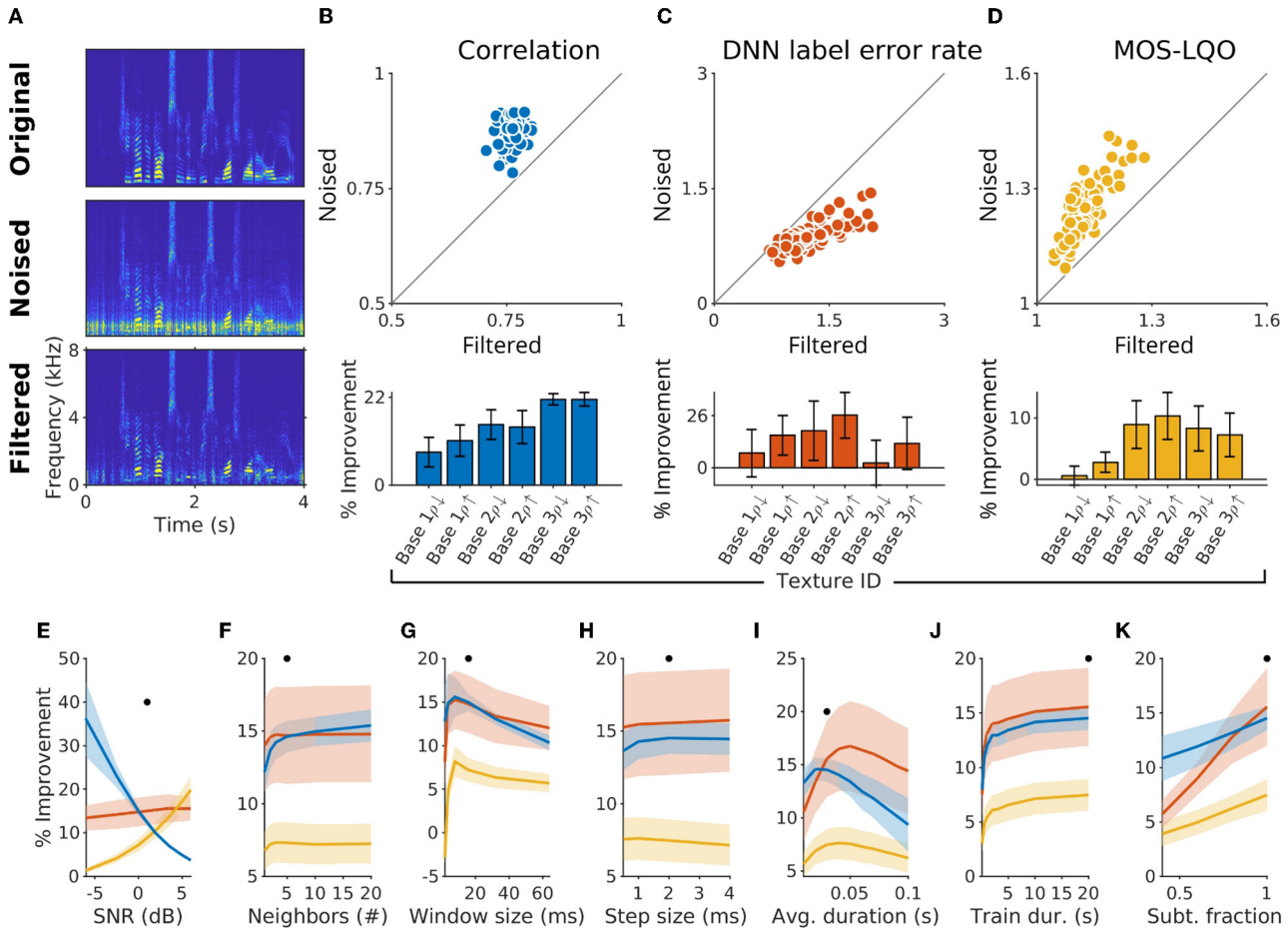


FIGURE 2
Performance of statistical filtering on speech in noise. **(A)** Sample spectrograms of original speech, speech within noise and filtered speech within noise. **(B)** Improvement via statistical filtering was first evaluated using pointwise correlation of the spectrograms (for *Base 2* $\rho\uparrow$; each dot corresponds to a sound clip; $n = 1{,}680$). In this case an average improvement of 13.9% (SD = 3.9%) was achieved, with generally positive, but varying improvements for the seven different textures used **(bottom)**. **(C)** Improvement in phoneme recognition error rate using an LSTM Deep Neural Network (DNN) trained on the TIMIT corpus. Improvement here corresponds to dots below the diagonal. Absolute values of improvement were similar in magnitude as for correlation **(bottom)**, with larger variation across the different texture types ($\mu = 10.1\%$; SD = 15.2%). **(D)** Improvement in mean opinion score—listening quality objective (MOS-LQO) estimated with the PESQ algorithm ($\mu = 6.2\%$; SD = 4.1). **(E–K)** Dependence of the improvement in speech quality on different parameters of the sounds [**(E)** signal-to-noise ratio and filtering algorithm, **(F)** number of neighbors, **(G)** window size for spectrogram creation, **(H)** step-size for moving the window, **(I)** local averaging duration in filtering the noisy speech signal, **(J)** training duration, **(K)** scaling of the estimated noise contribution, given as a fraction]. Overall, all measures improved across the range of tested parameters. We chose a value for each parameter as the default for the other parameters (black dot in each plot), which was also used in the subsequent figures.

sound and the given sound (Rix et al., 2001). In our case, comparisons to clean speech were made separately for noised speech and for filtered output of the noise reduction algorithm. The difference between the two scores obtained this way is interpreted as an objective estimate of the improvement in speech quality. In this study, we used a MATLAB wrapper function pesq_mex_vec.m provided with Sound Zone Tools (Donley, 2022), and a wideband version of the algorithm which maps raw PESQ score to MOS-LQO score for wideband sounds (ITU-T, 2007).

Comparison with Ephraim-Malah algorithm

To compare our method with an existing method we used the EM algorithm also commonly referred to as Minimum Mean Square Error-Short-Time Spectral Amplitude (MMSE-STSA) method which is a standard algorithm for single microphone noise reduction (Ephraim and Malah, 1985). It operates on short overlapping frames of the input signal in the frequency domain. By estimating the statistical properties of speech and noise, the algorithm computes a gain function that minimizes the mean square error between the estimated clean speech and the observed noisy signal. The gain function is determined based on the estimated speech presence probability in each frequency bin. The "ssubmmse.m" MATLAB routine from the VOICEBOX package (Brookes, 2002) was employed as the implementation of this algorithm. The default values were used for all user-specific parameters of the EM algorithm. See Table 2 for the full list of external software packages used in this study.

TABLE 2 External software.

Name	Version	Source	References
System: Ubuntu Linux	18.04.1	https://old-releases.ubuntu.com/releases/18.04.1	
MATLAB • System identification • Signal processing • Statistics and machine learning	R2019a/R2022b	https://nl.mathworks.com/products/matlab.html	MATLAB, 2022
Large time-frequency analysis toolbox	2.0	https://github.com/ltfat/ltfat	Pruša et al., 2014
Sound texture synthesis toolbox	1.7	https://mcdermottlab.mit.edu/Sound_Texture_Synthesis_Toolbox_v1.7.zip	McDermott and Simoncelli, 2011
NeurAudio statistical filtering toolbox		https://data.donders.ru.nl/collections/di/dcn/DSC_626840_0011_433	(This article)
Robust voice activity detection (rVAD)	2.0	https://github.com/zhenghuatan/rVAD	Tan et al., 2020
Sound zone tools	1.0.0	https://github.com/jdonley/SoundZone_Tools	Donley, 2022
Packages in the conda environment			
Python	3.7.7	https://www.python.org/downloads/release/python-377	van Rossum and Drake, 2009
Numpy	1.18.1	https://pypi.org/project/numpy/1.18.1	Harris et al., 2020
Scipy	1.4.1	https://docs.scipy.org/doc/scipy-1.4.1/reference/index.html	Virtanen et al., 2020
Scikit-learn	0.23.1	https://scikit-learn.org/0.23	Pedregosa et al., 2011
Tensorflow-gpu	1.14	https://www.tensorflow.org/install/pip	Abadi et al., 2016
Tensorpack	0.10.1	https://pypi.org/project/tensorpack	Wu, 2016
Cudatoolkit	10.0.130	https://anaconda.org/anaconda/cudatoolkit/files?version=10.0.130	n/a
cuDNN	7.6.5	https://developer.nvidia.com/rdp/cudnn-archive	Chetlur et al., 2014
Bob.ap	2.1.10	https://www.idiap.ch/software/bob/docs/bob/bob.ap/v2.1.10	Anjos et al., 2012
Editdistance	0.5.3	https://pypi.org/project/editdistance/0.5.3	Tanaka, 2019
Matplotlib	3.1.3	https://matplotlib.org/3.1.3/contents.html	Hunter, 2007
VOICEBOX		http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html	Brookes, 2002

Computational complexity

In order to evaluate the speed at which the algorithm can be run, we used a Desktop PC (AMD Threadripper 2920X, 12 cores) running the algorithm on 100 speech samples randomly selected from TIMIT dataset mixed with all seven texture types. To quantify the speed, the average time spent on running the processing of a single frame was divided by the step size (default value = 2 ms) to obtain a real-time factor. The average time spent on a single frame is estimated by estimating the time it takes to filter a given speech sample divided by the number of frames in the spectrogram, which is determined by the window (16 ms) and step size (2 ms) parameters. Real-time factors <1 indicate that the algorithm can run faster than real-time on our setup (see Figure 6).

Human experiments

To evaluate the human-perceivable change in sound quality, we performed an online experiment where we asked participants to rate speech and background components of the delivered sound. The experiment lasted on average about 1 h and 10 participants (six male, four female, average age: 33.6 y, SD = 7.6 y) took part in the study. Participants were recruited through Prolific (www.prolific.co), where we chose to recruit individuals with no hearing difficulties, hearing aids or cochlear implants, and those who spoke English as their first language. Experimental code was generated using PsychoPy3 Builder (Peirce et al., 2019) and hosted on Pavlovia (pavlovia.org). All participants gave written informed consent to take part in the experiment, which was approved by the Ethics Committee of the Faculty of Social Sciences at Radboud University Nijmegen.

To ensure that the participants using a variety of different hardware could hear the sounds in a comparable manner, and to check that they were using headphones as instructed, we started the session with a headphone screening test described in detail elsewhere (Woods et al., 2017). In this section, participants were asked to report which of the three pure tones was quietest,

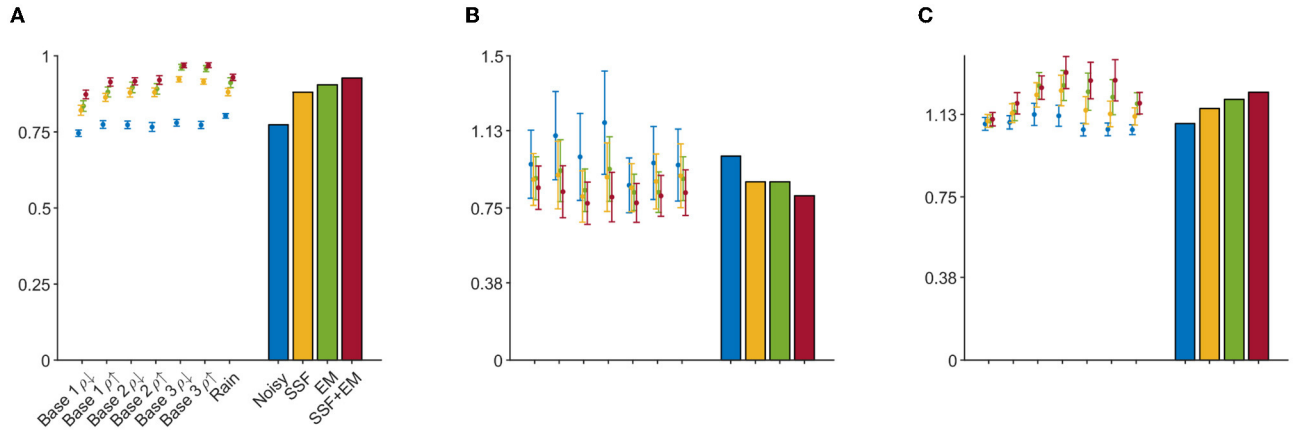


FIGURE 3
Comparison of individual and combined performances of SSF and EM. Combination of SSF and EM outperforms the separate filtering runs for correlation **(A)**, DNN label error rate **(B)**, and MOS-LQO score **(C)** for all textures and speech samples used in this study ($n = 7$ textures; $n = 1{,}680$ speech samples). Bar plots represent the mean performance for each filtering condition across all textures and speech samples ($n = 11{,}760$) and all differences between these means were highly significant ($p < \epsilon$), except for the DNN label error rate difference between SSF and EM ($p < 0.01$). Error bars here represent the SD, as SEMs are visually indistinguishable from 0 (due to the high number of speech samples).

with one of the tones presented 180° out of phase across the stereo channels. The task is trivial with headphones but gets harder to perform without headphones due to phase cancellation. Nine out of 10 participants were able to perform this task with 100% accuracy (n trials = 12). The outlier was included in the analysis due to the similarity of the behavioral results to other participants, suggesting that this individual was still engaged in the task. Participants were financially compensated for their time once the experiment had finished; no additional motivation was provided.

Experimental trials started with the presentation of a sound clip. After the sound played, a new screen with continuous vertical scales for speech and background ratings was shown (see Figure 3A). Speech rating scale ran from 1 (distorted) to 5 (clear), while the background rating scale ran in an analogous fashion from 1 (very quiet) to 5 (very loud). The participant could report their evaluation by clicking and adjusting the indicator point on the scale with a mouse. We tested a total of 518 sound clips which included filtered and noised versions of the same speech fragments mixed with different types of background noise. The order of sound delivery was randomized to avoid direct comparison of filtered and unfiltered versions of the same speech sample.

Results

We designed and implemented a fast noise-filtering algorithm (Statistical Sound Filter) SSF focussed on textural stimuli, characterized by time-independent statistics, and evaluated its performance on the TIMIT speech dataset corrupted by a set of semi-natural and natural background noises (Figure 1). Evaluation included both automated and human assessments of speech quality as well as an evaluation of SSF's run-time as a function of its various parameters.

Approach to filtering statistically governed sounds

A large fraction of naturally occurring background sounds can be characterized as acoustic textures, i.e., they are a composite of basic sounds whose temporal occurrence is only constrained statistically, with the additional limitation that these statistics are stationary/time-invariant (Figure 1A). Examples of auditory textures include the sounds of wind, waves, rain, fire, insect swarms, flocks of birds, and essentially all sounds where many similar entities produce similar sounds. We developed a noise filter that specifically approaches the removal of these auditory textures from target/foreground sounds, termed Statistical Sounds Filter (SSF). SSF is first trained on a section of acoustic texture (see Figure 4 for training on mixed speech-texture sounds), by assembling a library of spectrotemporal sounds from the training data. This includes both individual and composite occurrences of the constituent sounds (Figure 1B, top and bottom). SSF is then applied to speech embedded on a different realization/sample of the texture (Figure 1C), which would occur after the training data in real life. SSF then matches the feature library against a preprocessed version of the speech-in-noise sample for all time points, subtracts out the best match, and then recreates the filtered sound using the original phases (Figure 1D). The resulting sounds exhibit substantially and significantly reduced background noises and thus separation of the target sound, e.g., speech in the present testing.

Filtering performance as a function of algorithm parameters

To evaluate the performance of SSF we generated an array of artificial textures based on real-world sound textures (Figure 1A). These textures provided the benefit of being based on real sounds while at the same time allowing us to manipulate the background

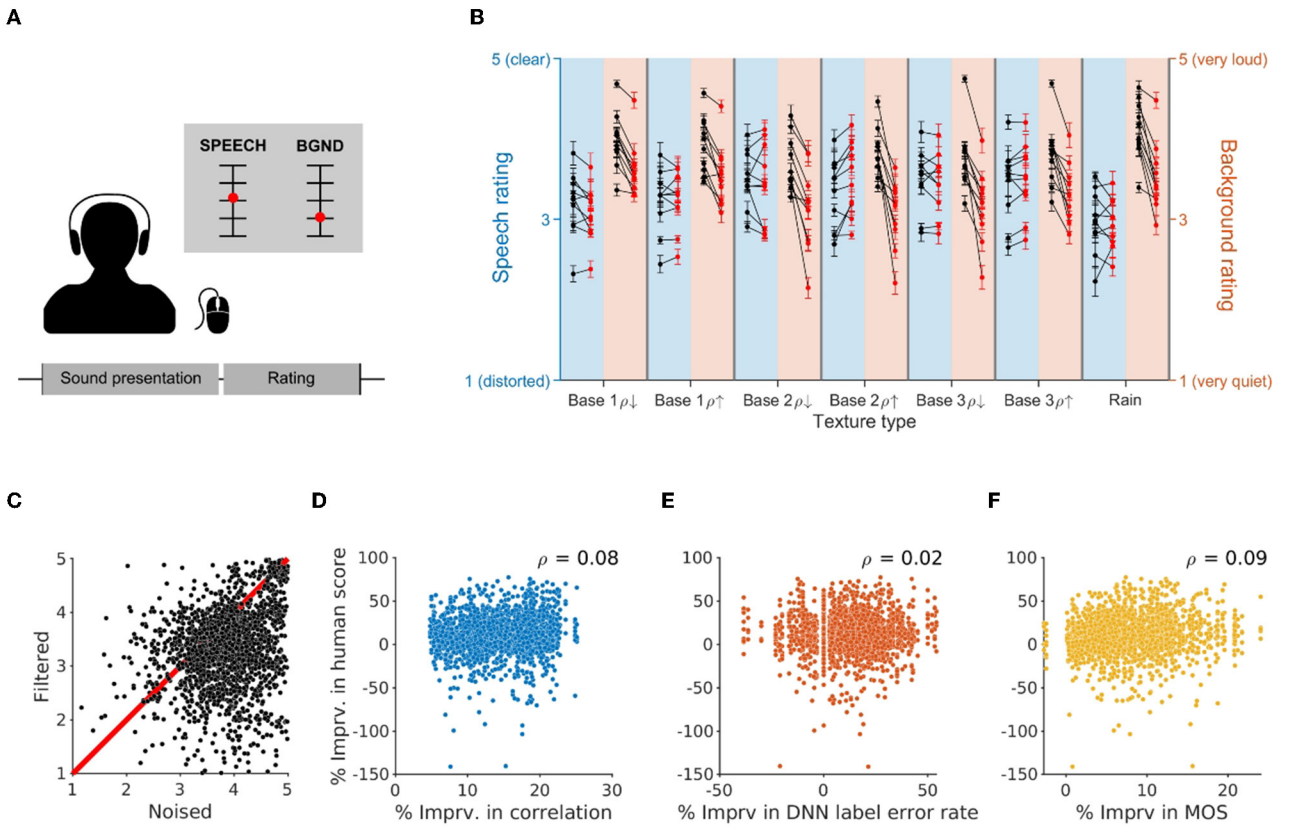


FIGURE 4
Human evaluation of filtering performance. **(A)** Experimental setup of the human experiment. At every trial, participants were asked to rate the speech and background components after the sound clip was delivered through headphones. The speech was rated on a scale from 1 (distorted) to 5 (clear), while the background was rated on a scale from 1 (very quiet) to 5 (very loud). **(B)** Evaluation of speech (blue) and background (orange) components of the presented sounds across different background noise conditions. Black dots represent mean ratings for noised sounds, while red dots stand for mean ratings for filtered sounds. Each line connecting the dots corresponds to an individual participant ($n = 10$). While speech ratings do not show any systematic change as a result of filtering, background ratings decrease, indicating consistent removal of background noise. **(C)** Background ratings provided for filtered and noised versions of the same sound clips, pooled across participants. Each dot represents two ratings provided in two separate, non-consecutive trials ($n = 2{,}010$ pairs). **(D–F)** Comparison of filtering-related change in human evaluation with the change in objective measures—correlation **(D)**, DNN label error rate **(E)**, and MOS **(F)**—reveals a very small correlation between objective and subjective evaluation metrics, suggesting that human evaluation used in this study reflects features that are not captured by these objective measures.

sound statistics parametrically to evaluate the influence on SSF's performance. The generated textures varied in their marginal and across-frequency correlation statistics and covered a large range of values in the respective parameter spaces, both spectrally and temporally (see Methods for details). We also included a natural texture (sound of rain) to exclude that the SSF's performance is limited to the peculiarities of artificially synthesized textures.

Initial testing was carried out on a set of parameters that allowed relatively fast and effective filtering of the background noise (Figure 2A). For this example run, SNR = 0 dB was used for mixing speech and noise. The performance was then quantified using three measures, (i) the spectrogram correlation, (ii) the label error rate of a DNN, and (iii) the MOS-LQO score.

The correlation coefficient was computed by taking a pointwise correlation between the spectrogram of the original (clean) sound clip and that of filtered (or noised) versions of the same speech fragment (Figure 2B). Across all speech and noise combinations, SSF achieved an average of 13.9% (SD = 3.9%, $p < 0.001$, Wilcoxon rank sum test) improvement in correlation metric. Statistical comparisons were carried out across the different texture types ($N = 7$) between speech-in-noise and filtered averages. Within texture type, almost all showed highly significant improvements (Table 3), which is unsurprising given the large number of samples tested ($N = 11{,}760$ total samples).

To evaluate the change in speech intelligibility automatically, we utilized a DNN-based phoneme recognition system trained on the TIMIT dataset (see Methods) and quantified the relative labeling error rate before and after filtering (Figure 2C, $\mu = 10.1\%$; SD = 15.2%, $p = 0.011$).

Lastly, we used a commonly used wideband PESQ algorithm to evaluate the speech quality and reported the results by transforming the raw PESQ score to mean opinion score—listening quality, as described in ITU-T P.862.1 (ITU-T, 2003) (Figure 2D, $\mu = 6.2\%$; SD = 4.1%, $p = 0.007$).

Next, we varied the main parameters of the algorithm to understand how each affects the performance, as measured by the above metrics. While for most parameters the performance changed in a comparable manner across different metrics

TABLE 3 Summary of improvement in objective evaluation metrics per texture for default parameters.

TextureID	Correlation			DNN LER			MOS-LQO		
	μ	Σ	p	μ	σ	p	μ	σ	p
$Base_1$ $\rho\uparrow$	10.154	2.357	$< \epsilon^*$	6.626	12.324	$< \epsilon$	1.277	1.591	$< \epsilon$
$Base_1$ $\rho\downarrow$	11.536	2.062	$< \epsilon$	16.481	12.146	$< \epsilon$	3.881	2.343	$< \epsilon$
$Base_2$ $\rho\uparrow$	13.758	2.995	$< \epsilon$	17.943	12.629	$< \epsilon$	8.012	3.224	$< \epsilon$
$Base_2$ $\rho\downarrow$	14.932	2.814	$< \epsilon$	21.643	11.865	$< \epsilon$	10.331	3.615	$< \epsilon$
$Base_3$ $\rho\uparrow$	18.340	1.160	$< \epsilon$	0.256	13.433	0.037	8.345	3.484	$< \epsilon$
$Base_3$ $\rho\downarrow$	18.493	1.218	$< \epsilon$	7.974	13.814	$< \epsilon$	6.645	3.179	$< \epsilon$
Rain	10.277	1.756	$< \epsilon$	−0.159	14.316	0.286	4.953	2.238	$< \epsilon$

*ϵ here is 10^{-15}.

(Figures 2F–K), varying the signal-to-noise ratio (SNR) affected our metrics in a clearly divergent manner (Figure 2E). At very low SNRs, our algorithm does not significantly improve the objective listening quality, but it is able to effectively improve the spectral representation of the speech as measured by pointwise correlations. Such divergent effects of SNR on the present performance metrics highlight the need for evaluation using multiple metrics that quantify separate aspects of the sound.

To determine if our method leverages a unique statistical aspect of the background noise, we compared its performance with an established method that uses the mean-square error short-time spectral amplitude (MMSE-STSA) estimator for enhancing noisy speech (Ephraim and Malah, 1985). The MMSE-STSA method uses a statistical model of the speech and noise spectra, and computes the gain function that minimizes the mean-square error between the estimated and true spectral amplitudes. Although the EM outperforms SSF when used alone at the SNR used in this filtering run (0 dB), combining it with our method (SSF → EM) significantly enhances performance across all metrics (Figures 3A–C). Since SSF has a fast processing time (as shown below), our results suggest that our method can effectively complement other standard methods to further reduce noise without adding excessive computational overhead.

Human listeners indicate consistent suppression of background noise

To get a more explicit evaluation of human-perceivable improvement as a result of our filtering algorithm we ran an online experiment with human participants ($n = 10$). Given the time limitation that comes with human experiments, we selected a representative subset of speech fragments from the TIMIT corpus with balanced features such as speaker gender, identity, dialect, and sentence type (Garofolo et al., 1993), and mixed the selected speech fragments with the aforementioned texture types (see Methods). At each trial, the participant was asked to rate speech and background components of the sound using separate linear scales running from 1 to 5 (Figure 4A). For the speech component, participants rated the quality of sound clips on a continuous scale from distorted (1) to clear (5). For background evaluation, participants reported their judgments on a scale from very quiet (1) to very loud (5).

Comparing individual ratings divided across texture types, we observed no significant change in speech ratings as a result of our filtering procedure ($\mu = -6.02\%$, SD = 7.35%, $p = 0.32$, Wilcoxon rank sum test, $n = 7$ filtered/noised pairs; Figure 4B). The ratings were also not significantly different for most texture types when the ratings were analyzed separately for each texture (Table 4). However, participants perceived the background component as consistently less loud after filtering ($\mu = -15.8\%$, SD = 3.47%, $p < 0.001$, Wilcoxon rank sum test, $n = 7$ filtered/noised pairs; Figure 4B).

To better visualize the variability in background ratings, we compared the matched ratings of sound clips with the same speech and noise components (Figure 4C). An additional source of variability here likely arises from the fact that the order of trials (and hence sound clips) in our experiment was completely randomized, preventing the participants from directly comparing filtered and noised versions of the same sound clip. This was done to avoid peculiarities of a given speech sample from affecting the evaluation and to encourage independent judgment of each sound sample.

Next, we compared the human evaluation to objective metrics described in the earlier section. To do this, we tallied the percent improvement in the human judgment of the background level to those computed by pointwise correlation, DNN label error rate, and MOS-LQO (Figure 4D). Even though our algorithm on average improves all four metrics, correlation coefficients across these measures of performance were very low, confirming that they capture different features of the sound than those evaluated by the human listeners.

Within-sample training achieved comparable performance to dedicated training data

Above, we trained the algorithm on a single, defined section of textural sound to standardize the algorithms library across samples. However, in real-life situations, such training data is not necessarily available. To improve the range of use cases for our method, we utilized an alternative, unsupervised training method that relies on voice activity detection (VAD). Briefly, VAD detects sections of the sound where human-voiced sounds are present. Focussing on the

TABLE 4 Summary of human evaluation results for each texture type.

TextureID	Speech [1 (distorted) → 5 (clear)]					Background [1 (very quiet) → 5 (very loud)]				
	μ (noised)	σ (noised)	μ (filtered)	σ (filtered)	*p*	μ (noised)	σ (noised)	μ (filtered)	σ (filtered)	*p*
$Base_1$ ρ ↑	3.197	0.867	3.079	0.903	0.053	3.979	0.569	3.623	0.631	$< \epsilon$
$Base_1$ ρ ↓	3.229	0.941	3.232	0.864	0.892	3.920	0.555	3.546	0.632	$< \epsilon$
$Base_2$ ρ ↑	3.507	0.827	3.460	0.865	0.436	3.684	0.679	3.128	0.815	$< \epsilon$
$Base_2$ ρ ↓	3.309	0.896	3.500	0.851	0.003	3.827	0.591	3.069	0.755	$< \epsilon$
$Base_3$ ρ ↑	3.497	0.864	3.457	0.899	0.522	3.765	0.666	3.156	0.772	$< \epsilon$
$Base_3$ ρ ↓	3.455	0.863	3.502	0.861	0.451	3.796	0.652	3.322	0.715	$< \epsilon$
Rain	2.961	0.897	2.945	0.882	0.793	4.121	0.623	3.521	0.701	$< \epsilon$

complement, i.e., sections that likely do not contain human voice, we create a within the sample training set, which we use to train SSF (Figure 5A). Considering the fact that the effect of training duration on the performance of our algorithm plateaus very fast (Figure 2J), we hypothesized that existing VAD methods should be able to extract sufficient amounts of training data from the gaps between bouts of speech in our sound clips. Consistently, we found that even though the performance of the algorithm was slightly reduced in comparison to the supervised training, the overall pattern of the results remained similar. Correlation with the clean speech improved on average by 14.8% (SD = 4%, $p < 0.001$, Wilcoxon rank sum test; Figure 5B), DNN label error rate was reduced by 13.7% (SD = 13.6%, $p < 0.01$, Wilcoxon rank sum test; Figure 5C), while MOS-LQO had an average of 7.66% improvement (SD = 4.48%, $p < 0.01$, Wilcoxon rank sum test; Figure 5D). Because VAD-based training is agnostic to the source of noise, we expect it to be better utilized in settings where noise is not stationary and cannot be obtained separately, such as cases where live filtering is required. Another option in a real scenario would be that the user selects certain time periods for rapidly (re)training SSF, instead of using an automatic selection.

Statistical filtering performs much faster than real-time

The speed with which an algorithm can be run is another factor determining the range of its use cases. We quantified the speed of execution on a desktop computer (AMD Threadripper 2920X, 12-core). As with other performance metrics, we varied the core parameters of the algorithm to get a detailed overview of the runtime of our algorithm (Figure 6). Runtime was quantified as the time it takes to process one frame of the sound spectrogram divided by the actual duration of that frame, referred to as the *real-time factor*, with values <1 indicating faster than real-time processing. With the default set of parameters, where each frame was 16 ms and the step size was 2 ms long, we obtained a real-time factor of $\mu = 0.0154$ (STD = 0.0011), i.e., ∼65× faster than real-time. The variables that had the strongest influence on processing speed were window and step sizes, as well as training duration. Given that the effect of these variables on performance metrics reported above plateaus very fast, the parameters can be tuned to run the algorithm extremely fast and effectively without compromising the filtering accuracy.

Discussion

We developed a dedicated method for noise reduction in the context of acoustic textures, exploiting their statistically stable composition from a limited set of constituent sounds. The algorithm represents the set of spectrotemporal features of the background texture and subtracts a pseudo-optimal match from the speech-in-noise mixture. Testing the algorithm on a set of semi-natural and natural textures, we found that the algorithm can effectively remove textural noise in a fast, efficient manner that leads to a perceptual improvement in human listeners.

Instantaneous statistics vs. a full statistical model

Given the statistically stable composition of textures, the most obvious choice for a filtering approach appears to be training a suitably designed statistical model of the texture, e.g., based on McDermott et al. (2013). We have experimented with both this model and Gaussian processes, however, we concluded that this approach was unsuitable for filtering for two main reasons: (1) Training and synthesis in these models are computationally intense and require a lot of data to constrain the models. These two aspects make them currently incompatible with the requirement of live processing, ideally on hearing aids. (2) If one wants to exploit the additive nature of the background noise and the target sound, a natural approach would be to synthesize future samples of the background noise, and subtract these from the composite sound. However, while these samples are individually statistically consistent with the background noise, they are not related to the current realization of it. Choosing a best match would thus require sampling a large variety of future samples and then subtracting the best match from the current sample, or projecting the latter onto the statistical model to separate noise and target sound. In our hands, neither of these approaches was fast enough to improve the quality of the target sound relative to the

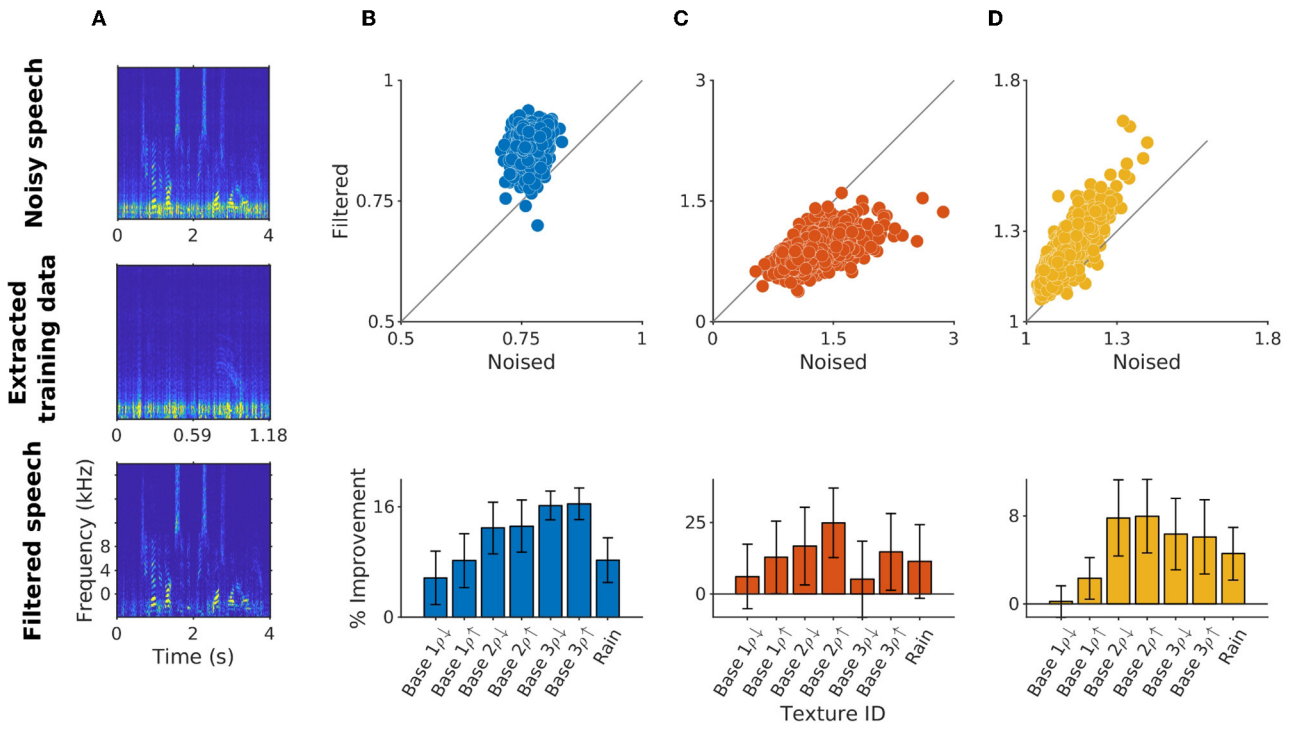


FIGURE 5
Training on automatically detected noise segments. **(A)** Sample spectrogram of original noisy speech and training data extracted using rVAD. The evaluation of the unsupervised version of the algorithm using correlation **(B)**, DNN label error rate **(C)** and MOS-LQO scores **(D)** exhibits an improvement pattern similar to the supervised version shown in Figure 2.

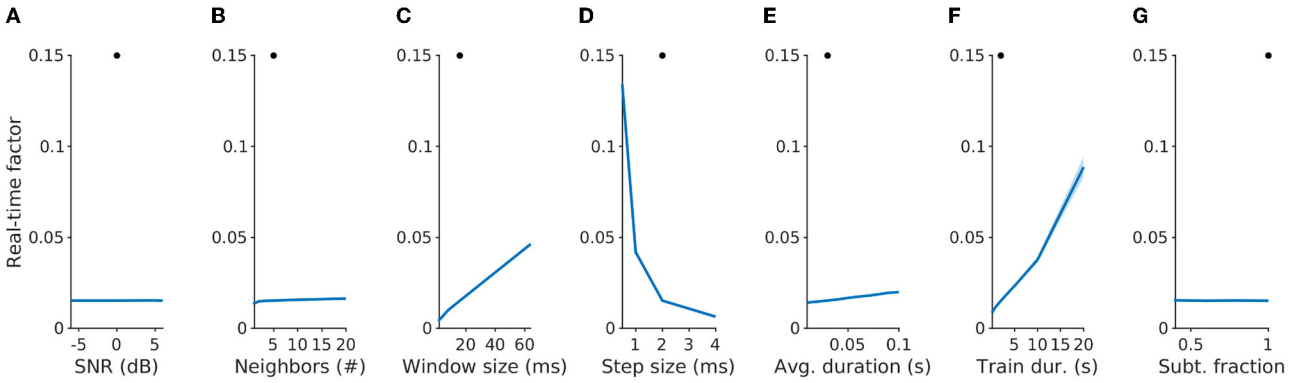


FIGURE 6
Statistical filtering can run in super-real-time on a current mobile phone. The default parameters for the present comparison were chosen on the basis of the filtering performance on speech in noise (see Figure 2). Filtering was run on a Threadripper 2920X processor, which is comparable in speed to a modern mobile processor (e.g., Apple M1, comparison based on Geekbench 5 performance). Performance is given in multiples of real-time (*Real-time factor*), i.e., smaller numbers indicating faster processing. Varying the parameters indicated that WindowSize **(C)**, Step-Size **(D)**, and training duration **(F)** have the largest influence on run-time. At the default parameters (black dots), the real-time factor is on average 0.0154 (S.D. = 0.0011). The current code is running non-compiled, hence, there remains room for optimization that would likely lead to substantial acceleration.

noise in a timely manner, however, see Liutkus et al. (2011) for a potential approach.

Conversely, the present approach is directly based on the expressed noise occurrences and utilizes them as a lexicon to compare against. While this approach is less general, it has the advantage of computational and data efficiency. In the future, we would like to combine a low-dimensional Gaussian process (GP) approach with SSF, since we think this might in combination remain computationally feasible and augment the performance of SSF by the inclusion of slower or marginal features that are potentially missed by SSF.

Relation of SSF filtering to human textural filtering

The interest in sound textures arises from the observation that they constitute a sizable fraction of naturally occurring sounds (Liu et al., 2022) and that humans can recognize and distinguish sounds based on their textural statistics alone (McDermott et al., 2013). This is reflected by the fact that sound textures can be represented and synthesized very effectively using a restricted number of summary statistics (McDermott and Simoncelli, 2011). The existence of such a compact representation raises an interesting possibility that the auditory system itself may utilize analogous compact representations and predictively suppress textural noise (Rabinowitz and King, 2011). The evidence from a perceptual discrimination task suggests that the human auditory system increasingly converges on time-averaged statistics of textures, instead of representing the temporal details of the individual acoustic events (McDermott et al., 2013).

As we mentioned in the previous section, even though time-averaged statistics carry sufficient information to resynthesize sound texture samples, an effective reduction of noise requires precisely matching the noise on a moment-by-moment basis. The statistical filtering that appears to be realized in the auditory cortex (Mesgarani et al., 2009; Khalighinejad et al., 2019) may achieve this by suitably transforming the sound, potentially using a cascade similar to the one proposed by McDermott et al. (2013). and then adapting on every level, as an extension to the principle used in Boubenec et al. (2017). The downside of this transformation is that it is not (easily) invertible and thus cannot be used to synthesize a sound from the filtered representation, which is essential in applications such as hearing aids. On the other hand, it might be sufficient for processing in speech recognition systems. The present method of matching samples of the sounds against a library of known spectrotemporal features of the texture is thus likely not reflective of neural processes but may be productively combined with them.

Generalizability of SSF to other classes of noises

One of the core issues facing any noise reduction algorithm is the generalization to other sounds. To assess how well SSF generalizes across different types of textures, we parametrically controlled the statistical features of the background noises on which the performance of the algorithm was tested, in addition to the inclusion of a natural texture. Among the tested background sounds we observed variability in algorithm performance, but an improvement was observed for all textures with different marginal and correlation statistics. These results suggest that SSF can generalize across a wide range of sound textures, though further studies utilizing the full set of statistics in the synthesis of sound textures can improve the granularity of the sampling.

Beyond this, an additional challenge can be that the definition of noise can be context-dependent (Liu et al., 2022): what may act as noise in one condition may carry information in another context. For instance, speech sounds, which are typically enhanced and considered as signals, are notoriously difficult to reduce when mixed together in a cocktail party situation (Middlebrooks et al., 2017). To address this issue, noise (or features thereof) can be defined in a supervised, user-driven manner or deduced using cues such as head direction, lip movements, etc. (Michelsanti et al., 2021). Babble noise associated with cocktail party situations was not included in our dataset due to our approach to sampling the noise space by texture synthesis, which is not conducive to synthesizing highly modulated speech sounds. Future applications of SSF would therefore require further testing of the algorithm with babble noise which can become more texture-like with a growing number of talkers.

In addition, the variability arising from dynamic changes in the background noise condition of the given scene poses another major challenge for noise reduction algorithms. More specifically, methods meant to be used in real-time situations should be able to adaptively reduce noise from sources that enter or leave the acoustic scene. VAD-based noise extraction can in principle address this problem by allowing continuous training data extraction-training-filtering cycles. The performance of the algorithm was found to plateau quickly (∼2–5 s, see Figure 2) as a function of adding longer training data sets, suggesting that the algorithm can plausibly be used in a real-time setting for continuous training. However, the time course of background noises varies greatly from continuous textures to impulse noise that happens very fast and poses a challenge to SSF which assumes some level of stationarity in the background conditions.

Comparison with other filtering techniques

By design, SSF is agnostic to the type of target sounds embedded in the noise. While this property imparts SSF its domain-neutrality, and a broader range of applicability, it also limits the improvement to the speech intelligibility, when such a use-case is desired. This is a common problem with noise reduction algorithms that aim to model the features of noise and subtractively remove them from the sound mixture. Previous studies showed that while such algorithms can decrease the listening effort, they do not necessarily improve speech quality or intelligibility at the same time (Sarampalis et al., 2009; Fiedler et al., 2021). In this study, speech quality rather than intelligibility was quantified in human experiments, and the metric that most closely approximated speech intelligibility was the DNN label error rate. We observed that the quality of speech was not degraded and the DNN label error rate decreased as a result of filtering. However, further experiments are needed to quantify the effect that SSF may have on speech intelligibility for human listeners.

Recent years have seen a lot of development of speech-denoising techniques based on machine learning methods, primarily artificial neural networks, in particular deep neural

networks (DNNs) (Michelsanti et al., 2021; Ochieng, 2022). These methods have been demonstrated to be highly effective in tackling speech-in-noise problems, partly because they can be trained to have a highly complex representation of speech which may enable them to selectively enhance speech. The present approach is more simplistic in nature, targeting the specific properties of sound textures. We think it has three concrete advantages over complex DNN systems:

(i) Rapid, targeted training: DNN systems require a substantial amount of time and resources to be trained. From the perspective of a hearing aid user, it might often be preferable to have an algorithm (such as SSF) available that can be quickly retrained to adapt to the current background sound, and thus specifically reduce disturbances from this source. Training using SSF requires only a few seconds of training data, and training completes closely after all training data has been processed (∼60 ms for 2 s of training data). As we have shown, supervised (Figure 2) and unsupervised training methods (Figure 4) can achieve similar levels of performance.

(ii) Domain-neutral: DNN systems are typically trained on a large set of speech sounds in the context of a certain set of noise sounds. This enables these systems to make use of the inherent predictability of speech in addition to the structures in the noise. In SSF, the regularities inside the target sound are not utilized in the filtering. While this likely limits the quality of filtering on the training set, it may generalize better to other target sounds, e.g., music or other sounds that are not consistent with the textural statistics.

(iii) Fast execution: SSF runs much faster (probably 10–100×) than real-time on the type of processors found in current mobile devices (e.g., multicore performance of the present desktop processor is only a factor 2 greater than an Apple A16 processor; Geekbench, 2022). It, hence, does not require a powerful GPU to run efficiently. This enables usage cases, where either the hearing aid processor in the hearing aid or a connected mobile phone runs the filtering in near real-time. While SSF is thus computationally lighter than DNN approaches, running it directly on a hearing aid may require the design of a specialized processor to stay within typical power limits and runtimes (Dr. Harzcos, audifon, personal communication).

Methodological limitations

The low computational complexity of spectral subtraction methods comes at a price of distortions that may arise from inaccuracies in noise estimation. Such distortions affect the speech as well as the noise components, creating a phenomenon known as musical noise (Loizou, 2013a), which is characterized by small, isolated peaks in the spectrum occurring randomly in the frequency bands at each time frame. A number of methods have been proposed in order to directly address musical noise (Goh et al., 1998; Lu and Loizou, 2008; Miyazaki et al., 2012). Although spectrally flooring negative values generated by subtraction to minimum values in adjacent frames (as was done in Boll, 1979) led to a small improvement in the MOS-LQO score, it significantly reduced the performance in other performance metrics. While our approach does not directly address the problem of musical noise, it indirectly reduces the overall likelihood of its occurrence by modeling the noise source specifically.

While the chosen DNN architecture for assessing the improvement of speech intelligibility was well-motivated, alternative approaches could have some additional value. Since real-world applications in speech recognition would choose more recent architectures (see Li, 2022 for a review), using such a system might provide estimates that are more in line with the human perceived evaluations and also translate better to current applications in speech recognition.

Lastly, our human experiments show that the participants do not perceive a reduction in speech quality as a result of filtering, suggesting that the speech component is not substantially distorted as a result of subtraction. However, since our test did not directly ask the participants to indicate the perceived background quality (only the level was asked), we cannot rule out the possibility of residual musical noise. The development of automatic methods for quantifying the amount of musical noise can therefore improve the evaluation of spectral-subtractive methods in the future.

Conclusions and future steps

We presented an efficient and dedicated spectral subtraction-based method for noise reduction in sound textures. The way of representing and estimating background noises was inspired by the fundamental feature of sound textures which are made up of spectrally similar sound events that tend to persist in the acoustic scene. We show that spectral subtraction performed based on the KNN search can effectively reduce this kind of noise, without causing significant distortion to the speech. Additionally, the algorithm runs much faster than real-time on conventional computing machines, suggesting that it can be integrated into devices that have limited computational power such as hearing aids. The speed of the algorithm also allows it to be potentially used in conjunction with other methods that can enhance the speech component and reduce the residual musical noise. Given that sound textures constitute a substantial subset of what is considered noise in human hearing, we believe closer attention to this class of sounds in development and testing may aid other noise reduction algorithms in the future in terms of generalizability.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://doi.org/10.34973/j6k7-j590.

Ethics statement

The studies involving humans were approved by Ethics Committee of the Faculty of Social Sciences at Radboud University Nijmegen. The studies were conducted in accordance with the

local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

AA, NS, and BE contributed to the conception and design of the study and wrote the first draft of the manuscript. NS and BE developed the first versions of the algorithm. AA performed the testing, fine-tuning, and statistical analysis. All authors contributed to the manuscript revision, read, and approved the submitted version.

Funding

BE acknowledges funding from an NWO VIDI Grant (016.VIDI.189.052) and an NWO Open Grant (ALWOP.346).

Acknowledgments

We would like to thank Dr. T. Harzcos (audifon Germany) for helpful discussions on the limitations of hearing aids.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fauot.2023.1226946/full#supplementary-material

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., et al. (2016). TensorFlow: large-scale machine learning on heterogeneous distributed systems. *arXiv [Preprint].arXiv:1603.04467*. doi: 10.48550/arXiv.1603.04467

Anjos, A., El-Shafey, L., Wallace, R., Günther, M., McCool, C., and Marcel, S. (2012). "Bob: a free signal processing and machine learning toolbox for researchers," in *Proceedings of the 20th ACM International Conference on Multimedia* (New York, NY: Association for Computing Machinery), 1449–1452.

Boll, S. (1979). Suppression of acoustic noise in speech using spectral subtraction. *IEEE Trans. Acoust. Speech Signal Process.* 27, 113–120. doi: 10.1109/TASSP.1979.1163209

Boubenec, Y., Lawlor, J., Górska, U., Shamma, S., and Englitz, B. (2017). Detecting changes in dynamic and complex acoustic environments. *eLife* 6, e24910. doi: 10.7554/eLife.24910.024

Braun, S., Gamper, H., Reddy, C. K. A., and Tashev, I. (2021). "*Towards Efficient Models for Real-Time Deep Noise Suppression*," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (Toronto, ON), 656–660.

Brookes, M. (2002). *VOICEBOX: Speech Processing Toolbox for MATLAB*. Available online at: http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html (accessed October 7, 2023).

Chetlur, S., Woolley, C., Vandermersch, P., Cohen, J., Tran, J., Catanzaro, B., et al. (2014). cuDNN: efficient primitives for deep learning. *arXiv [Preprint].arXiv: 1410.0759*. doi: 10.48550/arXiv.1410.0759

Culling, J. F., and Stone, M. A. (2017). "Energetic masking and masking release," in *The Auditory System at the Cocktail Party, Springer Handbook of Auditory Research*, eds J. C. Middlebrooks, J. Z. Simon, A. N. Popper, and R. R. Fay (Cham: Springer International Publishing), 41–73.

Donley, J. (2022). *Sound Zone Tools [MATLAB]*. Github package. Available online at: https://github.com/jdonley/SoundZone_Tools (Original work published 2015).

Ephraim, Y., and Malah, D. (1985). Speech enhancement using a minimum mean-square error log-spectral amplitude estimator. *IEEE Trans. Acoust. Speech Signal Process.* 33, 443–445. doi: 10.1109/TASSP.1985.1164550

Fiedler, L., Seifi Ala, T., Graversen, C., Alickovic, E., Lunner, T., and Wendt, D. (2021). Hearing aid noise reduction lowers the sustained listening effort during continuous speech in noise-A combined pupillometry and EEG study. *Ear Hear.* 42, 1590–1601. doi: 10.1097/AUD.0000000000001050

Garofolo, J. S., Lamel, L. F., Fisher, W. M., Fiscus, J. G., and Pallett, D. S. (1993). *DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus CD-ROM. NIST Speech Disc 1-1.1. NASA STI/Recon Technical Report No. 93.* Speech Corpus.

Geekbench (2022). *iPhone 14 Pro Max vs Generic - Geekbench Browser*. Available online at: https://browser.geekbench.com/v5/cpu/compare/19645606?baseline=19423140 (accessed February 1, 2023).

Goh, Z., Tan, K.-C., and Tan, T. G. (1998). Postprocessing method for suppressing musical noise generated by spectral subtraction. *IEEE Trans. Speech Audio Process.* 6, 287–292. doi: 10.1109/89.668822

Graves, A., Fernández, S., Gomez, F., and Schmidhuber, J. (2006). "Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks," in *Proceedings of the 23rd International Conference on Machine Learning (ICML '06)* (New York, NY: Association for Computing Machinery), 369–376.

Handel, S. (2006). "The transition between noise (disorder) and structure (order)," in *Perceptual Coherence: Hearing and Seeing*, ed S. Handel (Oxford: Oxford University Press).

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., et al. (2020). Array programming with NumPy. *Nature* 585, 357–362. doi: 10.1038/s41586-020-2649-2

Henry, F., Glavin, M., and Jones, E. (2021). Noise reduction in cochlear implant signal processing: a review and recent developments. *IEEE Rev. Biomed. Eng.* 16, 319–331. doi: 10.1109/RBME.2021.3095428

Huang, Y. A., and Benesty, J. (2012). A multi-frame approach to the frequency-domain single-channel noise reduction problem. *IEEE Trans. Audio Speech Lang. Process.* 20, 1256–1269. doi: 10.1109/TASL.2011.2174226

Hunter, J. D. (2007). Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.* 9, 90–95. doi: 10.1109/M.C.S.E.2007.55

ITU-T (2003). *P. 862.1: Mapping Function for Transforming P. 862 Raw Result Scores to MOS-LQO*. Geneva: International Telecommunication Union Recommendationss. Union Geneva, 24.

ITU-T (2007). *Wideband Extension to Recommendation P.862 for the Assessment of Wideband Telephone Networks and Speech Codecs*. International Telecommunication Union Recommendationss, 862.

Khalighinejad, B., Herrero, J. L., Mehta, A. D., and Mesgarani, N. (2019). Adaptation of the human auditory cortex to changing background noise. *Nat. Commun.* 10, 2509. doi: 10.1038/s41467-019-10611-4

Koole, A., Nagtegaal, A. P., Homans, N. C., Hofman, A., Baatenburg de Jong, R. J., and Goedegebure, A. (2016). Using the digits-in-noise test to estimate age-related hearing loss. *Ear Hear.* 37, 508–513. doi: 10.1097/AUD.0000000000000282

Lee, T., and Theunissen, F. (2015). A single microphone noise reduction algorithm based on the detection and reconstruction of spectro-temporal features. *Proc. R. Soc. Math. Phys. Eng. Sci.* 471, 20150309. doi: 10.1098/rspa.2015.0309

Li, J. (2022). Recent advances in end-to-end automatic speech recognition. *APSIPA Trans. Signal Inf. Process.* 11, 50. doi: 10.1561/116.00000050

Liu, F., Jiang, S., Kang, J., Wu, Y., Yang, D., Meng, Q., et al. (2022). On the definition of noise. *Humanit. Soc. Sci. Commun.* 9, 1–17. doi: 10.1057/s41599-022-01431-x

Liutkus, A., Badeau, R., and Richard, G. (2011). Gaussian processes for underdetermined source separation. *IEEE Trans. Signal Process.* 59, 3155–3167. doi: 10.1109/TSP.2011.2119315

Loizou, P. C. (2013a). *Spectral-Subtractive Algorithms, in: Speech Enhancement: Theory and Practice.*

Loizou, P. C. (2013b). *Speech Enhancement: Theory and Practice, 2nd Edn.* Boca Raton, FL: CRC Press. doi: 10.1201/b14529

Lu, Y., and Loizou, P. C. (2008). A geometric approach to spectral subtraction. *Speech Commun.* 50, 453–466. doi: 10.1016/j.specom.2008.01.003

MATLAB (2022). *MATLAB version 9.13.0.2049777 (R2022b).* Natick, MA: The Mathworks, Inc.

McDermott, J. H. (2009). The cocktail party problem. *Curr. Biol.* 19, R1024–R1027. doi: 10.1016/j.cub.2009.09.005

McDermott, J. H., Schemitsch, M., and Simoncelli, E. P. (2013). Summary statistics in auditory perception. *Nat. Neurosci.* 16, 493–498. doi: 10.1038/nn.3347

McDermott, J. H., and Simoncelli, E. P. (2011). Sound texture perception via statistics of the auditory periphery: evidence from sound synthesis. *Neuron* 71, 926–940. doi: 10.1016/j.neuron.2011.06.032

Mesgarani, N., David, S. V., Fritz, J. B., and Shamma, S. A. (2009). Influence of context and behavior on stimulus reconstruction from neural activity in primary auditory cortex. *J. Neurophysiol.* 102, 3329–3339. doi: 10.1152/jn.91128.2008

Michelsanti, D., Tan, Z.-H., Zhang, S.-X., Xu, Y., Yu, M., Yu, D., et al. (2021). An overview of deep-learning-based audio-visual speech enhancement and separation. *IEEEACM Trans. Audio Speech Lang. Process.* 29, 1368–1396. doi: 10.1109/TASLP.2021.3066303

Middlebrooks, J. C., Simon, J.Z., Popper, A.N., and Fay, R.R. (eds.). (2017). *The Auditory System at the Cocktail Party, Springer Handbook of Auditory Research.* Cham: Springer International Publishing.

Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N., and Terzopoulos, D. (2022). Image segmentation using deep learning: a survey. *IEEE Trans. Pattern Anal. Mach. Intell.* 44, 3523–3542. doi: 10.1109/TPAMI.2021.3059968

Miyazaki, R., Saruwatari, H., Inoue, T., Takahashi, Y., Shikano, K., and Kondo, K. (2012). Musical-noise-free speech enhancement based on optimized iterative spectral subtraction. *IEEE Trans. Audio Speech Lang. Process.* 20, 2080–2094. doi: 10.1109/TASL.2012.2196513

Ochieng, P. (2022). Deep neural network techniques for monaural speech enhancement: State of the art analysis. *arXiv[Preprint].arXiv:2212.00369.*

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., et al. (2019). PsychoPy2: experiments in behavior made easy. *Behav. Res. Methods* 51, 195–203. doi: 10.3758/s13428-018-01193-y

Pruša, Z., Søndergaard, P. L., Holighaus, N., Wiesmeyr, C., and Balazs, P. (2014). "The large time-frequency analysis toolbox 2.0," in *Sound, Music, and Motion, Lecture Notes in Computer Science*, eds M. Aramaki, O. Derrien, R. Kronland-Martinet, and S. Ystad (Cham: Springer International Publishing), 419–442.

Rabinowitz, N. C., and King, A. J. (2011). Auditory perception: hearing the texture of sounds. *Curr. Biol.* 21, R967–R968. doi: 10.1016/j.cub.2011.10.027

Rix, A. W., Beerends, J. G., Hollier, M. P., and Hekstra, A. P. (2001). "Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs," in *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.01CH37221). Presented at the 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.01CH37221)* (Salt Lake City, UT), Vol. 2, 749–752.

Sarampalis, A., Kalluri, S., Edwards, B., and Hafter, E. (2009). Objective measures of listening effort: effects of background noise and noise reduction. *J. Speech Lang. Hear. Res.* 52, 1230–1240. doi: 10.1044/1092-4388(2009/08-0111)

Tan, Z.-H., Sarkar, A., and kr., Dehak, N. (2020). rVAD: An unsupervised segment-based robust voice activity detection method. *Comput. Speech Lang.* 59, 1–21. doi: 10.1016/j.csl.2019.06.005

Tanaka, H. (2019). *Editdistance: Fast Implementation of the Edit Distance (Levenshtein Distance).* Software Package.

The Relaxed Guy (2014). *3 Hours of Gentle Night Rain, Rain Sounds to Sleep, Study, Relax, Reduce Stress, Help Insomnia.* Available online at: https://www.youtube.com/watch?v=q76bMs-NwRk (accessed September 12, 2023).

Tzirakis, P., Kumar, A., and Donley, J. (2021). Multi-channel speech enhancement using graph neural networks. *arXiv[Preprint].arXiv:2102.06934.* doi: 10.48550/arXiv.2102.06934

van Rossum, R. G., and Drake, F. (2009). Python 3 reference manual. *Scotts Val. CA Creat.* 10, 1593511.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* 17, 261–272. doi: 10.1038/s41592-019-0686-2

Woods, K. J. P., Siegel, M., Traer, J., and McDermott, J. H. (2017). Headphone screening to facilitate web-based auditory experiments. *Atten. Percept. Psychophys.* 79, 2064–2072. doi: 10.3758/s13414-017-1361-2

Wu, Y. (2016). *Tensorpack.* Github Package.

Xu, R., Wu, R., Ishiwaka, Y., Vondrick, C., and Zheng, C. (2020). "Listening to sounds of silence for speech denoising," in Advances in Neural Information Processing Systems (Virtual), 33, 9633–9648. Available online at: https://proceedings.neurips.cc/paper/2020/hash/6d7d394c9d0c886e9247542e06ebb705-Abstract.html

frontiers | Frontiers in Audiology and Otology

TYPE Original Research
PUBLISHED 23 November 2023
DOI 10.3389/fauot.2023.1275210

Check for updates

OPEN ACCESS

EDITED BY
Laura Coco,
San Diego State University, United States

REVIEWED BY
Jörg Bitzer,
Jade University of Applied Sciences, Germany
Ilze Oosthuizen,
University of Pretoria, South Africa

*CORRESPONDENCE
Klaudia Edinger Andersson
✉ kandersson@health.sdu.dk

RECEIVED 09 August 2023
ACCEPTED 30 October 2023
PUBLISHED 23 November 2023

CITATION
Andersson KE, Neher T and Christensen JH (2023) Ecological momentary assessments of real-world speech listening are associated with heart rate and acoustic condition.
Front. Audiol. Otol. 1:1275210.
doi: 10.3389/fauot.2023.1275210

COPYRIGHT
© 2023 Andersson, Neher and Christensen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Ecological momentary assessments of real-world speech listening are associated with heart rate and acoustic condition

Klaudia Edinger Andersson[1,2]*, Tobias Neher[1,2] and Jeppe Høy Christensen[3]

[1]Institute of Clinical Research, University of Southern Denmark, Odense, Denmark, [2]Research Unit for Oto-Rhino-Laryngology – Head & Neck Surgery and Audiology, Odense University Hospital & University of Southern Denmark, Odense, Denmark, [3]Eriksholm Research Centre, Oticon A/S, Snekkersten, Denmark

Background and aim: Ecological momentary assessment (EMA) can provide insights into the real-world auditory ecology of hearing aid (HA) users. To better understand what factors, influence the real-world listening experiences of this population, more detailed models of human auditory ecology and behavior are needed. Laboratory studies suggest that physiological measures are sensitive to different listening experiences, as changes in physiological signals (e.g., pupil dilation) have been associated with effortful listening. In addition, real-world heart rate (HR) has been shown to be sensitive to acoustic influences (e.g., sound pressure level, SPL, and signal-to-noise ratio, SNR). Here, we hypothesized that including physiological and acoustic data in models predicting EMA ratings can provide additional insights into real-world listening outcome. To test this, we collected and analyzed longitudinal data from individuals with normal hearing.

Method: Fifteen normal-hearing adults completed smartphone-based EMAs regarding their listening experiences during a 2-week period. When completing the EMAs, they had to indicate their current listening intent. The participants received a single HA each that they fastened to their collars. The HAs were used to collect continuous SPL and SNR data in the participants' daily environments. Wristbands worn by the participants were used to collect continuous HR data.

Results: Linear mixed-effects models with SPL, SNR, and HR as fixed effects and participant as random intercept showed that higher SPL and lower SNR were associated with lower (poorer) EMA ratings. Including listening intent in the analyses revealed increased HR in "speech communication" and "focused listening" situations to be associated with better EMA ratings relative to situations without any specific listening intent.

Conclusions: Our findings indicate that including *in-situ* HR and acoustic measurements can improve the prediction of real-world listening experiences. Further, they suggest that listening intent significantly impacts self-reported listening experiences and their association with physiological responses. Specifically, better listening experiences in speech communication situations are associated with higher HR.

KEYWORDS

ecological momentary assessment, smartphone, heart rate, objective measurement, hearing aids, data-logging

Introduction/Background

Hearing aid (HA) noise management features (i.e., directional microphones combined with noise reduction) are designed to respond to ambient acoustics and adapt their activation levels accordingly (Dillon, 2012). HA benefit in everyday listening is typically more prominent when environments are more auditorily demanding (Wu et al., 2019; Andersson et al., 2021). Everyday benefit from HA use also depends on the type and importance of intended listening activities and interacts with auditory demands (von Gablenz et al., 2021). For example, benefit from HA noise management may be different whether the user does or does not intend to listen to any target sounds (e.g., speech) in poor acoustic conditions. Depending on the nature of listening intent (i.e., active listening to speech or passive listening), the HA user might emphasize different aspects of the listening experience (i.e., understanding speech or reduced annoyance from background noise).

Thus, to maximize benefit, noise management solutions should not only be based on ambient acoustic information but also on information about how changes in the user's auditory demands interact with momentary listening intents to successfully complete the listening task. More generally, there is a need to increase our understanding of how the auditory ecology and behavior of HA users are associated with their listening experiences in daily life. Here, auditory ecology relates to the particular listening demands an individual faces in different surroundings (Jensen and Nielsen, 2005).

Traditionally, real-world listening experiences of HA users have been evaluated with retrospective questionnaires or interviews, while objective assessments of HA benefits typically rely on laboratory experiments under well-defined and controlled settings, that is, in specific non-naturalistic situations (e.g., Gnewikow et al., 2009). This means that laboratory outcomes do not translate effectively to the real world, where conditions are constantly changing. Greater ecological validity in hearing research can result in tests featuring more realistic sound scenarios, enhancing their applicability to real-world acoustics (Keidser et al., 2020). However, laboratory testing may not fully account for the influence of daily-life activities, interactions, and listening intentions in shaping individuals' real-world listening experiences (Pichora-Fuller et al., 2016). Instead, experience sampling methods have been proposed to better reflect real-world listening experiences. Generally, ecological momentary assessment (EMA) aims to collect *in-situ* self-reports in natural environments regarding behavior, motivation, experiences, thoughts, or feelings (Shiffman et al., 2008). Such *in-situ* reports can be collected from HA users which can then be linked to real-world acoustics obtained with data-logging (Andersson et al., 2021), and in combination with self-reported listening activities or intentions (von Gablenz et al., 2021).

Several EMA studies have provided insights into the listening experiences of HA users, their listening environments, and situations they typically encounter. The studies have consistently reported that adult HA users mostly encounter quite listening environments while listening experiences in noisy environments are being less frequently reported (Humes et al., 2018; Wu et al., 2019; Burke and Naylor, 2020; Schinkel-Bielefeld et al., 2020; Andersson et al., 2021; von Gablenz et al., 2021). When EMAs are extended with acoustic data-logging provided by HAs, individual assessments can be linked to real-world acoustic contexts. In this manner, specific HA technologies or features can be evaluated in real-world acoustic contexts they were designed for. For example, Andersson et al. (2021) showed that HA users significantly benefit from HA noise management as compared to default HA settings in listening situations dominated by speech or speech in noise signals. Findings like this can help reduce the incongruence between laboratory and real-world HA outcomes (e.g., Gnewikow et al., 2009).

Despite the benefits inherent to EMA methodology, it is important to consider some limitations. While speech communication-related situations account for roughly 50% of experienced listening situations (Pichora-Fuller et al., 2016) in daily life (Humes et al., 2018; Burke and Naylor, 2020; Andersson et al., 2021), self-reports in such situations can often be difficult to collect with EMA (Schinkel-Bielefeld et al., 2020; Wu et al., 2021). Schinkel-Bielefeld et al. (2020) concluded that EMAs regarding speech communication can be more difficult to collect as participants often experience such assessments as being inappropriate in social contexts. This suggests that the general EMA approach could be improved by extending it with objective non-invasive measures of listening-related factors that possibly correlate with real-world listening experiences during speech understanding.

Regarding situations related to speech communication, HA users typically struggle to hear in the presence of background noise (Henry and Heinz, 2012). In laboratory assessments of speech intelligibility in noisy conditions, a consistent finding is that HA users often need to exert greater listening effort compared to individuals with normal hearing (Ohlenforst et al., 2017). Listening effort is influenced by various factors, including the listener's hearing ability, the demands of the listening task (such as noisy or reverberant environments), and the listener's motivation to successfully complete the task, potentially receiving personal or social rewards (Pichora-Fuller et al., 2016). These multifaced aspects of listening experience have been described by the Framework for Understanding Effortful Listening (FUEL). Importantly, listening effort is not a static phenomenon as it can fluctuate throughout an activity based on both the task's demands, such as the difficulty of listening in noisy situations, and the listener's motivation or evaluation of the task's importance (Pichora-Fuller et al., 2016; Peelle, 2018). Also, the ramifications of persistent listening effort in everyday life are significant, potentially leading to fatigue, which can negatively impact the social lives of individuals with hearing impairment (Alhanbali et al., 2018). This underscores the importance of considering effort as a crucial aspect when assessing the benefit of hearing aids.

The assessment of listening effort and listening difficulties can be facilitated by physiological measurements (Mackersie and Calderon-Moultrie, 2016; Ohlenforst et al., 2017; Zekveld et al., 2018; Alhanbali et al., 2019; Francis et al., 2021; Giuliani et al., 2021). Moreover, non-invasive physiological recordings could also support the assessment of real-world listening experiences with EMA. In fact, there is increasing evidence for real-world sound exposure to be associated with changes in mean heart rate (HR).

Specifically, the study by Christensen et al. (2021) found that higher ambient sound pressure levels were associated with increases in HR, while El Aarbaoui et al. (2017) further documented that higher sound pressure levels also relate to decreases in HR variability. Moreover, increases in ambient signal-to-noise ratio were linked to decreases in HR, particularly in noisy environments (Christensen et al., 2021). In their respective studies, Christensen et al. (2021) used hearing aids and wrist-worn wearables to collect real-world environmental sound and HR data, while El Aarbaoui et al. (2017) had participants wear shoulder-mounted noise dosimeters and medical-grade heart-rate monitoring devices. However, the effect sizes reported in the two studies are almost identical with 0.154% increase in HR per 1 dB SPL (El Aarbaoui and Chaix, 2019) vs. 0.141% increase in HR per 1 dB SPL (Christensen et al., 2021), indicating that acoustic data-logging with HAs exhibit high face validity.

Thus, building upon the findings of Christensen et al. (2021), the present study aims to explore the association between environmental acoustic factors, physiological responses (i.e., HR), and self-reported listening experiences and intentions via EMA.

Previous research has shown that acoustic factors are related to both self-reported listening experiences (Andersson et al., 2021) and momentary HR (Christensen et al., 2021), but there is lack of evidence about how listening intentions impact these associations (von Gablenz et al., 2021). Thus, the aim was to link acoustic and HR data to self-reported listening experiences and intentions to enable a broader understanding of how such factors interact in relation to HA outcome.

We hypothesized that higher sound pressure level (SPL), lower signal-to-noise-ratio (SNR) and higher momentary HR, respectively, would be associated with poorer self-reported listening experiences. Further, we hypothesized that there would be a moderating effect of listening intention. That is, we hypothesized that the association between self-reported listening experiences and HR would be stronger in the case of specific listening intentions compared to non-specific listening. We tested these hypotheses on a group of normal-hearing individuals.

Materials and methods

Study design

The current study was evaluated by the Regional Committee on Health Research Ethics for Southern Denmark, and ethical approval was deemed unnecessary (i.e., a waiver was obtained). The present study was part of a larger project, which included two groups of individuals with normal and hearing loss, respectively. The results from the participants with hearing impairment will be reported elsewhere. For the current dataset, associations between the acoustic and HR data have been reported elsewhere (Christensen et al., 2023). All data were collected between March 2021 and June 2023. Each participant was paid 120 DKK/h for the time spent visiting the laboratory. All participants provided written informed consent.

To our knowledge, the current study was the first one to investigate potential associations between self-reported listening experiences, acoustic, and HR data. Given this, our study was primarily exploratory in nature. Our study resembled other EMA studies in terms of the number of prompted assessments and the duration of the data collection period (Holube et al., 2020). We chose a 2-week trial period, as previous research has suggested that longer data collection periods can become burdensome for participants, potentially resulting in decreased compliance (Schinkel-Bielefeld et al., 2020).

We applied a randomized crossover study design. Participants began either with a 2-week period of collecting objective data, including acoustic data and HR measurements, while also completing EMAs. Alternatively, they began with only the collection of objective data (without any EMA completions) for 1 week. The 1-week trial period without EMAs served as a control condition to assess if the mere act of completing EMAs influenced HR readings. Our analysis did not reveal a significant difference in HR between the two trial periods. In total, the data collection period lasted for 3 weeks and consisted of four visits to the laboratory. The first visit included hearing screening and comprehensive instructions in the use of the equipment as the participants started with the data collection afterwards Additionally, participants received a detailed paper guide explaining device usage and basic troubleshooting procedures. We encouraged participants to seek additional assistance if required. The remaining three visits occurred over the subsequent 3 weeks and were primarily dedicated to transferring data from participants' smartphones to a computer and verifying the correct storage of all data. The final visit encompassed the return of the equipment and the completion of a brief questionnaire concerning their participation in the study.

Participants

Participants with self-reported normal hearing were recruited from the student population of the University of Southern Denmark in Odense, Denmark. Individuals interested in participating were also encouraged to pass on flyers to friends and family members. The inclusion criteria included audiometric hearing thresholds $\leq$25 dB HL between 0.25 and 8 kHz for both ears. Self-reported health issues (e.g., a pacemaker) that are known to affect the cardiovascular system were defined as exclusion criteria. By including individuals with normal hearing, we avoided potential confounds due to hearing loss and HA use on our results, allowing them to be used for reference purposes, for example when evaluating the listening experiences of HA users in future studies.

Initially, 12 participants (four males, eight females) were enrolled in the study. Data from four participants had to be excluded due to technical issues ($n = 1$) or an insufficient number of completed self-reports ($n = 3$). Thus, seven additional participants were recruited and enrolled in the study. In total, 15 participants (five males, 10 females) completed the study. The age of these participants ranged from 23 to 35 years (mean: 27.7 years; SD: 3.9 years). The participants were screened with air-conduction pure-tone audiometry to confirm normal thresholds. All participants were familiar with using smartphones in their daily life. The majority of participants were university students ($n = 9$), and

among them, three held part-time jobs. The remaining participants were either employed full-time ($n = 5$) or part-time ($n = 1$).

Material and apparatus

Hearing aids

Each participant was provided with a single HA fixed to a metal clip that could be attached to the collar. The rationale for the collar placement was to ensure consistent daily wear by participants during data collection. Placing the HA behind the ear could lead to discomfort, such as occlusion or the inability to wear headphones, which might prompt participants to remove the HA, resulting in less acoustic data collected. Participants were instructed to keep the HA in the same position and ensure it was not obstructed by items like jackets to maintain reliable logging of acoustic data.

The HAs were small-size receiver-in-the-ear (mini-RITE) OPN S1 devices from Oticon A/S (Smørum, Denmark) with rechargeable batteries. No receivers or earpieces were used. The HAs were used solely for acoustic data-logging. Thus, familiarity with HA usage was not relevant for the current study. The participants were instructed in how to correctly place the HA in a charger and connect it to a smartphone via Bluetooth. They were instructed to charge it every evening to ensure enough battery power during the next day. The participants could monitor the battery level on the associated smartphone screen or app.

When turned on, the HAs continuously measured the SPL and estimated the SNR of the ambient sound environment. These data were transferred and stored on the Bluetooth-connected smartphone every 20 s. Both SPL and SNR were calculated in a broadband sense (0–10 kHz) with A-weighting applied (i.e., in dBA). A detailed description of the acoustic parameters can be found in Christensen et al. (2021). While charging, the HAs were automatically turned off, and so no acoustic data were logged then.

Wristbands

Garmin (Olathe, Kansas, USA) Vivosmart 3 and 4 wristbands were used to measure HR continuously. The wristbands were set up with default settings and connected via Bluetooth to a smartphone app for storage of these measurements on a beat-by-beat basis. Previously, the validity of commercial wristbands as compared to research-based electrocardiograms was investigated, leading to the conclusion that "different wearables are all reasonably accurate at resting and prolonged elevated HR, but that differences exist between devices in responding to changes in activity" (Bent et al., 2020). Also, the green LED sensor light (used in Garmin Vivosmart 3 and 4) is shown to be resistant to motion artifacts when measuring HR (Nelson et al., 2020). For each participant, the HR data exceeding the 95% percentile were excluded to reduce the effects of physical exercise on these measurements.

Smartphones

Each participant received an iPhone 7 smartphone (Apple, Cupertino, California, USA) that enabled a connection with the HAs and wristbands. The smartphones were used to perform the EMAs via an app and to store the collected HR and acoustic data.

Data collection

HR app

The HR data were stored in a research version of a commercially available app for HA management called "Oticon ON" (Oticon A/S, Smørum, Denmark). Besides the features that are available in the commercial app, the research version included two additional features: (1) connection and synchronization of the Garmin wearable with the smartphone, and (2) live tracking of the logged HR and SPL data from the HA. The participants were asked to use live tracking at least once a day to ensure connectivity between the smartphone, wristband, and HA. The HR and acoustic data shared the same timestamps.

EMA app

An iPhone app developed by Oticon A/S (Smørum, Denmark) was used to perform the EMAs, which were afterwards linked to the acoustic and HR data with the help of the timestamps. The participants were prompted pseudo-randomly during a day with app notifications to complete EMAs. The app prompting was enabled only when the HA was connected to the smartphone and when the EMA from the latest notification was completed. The notifications were sent every 1.5–2 h, but no more than eight times per day. The app notifications included both audible and vibratory alerts from the smartphone. Additionally, the wristbands were set to vibrate when the participants received prompts to improve compliance. The participants were also encouraged to self-initiate EMAs if they experienced a listening situation they considered interesting to assess. They were instructed to always think about the last 5 min of listening experiences when completing an EMA, regardless of whether it was initiated exactly at or sometime after a prompt, or whether it was self-initiated. Each EMA consisted of seven questions in total. To assist participants in their ratings, the answers to the first six questions were indicated using a slider on a line with five marks between two anchors (Table 1). The outcomes from the first six questions were coded as continuous numbers between 0 and 10 (these were not visible for the participants). The EMA app was mainly designed with HA users in mind and included an initial question regarding satisfaction with the sound from the HAs. The participants in the current study were instructed to use this question to rate how pleasant or unpleasant the sounds around them were. The last question had a selection menu where the participants could indicate their current listening intent. They were asked to select "streamed listening" when listening through headphones. Figure 1 and Table 1 lists the questions presented in the EMA app.

Data processing and statistical analyses

To reduce the number of variables, the reported listening intents were clustered into three categories according to the Common Sound Scenarios (CoSS) framework by Wolters et al. (2016). The rationale for including the CoSS framework was based on it being increasingly used in HA research (e.g., Wolters et al., 2016; Burke and Naylor, 2020; von Gablenz et al.,

TABLE 1 List of questions used in the app with corresponding abbreviations of the questions.

Question number	Question text	Anchor names/possible answers	Abbreviation of question text
1	Right now, how satisfied are you with the sound from your hearing aids?	Not satisfied ↔ Very satisfied	Satisfaction
2	Right now, how is it to focus on the sounds you want to hear?	Difficult ↔ Easy	Focusing on sounds
3	Right now, how is it to ignore sounds you don't want to hear?	Difficult ↔ Easy	Ignoring sounds
4	Right now, how is it to work out where different sounds are coming from?	Difficult ↔ Easy	Sound localization
5	Right now, how well can you hear what is going on around you?	Not very well ↔ Very well	Audibility
6	How noisy is it right now?	Very noisy ↔ Quiet	Noisiness
7	What are you listening to at the moment?	Choose one: One person talking People talking Music: live or sound system Music via streaming A streamed broadcast Sounds around me Nothing in particular	Listening intent

2021). The listening intents "one person talking" and "people talking" were classified as "speech communication" from the CoSS framework, whereas "nothing in particular" was classified as "non-specific." The other possible listening intents were classified as "focused listening."

The individual associations between HR, SPL, and SNR and EMA ratings were analyzed using linear mixed-effects (LME) models. The LME models accounted for multilevel data, that is, correlated observations nested within each participant and condition due to repeated measurements (Oleson et al., 2022). To account for individual baseline variability, the participants were defined as random effects (i.e., random intercepts) in the models (Barr et al., 2013). The continuous variables HR, SPL, and SNR were defined as individual fixed effects in the LME models, as defined below:

$$\begin{aligned} Y_{\text{EMA rating}_{ij}} &= \beta_0 + \beta_1 X_{\text{SPL}_{ij}} + \beta_2 X_{\text{SNR}_{ij}} + \beta_3 X_{\text{Heart rate}_{ij}} \\ &+ b_{0i} + e_{ij} \end{aligned} \quad (1)$$

Y denotes the response variable (i.e., the EMA ratings) for participant i and repeated observation j, β_0 is the intercept for the baseline EMA rating, the other β_n are the coefficients for the fixed effects, X are the fixed effects, b_0 is the random intercept for each participant, and e represents the residual.

Prior to the modeling, the fixed effects were converted into z-scores using the following general formula:

$$z - \text{score} = \frac{x - \mu}{\sigma} \quad (2)$$

Here, x denotes the raw value (e.g., dB SPL), μ is the mean, and σ refers to standard deviation. A z-score equal to 0 represents the observed grand average value (e.g., SPL) across all participants while a z-score equal to 1 represents an observed value one standard deviation from the grand average.

Initially, ratings from each EMA question were modeled separately, and models were compared using likelihood ratio tests to corresponding null models (i.e., intercept-only models) and simpler models which only included SPL and SNR as fixed effects (Harrison et al., 2018). This was done to test if the acoustic data significantly contributed to explaining the EMA ratings, and if this differed across EMA questions. Prior to any modeling, the ratings from question 6 were inverted to match the rating scales of the other EMA questions.

Secondly, we sought to investigate if associations between the predictor variables and EMA ratings were moderated by listening intent. Listening intent was included in interaction with SPL, SNR, and HR, respectively. In this case, the data from EMA questions 1–5 and those from question 6 (for description of each question see Table 1) were analyzed in two separate models since the former relate to the general listening experience while the latter (subjective rating of noisiness) relates to the environment. Furthermore, for the data from EMA questions 1–5, EMA question was included as a random effect term rather than as a fixed effect, since we were interested in the overall effect of listening intent generalized to all EMA questions:

$$\begin{aligned} Y_{\text{EMA rating}_{ij}} &= \beta_0 + \beta_1 X_{\text{SPL}_{ij}} + \beta_2 X_{\text{SNR}_{ij}} + \beta_3 X_{\text{Heart rate}_{ij}} \\ &+ \beta_4 X_{\text{listening intent}_{ij}} + \beta_5 X_{\text{SPL}_{ij}} X_{\text{listening intent}_{ij}} \\ &+ \beta_6 X_{\text{SNR}_{ij}} X_{\text{listening intent}_{ij}} \\ &+ \beta_7 X_{\text{Heart rate}_{ij}} X_{\text{listening intent}_{ij}} + b_{0i} + b_{1i} + e_{ij} \end{aligned} \quad (3)$$

Again, likelihood ratio tests were applied to compare the goodness-of-fit for the two resultant interaction models with simpler models that excluded listening intent.

Besides inspecting the coefficient magnitudes and error of each LME model, conditional and marginal R^2 effect sizes were considered, as these indicate whether the inclusion of the acoustic,

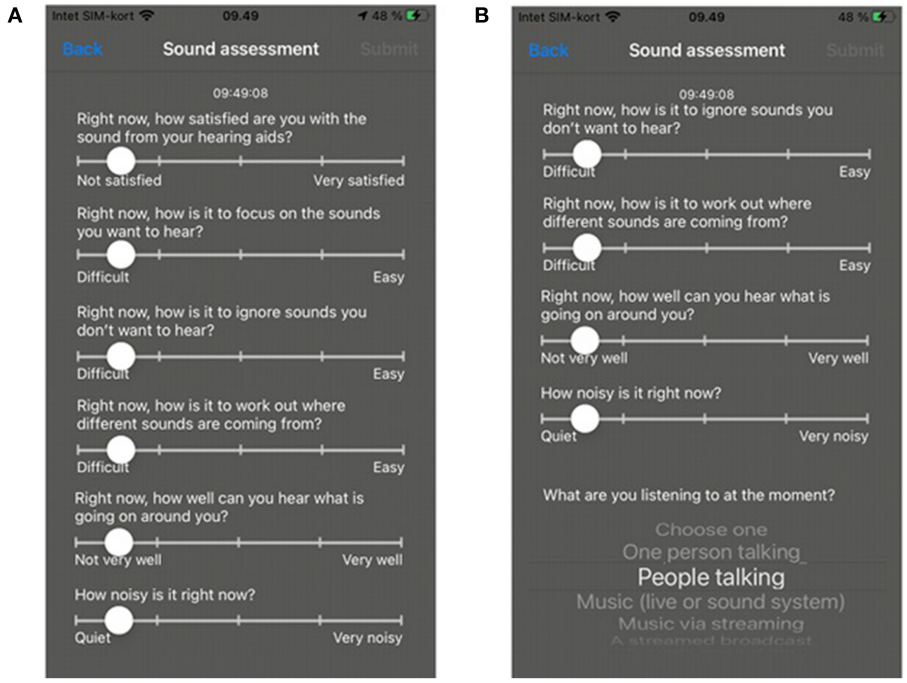


FIGURE 1
Screenshots of the Oticon EMA app. **(A)** Shows the welcome screen. **(B)** Show EMA questions with slider indicator and last question as a single-choice question. EMA, ecological momentary assessment.

HR, or listening intent data in the interaction model increased the proportion of explained variance (for more details, see Christensen et al., 2021).

All statistical analyses were conducted using R v.4.2.2 software. Visualization of descriptive statistics was done with the "ggplot2" package (v.3.3.6). The HR, SPL, and SNR data were averaged over a 5-min time window prior to each EMA completion. The density distributions of SPL, SNR, and HR were calculated and drawn by applying Gaussian kernel density estimate and non-parametric Kolmogorov-Smirnov tests were conducted to assess the differences in the distributions of SPL, SNR, and HR data, respectively, among the listening intent categories.

The LME analyses were performed with the "lmerTest—Tests in Linear Mixed Effects Models" package (v.3.1-3) with the use of the "sjPlot—Data Visualization for Statistics in Social Science" package (v.2.8.11) for tables and graphical plots of LME coefficients and interactions. Partial pseudo-R^2 was calculated using the "MuMIn—Multi-Model Inference" (v.1.46.0) package.

Results

Descriptive results

Distribution of EMAs

During the 2-week EMA data collection period, the participants submitted 1,521 EMAs in total with an average of 7.2 EMAs (SD = 4.4 EMAs) per day and participant. To assess the associations between the EMA ratings and acoustic factors, all data which stemmed from listening intents related to "streaming" (i.e., listening through headphones) were excluded. This resulted in 1,260 EMAs with an average of 6 EMAs (SD = 3.4 EMAs) per day and participant. Figure 2 shows the listening intents classified according to the CoSS framework. Figure 2A shows that the median number of submitted EMAs across the 2 weeks was 68 EMAs (after data exclusion). Figure 2B shows that most EMAs were submitted between 6:00 a.m. and 8:00 p.m.

Figure 3A shows the percentage distribution of self-reported listening intents during the EMAs across all original options (after exclusion of "streaming" related listening intents), whereas Figure 3B shows the listening intents classified according to the CoSS framework. As shown in Figure 3A, the option "nothing in particular" (31.3%) was the most selected listening intent followed by "sounds around me" (23.3%). Regarding the classifications based on the CoSS framework, "speech communication" accounted for most selected listening intents (i.e., 37.1%), whereas "focused listening" and "non-specific" accounted for 31.6 and 31.4%, respectively (Figure 3B).

Distribution of acoustic and HR data

Figure 4 shows density distributions of the acoustic data from the HA data-logging (i.e., SPL and SNR) and the HR data from the Garmin wristbands separated by CoSS category.

The highest median SPL (60.6 dB, SD = 9.6 dB) was obtained for "speech communication" listening intents (Figure 4A). The median SPL for "focused listening" and "non-specific" was 56.7 dB (SD = 10.9 dB) and 48.7 dB (SD = 9.8 dB), respectively (Figure 4A). In general, "non-specific listening" was characterized by lower SPL values as compared to "speech

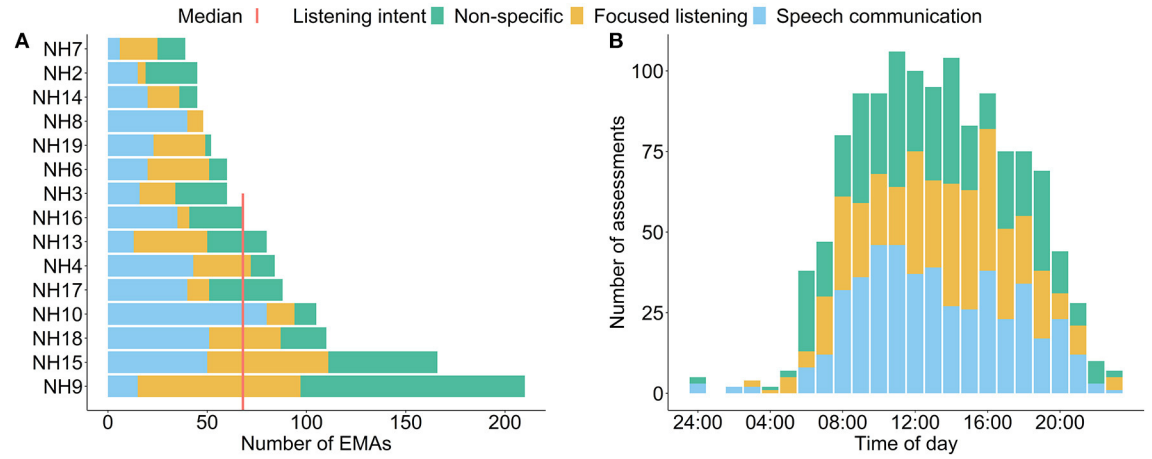


FIGURE 2
Number of submitted EMAs as a function of listening intents corresponding to CoSS classifications. **(A)** Number of EMAs submitted by each participant. The vertical red line represents the median value for submitted EMAs per participant but not divided into CoSS. **(B)** Stacked histograms showing the total number of EMAs across the time of day. EMA, ecological momentary assessment; CoSS, Common Sound Scenarios framework.

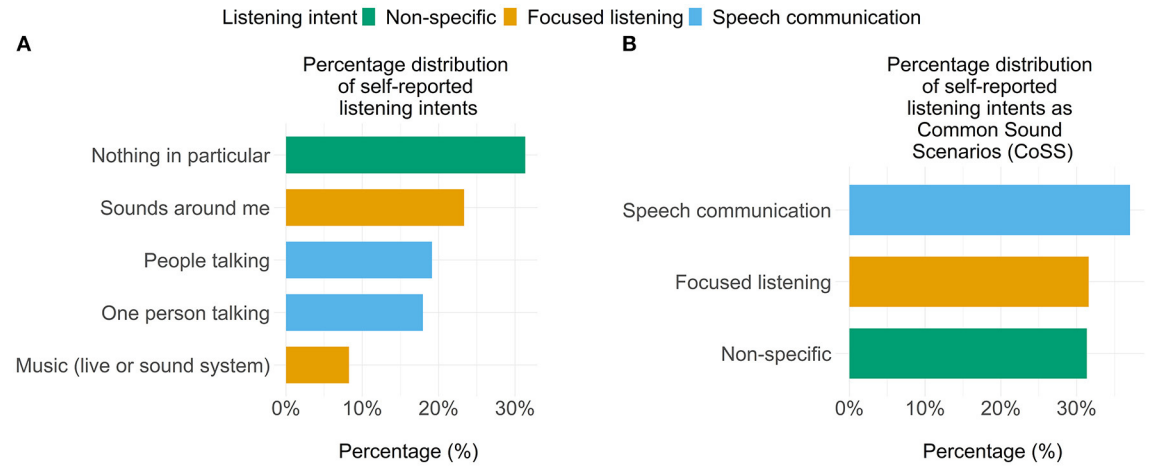


FIGURE 3
Percentage distribution as a function of listening intent in descending order. **(A)** Bar chart showing percentage of selected listening intents by participants. **(B)** Bar chart showing percentage of the selected listening intents classified as Common Sound Scenarios. Speech communication consists of "one person talking" and "people talking" listening intents, focused listening consists of "sound around me" and "music (live or sound system)" listening intents, and non-specific relates to "nothing in particular" listening intent.

communication" and "focused listening" intents (Figure 4A). Kolmogorov-Smirnov tests confirmed that the distribution of SPLs for "non-specific listening" was different from those for "focused listening" and "speech communication" ($D = 0.299$, $p < 0.001$, and $D = 0.468$, $p < 0.001$, respectively), and between "focused listening" and "speech communication" ($D = 0.215$, $p < 0.001$).

For SNR, the highest median value was estimated for "speech communication" (19.2 dB, SD = 4.4 dB), followed by "focused listening" (16.7 dB, SD = 4.2 dB) and "non-specific" (15.7 dB, SD = 4.7 dB) as shown in Figure 4B. Kolmogorov-Smirnov tests showed that the distributions of SNRs among the three listening intent categories were different from each other (all $p <$ 0.001). Further, Pearson's correlation coefficient r demonstrated a moderate correlation between SPL and SNR values ($r = 0.66$, $p < 0.001$).

The observed grand mean HR was 76.6 bpm (SD = 6.6 bpm). This corresponds well with the mean HRs of 80.2 bpm (SD = 14.8 bpm) and 78.5 bpm (SD = 15.1 bpm) reported by Avram et al. (2019) based on more than 15,000 adults aged 21–30 and 31–40 years, respectively. The median HR value for "speech communication" listening was slightly higher (77.6 bpm, SD = 6.3 bpm) than for "focused listening" and for "non-specific" (76 bpm, SD = 6.3, and 75.6 bpm, SD = 5.4, respectively) as shown in Figure 4C. The HR values mostly overlap across the three listening intent categories, with "non-specific" generally showing the highest

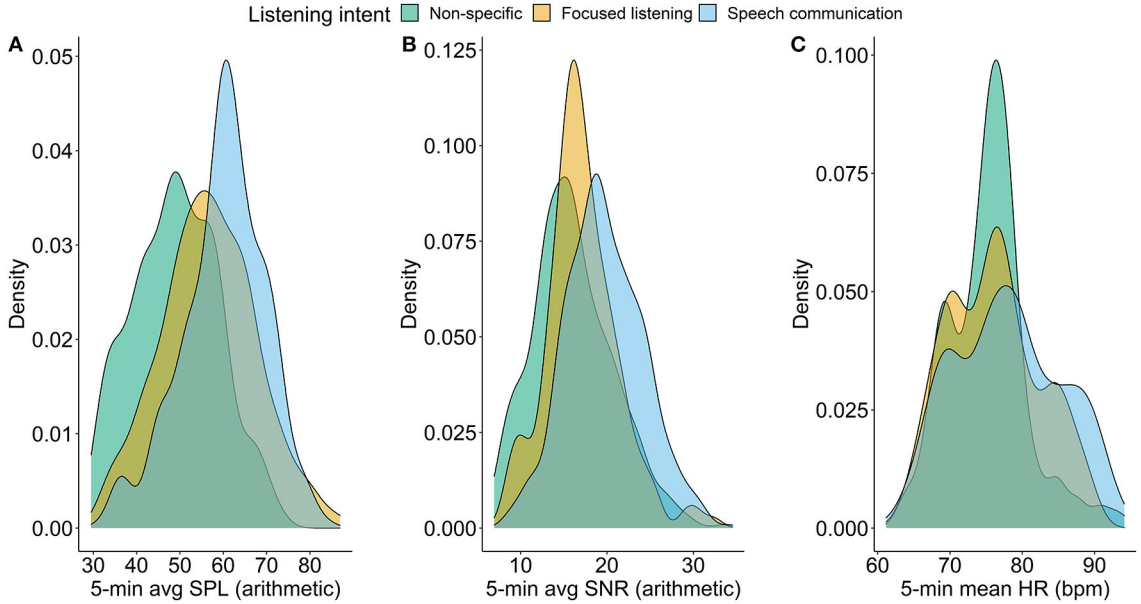


FIGURE 4
Density plots as a function of listening intent. Each data point represents a 5-min average corresponding to the time window for each ecological momentary assessment completion. **(A)** Averaged values for sound pressure levels (SPLs). **(B)** Averaged values for signal-to-noise ratio (SNR). **(C)** Averaged values for momentary HR (HR).

density around mean HR values (Figure 4C). A Kolmogorov-Smirnov test revealed that the distributions of the HR data for the listening intent categories were significantly different from each other ("non-specific" vs. "speech communication": $D = 0.294$, $p < 0.001$; "non-specific" vs. "focused listening": $D = 0.156$, $p < 0.001$; "focused listening" vs. "speech communication": $D = 0.173$, $p < 0.001$).

Correlations between ratings from individual EMA questions

To assess the validity of modeling each EMA question separately, we analyzed the responses from the different EMA questions in terms of multicollinearity. Figure 5 depicts how strongly the responses from the different EMA questions were correlated. For that purpose, Spearman's correlation coefficients (marked with bold) were calculated for each participant and then averaged. The numbers below the coefficients show the total number of significant coefficients ($p < 0.05$) per comparison. The average correlations showed weak to moderate associations between the different EMA questions, which indicates that the participants could differentiate between the different EMA dimensions.

Associations between acoustic, HR, and EMA data

To recapitulate, we hypothesized that higher SPL, lower SNR, and higher momentary HR, respectively, would be associated with poorer self-reported listening experience (i.e., lower EMA ratings). Figure 6 depicts the fixed-effects coefficients from the LME models for the individual EMA questions. There were significant and negative effects of SPL for all EMA questions ($p < 0.001$ for all EMA questions). Thus, higher SPLs were found to be associated with lower EMA ratings. This effect was strongest for EMA question 6 (rating of noisiness of the surroundings). Except for question 1, the fixed-effect coefficients for SNR were all positive and statistically significant ($p < 0.01$ for question 4 and $p < 0.001$ for questions 2–3 and 5–6). For HR, no significant coefficients were observed. Coefficient estimates and effect sizes regarding estimates from the LME models can be found in Table 2. While the applied LME models corrected for inter-individual differences in baselines, the imbalance in amount of EMAs per participant (see Figure 2) could have influenced the slope estimates. We therefore re-fitted the LME models to ratings from all EMA questions using only a subset of data from NH15 and NH9. Specifically, we randomly selected 146 EMAs from NH9 and NH15 as this corresponds to the number of EMAs from the participants with the 3rd most completed EMAs. Inspecting the re-fitted models, we only identified negligible and unsystematic changes in coefficients, while the direction and significance of effects were unchanged.

Moderating effect of listening intent

Our second hypothesis was that the strength of the associations between the acoustic and physiological data and the self-reported listening experiences would be influenced by the individual's listening intent. To test this, we included self-reported listening intent in the LME models as a fixed effect in interaction with SPL, SNR, and momentary HR. This improved the goodness-of-fit for

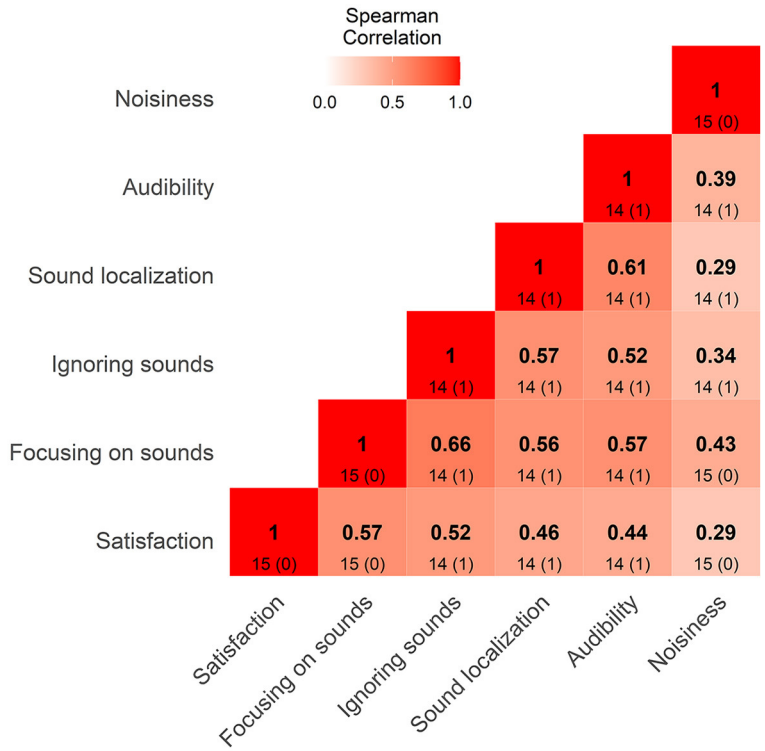


FIGURE 5
Correlation matrix displays Spearman's rho averaged correlation coefficients (marked with bold) for rating ratings for different ecological momentary assessment questions. The matrix shows the correlation between all the possible pairs of rating values. Red color represents positive correlation coefficients. The values below correlation coefficients indicate number of statistically significant correlations with significance level of 0.05, whereas values in the parentheses indicate number of p-values where the correlation could not be estimated due to no variance in ratings.

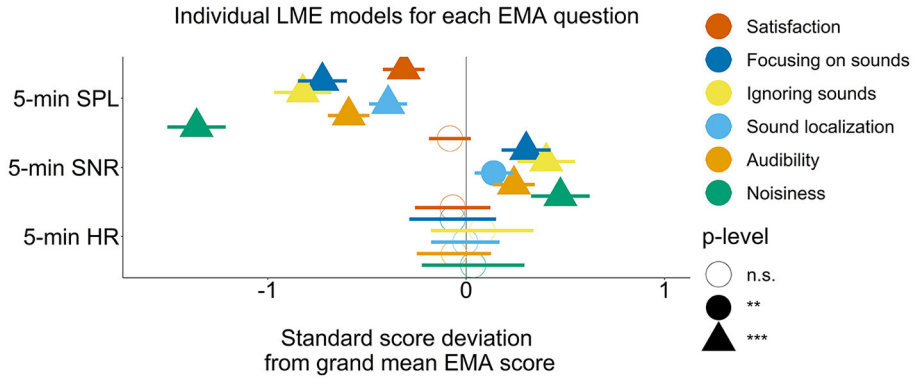


FIGURE 6
Plots of standard deviation from grand mean EMA rating with 95%-confidence intervals for fixed effects in LME models. Individual LME model estimated fixed-effects coefficients for each EMA question. Significance levels for p-values: n.s., non-significant; **p < 0.01, ***p < 0.001. LME, linear mixed effects; EMA, ecological momentary assessment; SPL, sound pressure level; SNR, signal-to-noise ratio; HR, heart rate.

both models [model with EMA questions 1–5 as random effect: $\chi^2_{(8)} = 153.4$, $p < 0.001$; model of EMA question 6: $\chi^2_{(8)} = 110.2$, $p < 0.001$]. The model that included listening intent explained 1.8 percentage point (total: 34.6% increase) more variance (EMA questions 1–5) compared to the models with only SPL, SNR, and HR. For EMA question 6 (rating of noisiness of the surroundings), the increase in explained variance was 6 percentage point (total: 31% increase).

The coefficients from these interaction models (see Table 2) revealed that, for the baseline condition (non-specific listening), EMA ratings were significantly and negatively associated with SPL (as in the simpler model without listening intent) for the general listening experience (questions 1–5). Also, higher SPL was associated with ratings of more noise in the surroundings for the baseline condition (for question 6). The association between SNR and EMA ratings was not statistically significant

10.3389/fauot.2023.1275210

TABLE 2 Estimates for fixed-effect coefficients, confidence intervals, and *p*-values from linear mixed-effects models testing associations between acoustic data (SPL, SNR), HR, and EMA rating.

Coefficient	EMA question 1: satisfaction		EMA question 2: focusing on sounds		EMA question 3: ignoring sounds	
	Estimate (rating) CI (95%)	*P*-value	Estimate (rating) CI (95%)	*P*-value	Estimate (rating) CI (95%)	*P*-value
Intercept	8.36 (7.60 to 9.13)	**<0.001**	8.59 (7.92 to 9.26)	**<0.001**	8.32 (7.45 to 9.19)	**<0.001**
SPL (SD)	−0.31 (−0.42 to −0.21)	**<0.001**	−0.72 (−0.85 to −0.60)	**<0.001**	−0.82 (−0.97 to −0.68)	**<0.001**
SNR (SD)	−0.08 (−0.19 to 0.02)	0.127	0.30 (0.18 to 0.43)	**<0.001**	0.40 (0.26 to 0.55)	**<0.001**
HR (SD)	−0.07 (−0.26 to 0.12)	0.484	−0.07 (−0.29 to 0.15)	0.542	0.08 (−0.18 to 0.34)	0.540
Random effects						
σ^2	1.66		2.29		3.16	
τ_{00}	1.95_{ID}		1.48_{ID}		2.51_{ID}	
ICC	0.54		0.39		0.44	
N	13_{ID}		13_{ID}		13_{ID}	
Observations	1,130		1,130		1,130	
Marginal R^2/Conditional R^2	0.037/0.557		0.080/0.441		0.068/0.481	
Coefficient	EMA question 4: sound localization		EMA question 5: audibility		EMA question 6: noisiness	
	Estimate (rating) CI (95%)	*P*-value	Estimate (rating) CI (95%)	*P*-value	Estimate (rating) CI (95%)	*P*-value
Intercept	8.95 (8.28 to 9.62)	**<0.001**	8.84 (8.21 to 9.47)	**<0.001**	7.24 (6.50 to 7.98)	**<0.001**
SPL (SD)	−0.39 (−0.49 to −0.30)	**<0.001**	−0.59 (−0.70 to −0.49)	**<0.001**	−1.36 (−1.51 to −1.21)	**<0.001**
SNR (SD)	0.14 (0.04 to 0.23)	**0.005**	0.24 (0.14 to 0.35)	**<0.001**	0.47 (0.33 to 0.62)	**<0.001**
HR (SD)	−0.00 (−0.18 to 0.17)	0.962	−0.06 (−0.25 to 0.13)	0.518	0.04 (-0.22 to 0.29)	0.790
Random effects						
σ^2	1.38		1.65		3.27	
τ_{00}	1.50_{ID}		1.32_{ID}		1.80_{ID}	
ICC	0.52		0.44		0.35	
N	13_{ID}		13_{ID}		13_{ID}	
Observations	1,130		1,130		1,130	
Marginal R^2/Conditional R^2	0.034/0.537		0.069/0.482		0.195/0.480	

SPL, SNR, and HR were modeled as fixed effects, while the participants and listening intents were modeled as random effects. *P*-values in bold indicate that they are below the significance level (alpha = 0.05). SPL, sound level pressure; SNR, signal-to-noise-ratio; HR, heart rate; SD, standard deviation; EMA, ecological momentary assessment. σ^2, variance; τ_{00}, between-subject variance; ICC, interclass correlation coefficient; N, sample size; ID, participant ID as random effect; Marginal R^2, variance of fixed effects; Conditional R^2, variance of both the fixed and random effects coefficient; N, sample size; ID, participant ID as random effect; Marginal R^2, variance of fixed effects; Conditional R^2, variance of both the fixed and random effects.

for the baseline condition in both models. Further, the model for EMA questions 1–5 showed a significant negative association between HR and EMA ratings for non-specific listening (baseline condition). Interestingly, the models revealed several significant interactions. These can be seen in Figure 7, which shows model predictions for the associations between the EMA ratings and SPL, SNR, or HR in interaction with listening intent. There was a significant interaction between listening intent and SPL for EMA questions 1–5 (Table 2). That is, ratings performed during "speech communication" and "focused listening" were stronger (and negatively) associated with SPL than ratings made during "non-specific" listening (Figure 7A). The LME model for question 6 also revealed that the higher SPL was associated with the perception of increased noise (i.e., higher ratings) when listening focused as compared to the baseline condition (Figure 7B). Moreover, there were significant interactions between listening intent and SNR for all EMA questions. More precisely, ratings during "speech communication" and "focused listening" were more strongly (and positively) associated with SNR than during "non-specific" listening (Figure 7C). Also, the participants rated their surroundings as being less noisy (Figure 7D) when SNR values were higher while listening focused or to speech than when listening passively (i.e., "non-specific" listening).

Lastly, the models revealed that HR associated more with ratings during "speech communication" and "focused listening" than during "non-specific" listening for the EMA questions related to the general listening experience (Figure 7E and Table 3). For perceived noisiness (EMA question 6), ratings associated with HR

only for "speech communication" (Figure 7E and Table 3). In other words, when listening actively (to speech or focused), increased HR associated with better general listening experiences, while the perception of ambient noise during speech listening were associated with higher HR.

We again re-fitted the LME models for self-reported listening intents with data consisting of only a random subset (146 EMAs) of data from participants NH9 and NH15 to evaluate if the imbalance in number of completed EMAs among participants (Figure 2) influenced the results. As with the simpler LME models, we found only minor unsystematic changes in coefficient magnitudes but no alteration of the direction and significance of effects.

Discussion

The current study explored how acoustic factors and HR measurements in interaction with self-reported listening intents relate to real-world listening experiences in young adults with normal hearing.

Across listening intents, ambient SPL significantly and negatively associated with ratings from all EMA questions, which indicates that increased loudness during EMA completion is related to poorer listening experiences and increased perception of noisiness in the surroundings. Not surprisingly, the effect was strongest for EMA question 6 that asked participants to rate the noisiness of their surroundings on a scale from quiet to very noisy. Overall, these results indicate that the participants used the EMA scale correctly and that their experiences were reflected by the logged acoustic factors. This can be supported by significant and positive associations between SNR and the ratings for almost all individual EMA questions (except question 1). This indicates that the participants were sensitive to both loudness-related factors and the relative levels of modulated sound and background noise (in terms of higher SNR being associated with better ratings). These general patterns correspond well with results from previous EMA studies performed with hearing-device users that included acoustic data-logging (Andersson et al., 2021; Bosman et al., 2021; Pasta et al., 2022).

While EMAs in the current study are associated with mostly positive 5-min SNR values (Figure 4), this is also expected from previous investigations into typically encountered sound environments during daily life. For example, Pearsons et al. (1977) reported participants with normal hearing experienced mostly positive SNR levels while others report that hearing-aid users experience only few (<8%) moments in daily life with negative SNR (Smeds et al., 2015; Wu et al., 2018). Also, the study by Pearsons et al. (1977), demonstrated that increased ambient SPL was typically associated with decreased SNR for individuals with normal hearing. It should be noted that caution should be taken when comparing the absolute values of the reported SNRs with those from studies involving different measurement methodology as differences in estimation approach, frequency weighting and temporal averaging all could influence the levels.

In the current study, SPL and SNR showed to have positive moderate correlation. This is most likely because higher SPL only associated with decreased SNR in the presence of noise as reported by Christensen et al. (2021), while situations with both loud and clear sound would lead to positive correlations among SPL and SNR.

Moderating effect of listening intent

As hypothesized, we found that higher SPL and lower SNR were associated with poorer EMA ratings, and that this was moderated by listening intent. Namely, for listening intents related to "speech communication" or "focused listening," higher ambient SPL and lower ambient SNR led to poorer EMA ratings relative to "non-specific listening." Our study appears to be the first to report associations between acoustic data from HAs and subjective ratings of real-world listening experiences separated by listening intent. It supports previous findings concerning the relationship between acoustic factors and reported listening experiences alone (Andersson et al., 2021), and the introduction of the novel assessment dimension (listening intent) highlights the value of understanding not only in what conditions the participants are listening in, but also what their intentions are, to better account for their experiences. Even when encountering similar listening conditions as shown in Figure 4, individuals are reporting different assessments of their listening experiences depending on their listening intents.

While mean HR did not associate with EMA ratings when disregarding listening intents, the opposite was true when listening intents were included as a moderating factor. HR associated significantly with EMA ratings for questions 1–5 (general listening experiences) and 6 (perception of noisiness) when listening to speech or during "focused listening" but not for "non-specific" listening (Figure 7 and Table 3). This finding further highlights the value of self-reported listening intent, which here helps reveal associations between physiological responses and listening experiences.

Our findings support the theoretical framework of effortful listening offered by FUEL (Pichora-Fuller et al., 2016), which describes that increased listening effort is dependent on motivational factors (e.g., to understand what is being said) and increased listening demands (e.g., more challenging acoustic conditions). In accordance with the FUEL framework (Pichora-Fuller et al., 2016), we interpret the direction of the association (i.e., higher HR being related to better ratings) as indicative of the fact that increased HR reflects an increased willingness to put more effort into listening driven by motivation to hear what is going on (related to "focused listening") or by understanding what is being said (related to "speech communication"), ultimately leading to an achievement of listening success (e.g., better ratings). Further, the study by von Gablenz et al. (2021) has reported that the importance of hearing well was mostly related to "speech communication" listening intents, which can be linked with the motivation for achieving listening success. It seems reasonable to assume that questions related to specific listening intents or activities can reflect motivational factors when assessing real-world listening experiences. This corresponds well to the FUEL framework (Pichora-Fuller et al., 2016) which states that during increased listening demands (in terms of higher SPL and lower SNR) individuals are willing to invest or maintain certain level

10.3389/fauot.2023.1275210

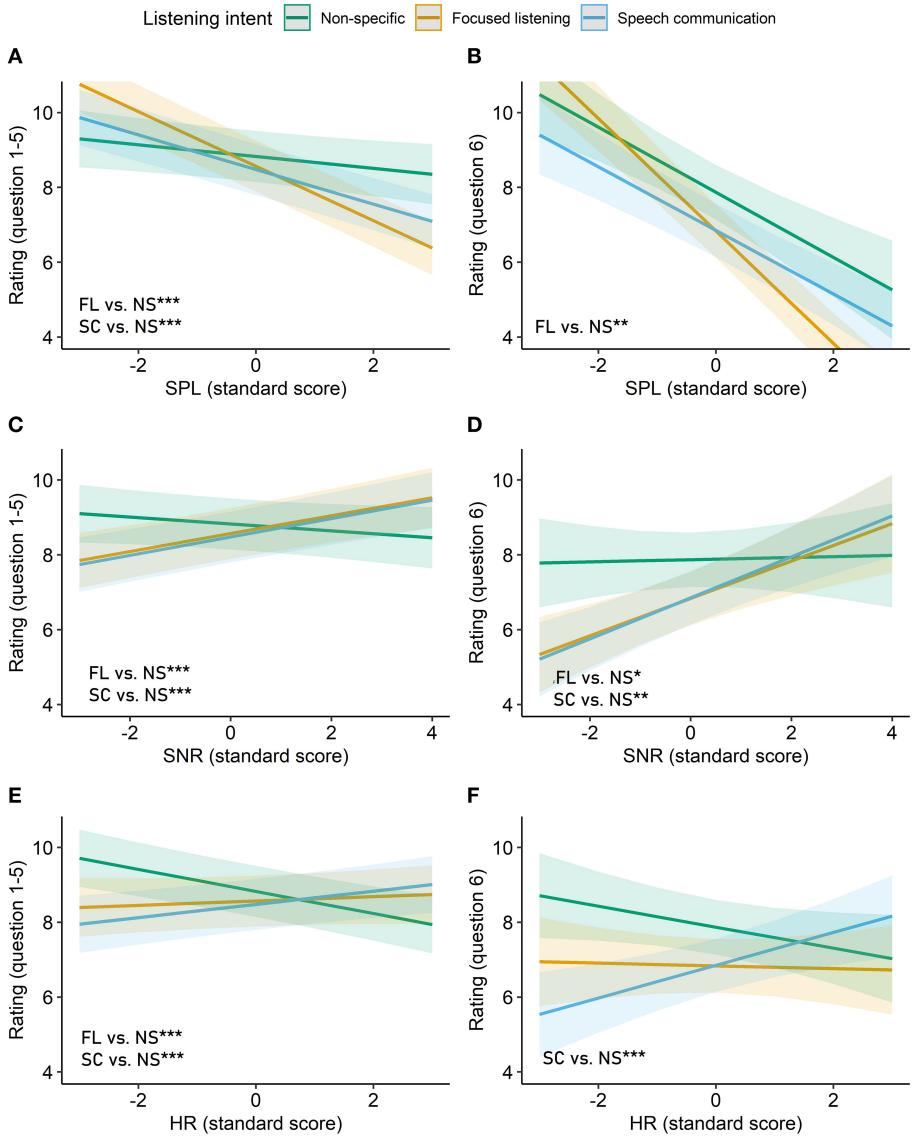


FIGURE 7
Interaction effects plot for standard deviation from grand mean EMA rating across different listening intents ("non-specific" as reference). **(A, C, E)** Represent LME model with participants and EMA questions (questions 1–5) as random effect. **(B, D, F)** Represent LME model of EMA question 6. Panel A and B show interaction effects plot for changes in SPL (sd). **(C, D)** Show interaction effects plot for changes in SNR (sd). **(E, F)** Show interaction effects plot for changes in HR (sd). The green lines represent "non-specific" listening intent. The yellow lines represent "focused listening" intent. The blue lines represent "speech communication" listening intent. Significant interactions are showed with *p*-values. Significance levels for *p*-values: $*p < 0.05$, $**p < 0.01$, $***p < 0.001$. EMA, ecological momentary assessment; SPL, sound pressure level; SNR, signal-to-noise ratio; HR, heart rate.

of listening effort. This willingness to exert effort is driven by motivation, such as the desire to comprehend speech effectively and engage in social interactions.

Limitations

We expected high compliance due to EMA prompts being sent out by both the smartphones and the wristbands every morning and throughout the day. Five participants reported intermittent technical issues with the Bluetooth connection between their HA and smartphone, which could have resulted in a reduced number of prompts and submitted EMAs. This is an inherent weakness of EMA studies, where participants are responsible for data collection and researchers have limited possibility to monitor the process. Only one participant felt that the EMA app sent too many notifications. Burke and Naylor (2020) reported that their participants with normal hearing ($N = 20$) completed 1,007 EMAs in total during a 2-week data collection period, which on average corresponded to 3.6 EMAs per participant per day. In the current study, compliance was much higher despite exclusion of EMAs related to "streamed" listening activities and possible technical issues (i.e., 6 EMAs per participant per day). Possibly, weekly visits to the laboratory combined with tactile notifications from the wristbands increased the study engagement. Moreover, as acoustic data-logging could not be monitored during

TABLE 3 Estimates for fixed-effect coefficients, confidence intervals, and *p*-values from linear mixed-effects models testing associations between acoustic data (SPL, SNR), HR, listening intent, and EMA ratings across EMA questions 1–5, and EMA question 6, respectively.

	EMA questions 1–5: general listening experience		EMA question 6: noisiness	
Coefficient	Estimate rating CI (95%)	*P*-value	Estimate rating CI (95%)	*P*-value
Intercept	8.82 (8.14 to 9.50)	**<0.001**	7.87 (7.14–8.60)	**<0.001**
SPL (SD)	−0.16 (−0.28 to −0.03)	**0.015**	−0.87 (−1.20 to −0.54)	**<0.001**
SNR (SD)	−0.09 (−0.21 to 0.02)	0.123	0.03 (−0.27 to 0.33)	0.852
HR (SD)	−0.29 (−0.41 to −0.18)	**<0.001**	−0.28 (−0.58 to 0.02)	0.066
Listening intent (FL vs. NS)	−0.26 (−0.38 to −0.14)	**<0.001**	−1.03 (−1.35 to −0.72)	**<0.001**
Listening intent (SC vs. NS)	−0.35 (−0.47 to −0.22)	**<0.001**	−1.02 (−1.34 to −0.70)	**<0.001**
SPL (SD): FL vs. NS	−0.57 (−0.72 to −0.42)	**<0.001**	−0.63 (−1.03 to −0.23)	**0.002**
SPL (SD): SC vs. NS	−0.31 (−0.46 to −0.15)	**<0.001**	0.02 (−0.39 to 0.43)	0.927
SNR (SD): FL vs. NS	0.33 (0.18 to 0.48)	**<0.001**	0.47 (0.07 to 0.87)	**0.020**
SNR (SD): SC vs. NS	0.34 (0.20 to 0.48)	**<0.001**	0.52 (0.15 to 0.89)	**0.006**
HR (SD): FL vs. NS	0.35 (0.23 to 0.47)	**<0.001**	0.24 (-0.07 to 0.56)	0.130
HR (SD): SC vs. NS	0.47 (0.36 to 0.58)	**<0.001**	0.72 (0.43 to 1.00)	**<0.001**
Random effects				
σ^2	2.21		2.99	
τ_{00}	1.35_{ID}		1.57_{ID}	
	0.07_{EMA}			
ICC	0.39		0.34	
N	13_{ID}		13_{ID}	
	5_{EMA}			
Observations	5,650		1,130	
Marginal R^2/Conditional R^2	0.072/0.435		0.255/0.511	

SPL, SNR, HR, and listening intent (referenced to "non-specific") were modeled as fixed effects, while the participants were modeled as random effects for EMA question 6. Model analyzing across EMA question 1–5 included participants and question number as random effects. ":" denotes an interaction. *P*-values in bold indicate that they are below the significance level (alpha = 0.05). SPL, sound level pressure; SNR, signal-to-noise-ratio; HR, heart rate; SD, standard deviation; EMA, ecological momentary assessment; NS, "non-specific"; FL, "focused listening"; SC, "speech communication." σ^2, variance; τ_{00}, between-subject variance; ICC, interclass correlation coefficient; N, sample size; ID, participant ID as random effect; Marginal R^2, variance of fixed effects; Conditional R^2, variance of both the fixed and random effects.

the study, noise, as produced by movement or contact with the HA microphone might have impacted the reported values. To minimize such issues, we provided detailed instructions to our participants in terms of proper handling of the test equipment, but we are unable to verify compliance. However, we expect that potential contributions from such noise artifacts would be spurious in time and not systematically correlated with listening intents. Thus, by relying on averaging several acoustic samples (e.g., 5-min averages) and performing association analysis (rather than statistically assess absolute levels), we believe confounds have been mitigated.

Another limitation for the current study was a sudden COVID-19 lockdown, which occurred while the data collection was ongoing (applying for eight participants included in the analyses). This resulted in less social interaction and thus reduced diversity of the listening situations experienced by the participants. It could have also affected their motivation for completing EMAs over time. Also, some participants reported that they were not always able to complete EMAs as this was inappropriate for them to do while being at work. This is relevant to consider when designing EMA studies with younger individuals as they may encounter other social contexts than older adults.

In future studies, we suggest exploring additional dimensions of motivational factors to further deepen the understanding of how listening experiences are related to intents in interaction with intrinsic motivational factors. This could involve examining how individual perceptions of the importance of hearing well in various situations relate to these experiences, a factor not explored in our current study. Additionally, future research could explore the feasibility of collecting other physiological measures besides HR in real-world settings and how they might be linked to EMA outcomes.

Conclusions

Real-world listening situations are characterized by high variability, and similar acoustic conditions can result in different self-reported listening experiences at the individual

level. The current study found that acoustic and HR data-loggings can improve the prediction of real-world self-reported listening experiences in young adults with normal hearing. Furthermore, it found that listening intent can influence self-reported real-world listening experiences, and that listening intent is associated with both acoustic factors and HR measurements. Overall, increased HR was associated with better self-reported listening experiences during "speech communication" as compared to non-specific listening situations. These findings indicate that the value of including *in-situ* HR measures in EMAs depend on the ability to also discriminate between listening intentions.

Data availability statement

There are ethical restrictions on publicly sharing the dataset. The consent given by users did not explicitly detail sharing of the data in any format; this limitation is in keeping with EU General Data Protection Regulation and is imposed by the Research Ethics Committees of the Capital Region of Denmark. Data can be obtained by contacting the corresponding author and signing a nondisclosure agreement.

Ethics statement

The requirement of ethical approval was waived by Regional Committee on Health Research Ethics for Southern Denmark for the studies involving humans because Regional Committee on Health Research Ethics for Southern Denmark. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

KA: Conceptualization, Formal analysis, Investigation, Methodology, Visualization, Writing—original draft, Writing—review & editing. TN: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing—review & editing. JC: Conceptualization, Data curation, Formal analysis, Methodology, Software, Supervision, Visualization, Writing—review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The current study was supported by a Ph.D. stipend from the William Demant Foundation (case no. 19-4068).

Conflict of interest

JC was employed by Oticon A/S.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Alhanbali, S., Dawes, P., Lloyd, S., and Munro, K. J. (2018). Hearing handicap and speech recognition correlate with self-reported listening effort and fatigue. *Ear Hear.* 39, 470–474. doi: 10.1097/AUD.0000000000000515

Alhanbali, S., Dawes, P., Millman, R. E., and Munro, K. J. (2019). Measures of listening effort are multidimensional. *Ear Hear.* 40, 1084–1097. doi: 10.1097/AUD.0000000000000697

Andersson, K. E., Andersen, L. S., Christensen, J. H., and Neher, T. (2021). Assessing real-life benefit from hearing-aid noise management: SSQ12 questionnaire versus ecological momentary assessment with acoustic data-logging. *Am. J. Audiol.* 30, 93–104. doi: 10.1044/2020_AJA-20-00042

Avram, R., Tison, G. H., Aschbacher, K., Kuhar, P., Vittinghoff, E., Butzner, M., et al. (2019). Real-world HR norms in the Health eHeart study. *Npj Digit. Med.* 2, 58. doi: 10.1038/s41746-019-0134-9

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *J. Mem. Lang.* 68, 255–278. doi: 10.1016/j.jml.2012.11.001

Bent, B., Goldstein, B. A., Kibbe, W. A., and Dunn, J. P. (2020). Investigating sources of inaccuracy in wearable optical HR sensors. *Npj Digit. Med.* 3, 18. doi: 10.1038/s41746-020-0226-6

Bosman, A. J., Christensen, J. H., Rosenbom, T., Patou, F., Janssen, A., and Hol, M. K. S. (2021). Investigating real-world benefits of high-frequency gain in bone-anchored users with ecological momentary assessment and real-time data logging. *J. Clin. Med.* 10, 3923. doi: 10.3390/jcm10173923

Burke, L. A., and Naylor, G. (2020). Daily-life fatigue in mild to moderate hearing impairment: an ecological momentary assessment study. *Ear Hear.* 41, 1518–1532. doi: 10.1097/AUD.0000000000000888

Christensen, J., Andersson, K., and Neher, T. (2023). "Distinct influence of everyday noise on cardiovascular stress," in *INTER-NOISE and NOISE-CON Congress and Conference Proceedings, Vol. 265* (Glasgow), 242–247. doi: 10.3397/IN_2022_0038

Christensen, J. H., Saunders, G. H., Porsbo, M., and Pontoppidan, N. H. (2021). The everyday acoustic environment and its association with human HR: evidence from real-world data logging with hearing aids and wearables. *R. Soc. Open Sci.* 8, 201345. doi: 10.1098/rsos.201345

Dillon, H. (2012). *Hearing Aids, 2nd Edn*. Sydney, NSW: Boomerang Press.

El Aarbaoui, T., and Chaix, B. (2019). The short-term association between exposure to noise and heart rate variability in daily locations and mobility contexts. *J. Exposure Sci. Environ. Epidemiol.* 30, 383–393. doi: 10.1038/s41370-019-0158-x

El Aarbaoui, T., Méline, J., and Brondeel, R., Chaix, B. (2017). Short-term association between personal 566 exposure to noise and HR variability: the RECORD MultiSensor Study. *Environ. Pollut.* 231, 703–711. doi: 10.1016/j.envpol.2017.08.031

Francis, A. L., Bent, T., Schumaker, J., Love, J., and Silbert, N. (2021). Listener characteristics differentially affect self-reported and physiological measures of effort

associated with two challenging listening conditions. *Attent. Percept. Psychophys.* 83, 1818–1841. doi: 10.3758/s13414-020-02195-9

Giuliani, N. P., Brown, C. J., and Wu, Y.-H. (2021). Comparisons of the sensitivity and reliability of multiple measures of listening effort. *Ear Hear.* 42, 465–474. doi: 10.1097/AUD.0000000000000950

Gnewikow, D., Ricketts, T., Bratt, G. W., and Mutchler, L. C. (2009). Real-world benefit from directional microphone hearing aids. *J. Rehabil. Res. Dev.* 46, 603. doi: 10.1682/JRRD.2007.03.0052

Harrison, X. A., Donaldson, L., Correa-Cano, M. E., Evans, J., Fisher, D. N., Goodwin, C. E. D., et al. (2018). A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ* 6, e4794. doi: 10.7717/peerj.4794

Henry, K. S., and Heinz, M. G. (2012). Diminished temporal coding with sensorineural hearing loss emerges in background noise. *Nat. Neurosci.* 15, 1362–1364. doi: 10.1038/nn.3216

Holube, I., von Gablenz, P., and Bitzer, J. (2020). Ecological momentary assessment in hearing research: current state, challenges, and future directions. *Ear Hear.* 41, 79S—90S. doi: 10.1097/AUD.0000000000000934

Humes, L. E., Rogers, S. E., Main, A. K., and Kinney, D. L. (2018). The acoustic environments in which older adults wear their hearing aids: insights from datalogging sound environment classification. *Am. J. Audiol.* 27, 594–603. doi: 10.1044/2018_AJA-18-0061

Jensen, N. S., and Nielsen, C. (2005). "Auditory ecology in a group of experienced hearing-aid users: Can knowledge about hearing-aid users' auditory ecology improve their rehabilitation?," in *Proceedings of the 21st Danavox Symposium* (Kolding), 235–258.

Keidser, G., Naylor, G., Brungart, D. S., Caduff, A., Campos, J., Carlile, S., et al. (2020). The quest for ecological validity in hearing science: what it is, why it matters, and how to advance it. *Ear Hear.* 41, 5S—19S. doi: 10.1097/AUD.0000000000000944

Mackersie, C. L., and Calderon-Moultrie, N. (2016). Autonomic nervous system reactivity during speech repetition tasks: HR variability and skin conductance. *Ear Hear.* 37, 118S—125S. doi: 10.1097/AUD.0000000000000305

Nelson, B. W., Low, C. A., Jacobson, N., Areán, P., Torous, J., and Allen, N. B. (2020). Guidelines for wrist-worn consumer wearable assessment of HR in biobehavioral research. *Npj Digit. Med.* 3, 90. doi: 10.1038/s41746-020-0297-4

Ohlenforst, B., Zekveld, A. A., Jansma, E. P., Wang, Y., Naylor, G., Lorens, A., et al. (2017). Effects of hearing impairment and hearing aid amplification on listening effort: a systematic review. *Ear Hear.* 38, 267–281. doi: 10.1097/AUD.0000000000000396

Oleson, J. J., Jones, M. A., Jorgensen, E. J., and Wu, Y.-H. (2022). Statistical considerations for analyzing ecological momentary assessment data. *J. Speech Lang. Hear. Res.* 65, 344–360. doi: 10.1044/2021_JSLHR-21-00081

Pasta, A., Petersen, M. K., Jensen, K. J., Pontoppidan, N. H., Larsen, J. E., and Christensen, J. H. (2022). Measuring and modeling context-dependent preferences for hearing aid settings. *User Model. User Adapt. Interact.* 32, 977–998. doi: 10.1007/s11257-022-09324-z

Pearsons, K. S., Bennett, R. L., and Fidell, S. A. (1977). *Speech Levels in Various Noise Environments Research Reporting Series: Environmental Health Effects Research*. Office of Health and Ecological Effects, Office of Research and Development, U.S. EPA.

Peelle, J. E. (2018). Listening effort: how the cognitive consequences of acoustic challenge are reflected in brain and behavior. *Ear Hear.* 39, 204–214. doi: 10.1097/AUD.0000000000000494

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., Humes, L. E., et al. (2016). Hearing impairment and cognitive energy: the framework for understanding effortful listening (FUEL). *Ear Hear.* 37, 5S—27S. doi: 10.1097/AUD.0000000000000312

Schinkel-Bielefeld, N., Kunz, P., Zutz, A., and Buder, B. (2020). Evaluation of hearing aids in everyday life using ecological momentary assessment: what situations are we missing? *Am. J. Audiol.* 29, 591–609. doi: 10.1044/2020_AJA-19-00075

Shiffman, S., Stone, A. A., and Hufford, M. R. (2008). Ecological momentary assessment. *Annu. Rev. Clin. Psychol.* 4, 1–32. doi: 10.1146/annurev.clinpsy.3.022806.091415

Smeds, K., Wolters, F., and Rung, M. (2015). Estimation of signal-to-noise ratios in realistic sound scenarios. *J. Am. Acad. Audiol.* 26, 183–196. doi: 10.3766/jaaa.26.2.7

von Gablenz, P., Kowalk, U., Bitzer, J., Meis, M., and Holube, I. (2021). Individual hearing aid benefit in real life evaluated using ecological momentary assessment. *Trends Hear.* 25, 233121652199028. doi: 10.1177/2331216521990288

Wolters, F., Smeds, K., Schmidt, E., Christensen, E. K., and Norup, C. (2016). Common sound scenarios: a context-driven categorization of everyday sound environments for application in hearing-device research. *J. Am. Acad. Audiol.* 27, 527–540. doi: 10.3766/jaaa.15105

Wu, Y.-H., Stangl, E., Chipara, O., Hasan, S. S., DeVries, S., and Oleson, J. (2019). Efficacy and effectiveness of advanced hearing aid directional and noise reduction technologies for older adults with mild to moderate hearing loss. *Ear Hear.* 40, 805–822. doi: 10.1097/AUD.0000000000000672

Wu, Y.-H., Stangl, E., Chipara, O., Hasan, S. S., Welhaven, A., and Oleson, J. (2018). Characteristics of real-world signal to noise ratios and speech listening situations of older adults with mild to moderate hearing loss. *Ear Hear.* 39, 293–304. doi: 10.1097/AUD.0000000000000486

Wu, Y.-H., Xu, J., Stangl, E., Pentony, S., Vyas, D., Chipara, O., et al. (2021). Why ecological momentary assessment surveys go incomplete: when it happens and how it impacts data. *J. Am. Acad. Audiol.* 32, 16–26. doi: 10.1055/s-0040-1719135

Zekveld, A. A., Koelewijn, T., and Kramer, S. E. (2018). The pupil dilation response to auditory stimuli: current state of knowledge. *Trends Hear.* 22, 2331216518777174. doi: 10.1177/2331216518777174

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 21 December 2023
DOI 10.3389/fnins.2023.1292395

Check for updates

OPEN ACCESS

EDITED BY
Faheema Mahomed-Asmail,
University of Pretoria, South Africa

REVIEWED BY
Kumar Seluakumaran,
University of Malaya, Malaysia
Yuanchia Chu,
Taipei Veterans General Hospital, Taiwan

*CORRESPONDENCE
Xinxing Fu
✉ xinxing.fu@research.uwa.edu.au
Shuo Wang
✉ shannonwsh@aliyun.com

RECEIVED 11 September 2023
ACCEPTED 06 December 2023
PUBLISHED 21 December 2023

CITATION
Liu H, Fu X, Li M and Wang S (2023) Comparisons of air-conduction hearing thresholds between manual and automated methods in a commercial audiometer. *Front. Neurosci.* 17:1292395. doi: 10.3389/fnins.2023.1292395

COPYRIGHT
© 2023 Liu, Fu, Li and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Comparisons of air-conduction hearing thresholds between manual and automated methods in a commercial audiometer

Hui Liu[1,2], Xinxing Fu[1,2,3,4]*, Mohan Li[1,2] and Shuo Wang[1,2]*

[1]Department of Otolaryngology, Head and Neck Surgery, Beijing Tongren Hospital, Capital Medical University, Beijing, China, [2]Key Laboratory of Otolaryngology, Head and Neck Surgery, Ministry of Education, Beijing Institute of Otolaryngology, Beijing, China, [3]Medical School, The University of Western Australia, Crawley, WA, Australia, [4]Ear Science Institute Australia, Subiaco, WA, Australia

Objective: To investigate the correlation of air-conduction thresholds between automated audiometry in a non-isolated environment and manual audiometry in participants with normal hearing and different degrees of hearing loss.

Methods: Eighty-three participants aged 11–88 years old underwent automated pure-tone audiometry in a non-acoustically isolated environment, and the results were compared with those of manual pure-tone audiometry performed in a standard acoustically isolated booth, with the order of testing randomised. Six frequencies of 250, 500, 1,000, 2000, 4,000 and 8,000 Hz were tested.

Results: All 166 ears were completed and 996 valid hearing threshold data were obtained, with 28 data exceeding the 95% confidence interval in the Bland–Altman plot, accounting for 2.81% of all data. The means and standard deviations of the differences for the six frequencies from 250 to 8,000 Hz were, respectively, 0.63 ± 5.31, 0.69 ± 4.50, 0.45 ± 4.99, 0.3 ± 6.2, -0.15 ± 4.8, and 0.21 ± 4.97 dB. The correlation coefficients of the two test results for normal hearing, mild, moderate, severe and above hearing loss groups were 0.95, 0.92, 0.97, and 0.96, respectively. The correlation coefficient of the automated and manual audiometry thresholds for the age groups under 40 years, 40–60 years, and 60 years above, were 0.98, 0.97 and 0.97, respectively, with all being statistically significant ($p < 0.01$). The response time of the three age groups were 791 ± 181 ms, 900 ± 190 ms and $1{,}063 \pm 332$ ms, respectively, and there was a significant difference between the groups under 40 years and over 60 years.

Conclusion: There was good consistency between automated pure-tone audiometry in a non-acoustically isolated environment and manual pure-tone audiometry in participants with different hearing levels and different age groups.

KEYWORDS

automated audiometry, pure-tone audiometry, KUDUwave, response time, non-soundproof booth

1 Introduction

WHO estimates by 2050, nearly 2.5 billion people will suffer from some degree of hearing loss, of whom at least 700 million will need rehabilitation services (World Health Organization, 2021). In addition to its impact on interpersonal communication, psychosocial well-being and quality of life, hearing loss has a significant socio-economic impact. In children, hearing loss can limit language development and lead to difficulties in social integration and access to education, with significant impacts on the family; in adults, hearing loss can lead to higher unemployment rates and social isolation (Kramer et al., 2006). In older adults, hearing loss is also associated with cognitive decline and dementia (Livingston et al., 2017). In China, according to the results of the second national sample survey of people with disabilities, there are 27.8 million people with hearing disabilities, ranking first among the five major disabilities (Sun et al., 2008). In recent years, the number of people with hearing loss has increased with the increase in population aging. Early detection, diagnosis, and intervention can reduce the socioeconomic burden of hearing loss.

Pure-tone audiometry is the most basic and important method of assessing hearing loss. Traditional manual testing methods for pure-tone audiometry require three conditions to be met: a compliant acoustic isolation room, calibrated audiometers, and professionally trained audiologists. In China, most tertiary hospitals in first and second-tier cities can fulfil these conditions for testing, but in remote and economically underdeveloped areas, there are a limited number of hospitals that can fulfil the conditions for testing, which means that it is difficult for many people to access hearing healthcare, and at the same time, the large group of patients puts tertiary hospitals under even greater pressure.

Automated pure-tone audiometry means hearing threshold testing where the testing process is automated with no or minimal staff involvement (Wasmann et al., 2022). A growing body of research suggests that automated pure-tone audiometry can be useful in mass hearing screening, in remote and economically underdeveloped areas (Visagie et al., 2015; Eksteen et al., 2019; Sandström et al., 2020). There are three approaches to automate audiometry, including software solutions such as the AMTAS (Automated Method for Testing Auditory Sensitivity) and the Home Hearing Test (HHT); hardware solutions such as the KUDUwave portable audiometer; and smartphone/tablet solutions such as the hearScreen and hearTest application (Shojaeemend and Ayatollahi, 2018).

An automated pure tone audiometer for complete diagnostic testing purposes needs to include air conduction testing, bone conduction testing, masking techniques, and controling the noise attenuation. The KUDUwave 5,000 audiometer (hereinafter referred to as KUDUwave) is a portable audiometer that performs air-conducted and bone-conducted pure tone hearing threshold tests in automated and manual modes, with masking when required, by insert earphones covered by circumaural earcups to increase ambient noise attenuation, and continuous monitoring of ambient noise and determination of the amount of attenuation by using microphones inside and outside the circumaural earcups. This combination of attenuation and monitoring allows hearing tests to be performed in non-acoustically isolated environments, ensuring that pure tone thresholds can be tested down to 0 dB HL at maximum permissible ambient noise levels (MPANLs) of 70, 69, 58, 53, 50, 59, and 59 dB SPL for octaves from 0.125-8 k Hz. The audiometers are connected to a computer via a USB port with Internet access for remote hearing tests.

Existing studies reported good correlations between the results of automated and traditional manual pure-tone audiometry, both in adults and children using KUDUwave in sound-insulated and non-insulated environments. Swanepoel et al. conducted automated pure-tone audiometry in non-sound-insulated environments using KUDUwave in 23 adults with normal hearing (Swanepoel De et al., 2015) and in 149 children (Swanepoel De et al., 2013), and obtained reliable results when compared to traditional manual pure-tone audiometry. Maclennan-Smith et al. (Maclennan-Smith et al., 2013) performed automated versus manual testing of the KUDUwave on 147 older adults with normal hearing or varying degrees of hearing loss. The automated test was performed in a normal room, and the manual test was performed in an acoustically insulated room, with 95% of the threshold difference in air-conducted (250–8,000 Hz) and 86% of the threshold difference in bone-conducted (250–4,000 Hz) were within 5 dB. Swanepoel et al. (Swanepoel De et al., 2010a) and Visagie et al. (Visagie et al., 2015) also reported remote pure-tone audiometry using KUDUwave that reliable test results were obtained.

Governder and Mars (Govender and Mars, 2018a) conducted hearing screening in a group of children aged 6–12 years in a rural primary school and those who failed the screening underwent diagnostic audiometry, both screening and diagnostic audiometry were conducted using KUDUwave. The results showed high specificity (100%) but low sensitivity (65.2%) for automated pure-tone audiometric screening. The 1,500 ms suggested by KUDUwave was used as the reaction time, and Governder and Mars concluded that this reaction time might be insufficient for child subjects. It is proposed that the response time of the subjects should be investigated, and the parameters of the device should be adjusted. Storey et al. (Storey et al., 2014) measured 31 subjects (aged 15 to 80 years) with different degrees of hearing loss using the KUDUwave in quiet and noisy environments, most of the thresholds obtained were within ±5 dB of the results of the manual pure-tone audiometry in an acoustic chamber (89 and 92% in quiet and noisy environments, respectively). However, thresholds obtained with the KUDUwave in 5% of the test ears showed large differences compared to clinical audiometers, with differences in thresholds up to 60 dB.

This study analysed subjects of different ages and degrees of hearing loss in groups and reported the response times of subjects of different ages. It is expected to provide evidence for setting parameters for automated pure-tone audiometry.

2 Materials and methods

2.1 Participants

Eighty-three participants, 41 males and 42 females, aged 11–88 years (median age was 57 years), were enrolled from the clinical audiology centre, Department of Otorhinolaryngology, Head and Neck Surgery, Beijing Tongren Hospital. Inclusion criteria: ability to understand the test requirements and cooperate in completing the test; including normal hearing and varying degrees of sensorineural, conductive and mixed hearing loss. Exclusion criteria: known cognitive impairment and inability to understand the test requirements. This research project was approved by the Medical

Ethics Committee for Clinical Research of Beijing Tongren Hospital, and informed consent was obtained from the participants before the tests.

2.2 Equipment

Both manual and automated pure-tone audiometry were performed using the KUDUwave 5,000 (GeoAxon, Pretoria) clinical audiometer. KUDUwave was connected to a computer via a USB port, and the test procedure was operated by software installed on a laptop computer. Before testing with the KUDUwave, it was calibrated according to ISO 389-2: 1994. The B&K 2240 (Brüel & Kjær, Denmark) sound level meter was used to monitor the clinic's environmental noise, recording the average and maximum noise values.

2.3 Test methods

An otoscopic examination of the subject's external ear canal was conducted to remove possible cerumen obstruction. All participants were tested for pure-tone air-conduction hearing thresholds by manual and automated methods, in a randomised order, with adequate rest given between each test. Test frequencies were 250, 500, 1,000, 2000, 4,000 and 8,000 Hz. The test requirements were fully explained to the participant before the tests. The KUDUwave insert earphones were fully into the ear canal and the end flush with the tragus, and then the circumaural earphones were placed over the insert earphones. Before the test, a pure-tone signal sound was given manually for the subject to practice. The subjects were instructed to press the button as soon as they heard the pure tone, and to perform manual or automated audiometry after the subjects had fully understood the test requirements.

Manual pure-tone audiometry was performed in a standard double-walled soundproof booth with the KUDUwave. The manual test determined the hearing threshold according to the method specified in ISO 8253-1:2010. The automated pure-tone audiometry was conducted using the shortened ascending method (ISO 8253-1:2010). The initial sound intensity for each frequency was 30 dB HL, and the sound duration was 1,000 ms. The waiting response time is 2000 ms, i.e., it's considered to be a valid response to press the transponder button within 2000 ms from the time the tone is given, otherwise it will be marked as a false positive response. At the end of the test, a pure-tone audiogram was automatically generated, while KUDUwave reported the percentage of false positives, the noise monitoring value, the number of times the subject responded to the signal, and the response time to press the button. The automated test was conducted in a general clinic room, and the average and maximum values of ambient noise were monitored with a sound level meter during the test. A comparison of manual and automated hearing testing protocols is shown in eTable 1. To avoid the audiologist referring to the results of the first test for a second test, separate audiologists operated manual and automated tests and were unaware of each other's results. Meanwhile, in order to minimize the variability, the instructions remained the same between the two audiologists. The test procedure is shown in eFigure 1.

2.4 Data processing

Descriptive measures illustrated the difference between the thresholds of manual and automated.

audiometry, described as mean ± SD. An independent samples t-test was performed on the difference in thresholds from 250 to 8,000 Hz obtained by the two testing methods, with $p < 0.05$ as the criterion for significance. Pearson correlation tests were used to assess whether there was a correlation between manual and automated test results. One-way ANOVA was used to compare the thresholds between different age groups and groups with different hearing loss degrees. The post-hoc power analysis was run to confirm the sample size. The difference between the two test methods was analysed using Bland–Altman plots. All statistical analyses were performed by SPSS 25 (SPSS Inc., Chicago, Illinois, USA).

3 Result

A total of 166 ears were obtained from 83 participants, with 6 frequencies tested in each ear for a total of 996 data. A total of 28 data exceeded the 95% upper and lower limits of the Bland–Altman plots, accounting for 2.8% of all the data, less than 5%, indicating good consistency between manual and automated pure-tone audiometry results (Figure 1). The Bland–Altman plots of the separate frequencies from 250, 500, 1,000, 2000, 4,000 and 8,000 Hz were listed as supplementary material (eFigures 2A-F). The post-hoc power analysis was run to confirm the sufficient sample size, the calculation results were listed as supplementary material (eTable 2).

The difference between the automated and manual pure-tone audiometry results and the absolute value of the difference were shown in Table 1, the maximum value of the absolute value of the difference at each frequency is between 20 and 35 dB. The distribution of the absolute difference between the manual and the automated thresholds was shown in eFigure 3.

All participants were divided into four groups according to better ear average hearing level (mean values of hearing thresholds at four frequencies, 500, 1,000, 2000, and 4,000 Hz), the normal group (≤25 dB HL), the mild hearing loss group (26–40 dB HL), the moderate hearing loss group (41–60 dB HL), and the severe and above hearing loss group (≥61 dB HL) (Olusanya et al., 2019). The correlation coefficient (r) between the automated and the manual test in the mild hearing loss group was 0.92 while the correlation coefficients were equal to or greater than 0.95 in all other groups, all with significance ($p < 0.01$). Threshold differences were not statistically significant between the groups ($p > 0.05$). The correlation between the automated and manual test results for participants with different hearing levels is shown in Table 2.

All participants were also divided into three groups according to age, the group under 40 years, 40–60 years and over 60 years, and the correlation coefficient (r) between the automated and the manual test for each group was greater than 0.9, all with significance ($p < 0.01$) (Table 3). The demographic of the participants is shown in eTable 3.

The test durations for automated and manual pure tone audiometry were 320 ± 42 s and 281 ± 90 s, respectively, with the manual test time being less than the automatic test time ($p < 0.05$). The average value of ambient noise in the general clinic was

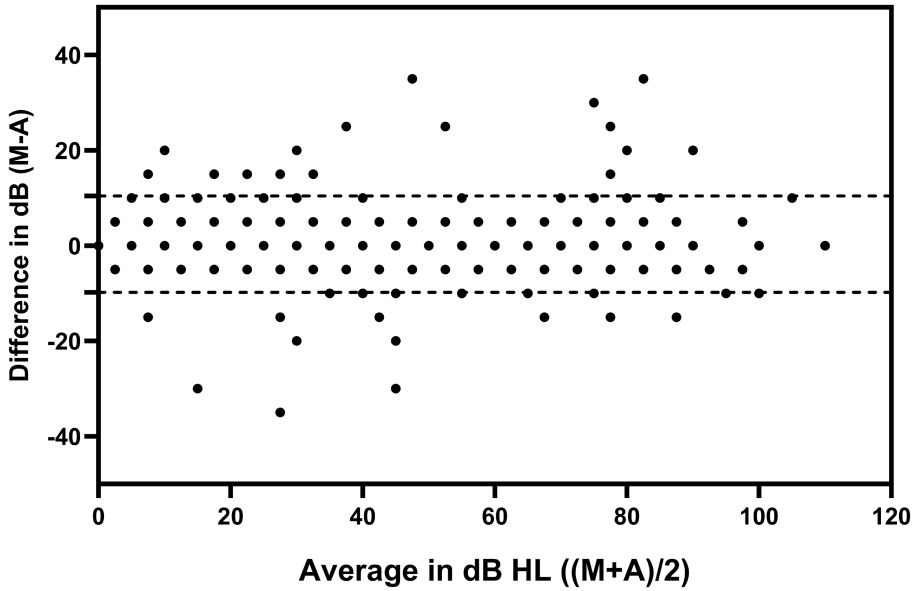


FIGURE 1
Bland–Altman plots of the results of the automated and manual pure-tone audiometry. The dotted lines in the plots are the upper and lower limits of the 95% confidence intervals. M: manual pure-tone audiometry threshold; A: automated pure-tone audiometry threshold. Dots were pooled across both test ears and all test frequencies.

TABLE 1 The difference between manual and automated audiometry thresholds for each frequency.

Hz	250	500	1,000	2000	4,000	8,000	Total
M difference in dB (SD)	0.63 (5.31)	0.69 (4.50)	0.45 (4.98)	0.30 (6.22)	−0.15 (4.84)	0.21 (4.97)	0.36 (5.16)
Abs M difference in dB (SD)	3.16 (4.31)	2.74 (3.63)	2.74 (4.18)	3.31 (5.27)	2.20 (4.32)	2.44 (4.33)	2.77 (4.37)
Maximum of Abs difference	25	20	35	30	35	35	35

M difference: the average value of the difference between the manual and the automated thresholds (manual minus automated values); M difference: the average of the absolute value of the difference between the manual and the automated threshold; Maximum of Abs difference: the maximum of the absolute value of the difference between the manual and the automated threshold.

41.5 ± 4.6 dB(A), and the maximum value of ambient noise was 66.2 ± 7.2 dB(A).

False positives of automated audiometry were reported from KUDUwave, and the results were listed as supplementary material (eTable 4). The response time for subjects to press the transponder button increased with age (Figure 2), and the overall response time for all subjects was 941.5 ± 279.3 ms. The response time was 791.5 ± 181.2 ms in the age group below 40 years, 900.4 ± 190.9 ms in the age group 40–60 years, and 1063.1 ± 332.3 ms in the group 60 years above. There was a statistical difference in response time between the under 40 years group and the over 60 years group ($p < 0.01$). The results were listed as supplementary material (eTable 5).

4 Discussion

In our previous study, a comparison of automated and manual pure-tone audiometry was performed on normal hearing subjects, and a good correlation was found between the results of the two tests (Liu et al., 2021), which is consistent with the findings of other studies (Swanepoel De et al., 2010b; Mahomed et al., 2013; Corry et al., 2017). However, there are fewer studies on the correlation between automated and manual pure-tone audiometry in participants with hearing loss. Measurement bias might be introduced by including only participants with normal hearing (Rutjes et al., 2006). Participants with normal hearing are known to have hearing within a specific range, which will reduce the possible range of variation between the two diagnostic techniques. Reducing the measurement bias by testing on clinical patients would provide a more valid estimate of the accuracy of automated pure-tone audiometry in practice (Whitton et al., 2016). The participants in this study were drawn from the clinical audiology centre, covering a wide range of hearing loss conditions of varying degrees and natures, and the range of ages is large, simulating the more common situations that occur in clinical audiometry.

The overall consistency between automated and manual audiometry was good, with an outlier of 2.8%, less than 5%. The absolute maximum value of the threshold difference between the two methods ranged from 20 to 35 dB. These results are consistent with

TABLE 2 Correlation of manual and automated pure-tone audiometry thresholds for different hearing levels.

	Groups by hearing level				Total (n = 83)
	Normal (n = 26)	Mild (n = 14)	Moderate (n = 30)	Severe and above (n = 13)	
M difference in dB (SD)	0.77 (3.94)	1.19 (6.43)	−0.35 (3.60)	−0.76 (5.29)	0.20 (4.60)
r	0.95	0.92	0.97	0.96	0.98
p-values	<0.01	<0.01	<0.01	<0.01	<0.01

M difference: the average value of the difference between the manual and the automated thresholds (manual minus automated values); r: correlation coefficient.

TABLE 3 Correlation between manual and automated pure-tone audiometry thresholds in different age groups.

	Age groups			Total (n = 83)
	<40 years (n = 21)	40 ~ 60 years (n = 27)	>60 years (n = 35)	
M difference in dB (SD)	0.83 (4.88)	−0.20 (5.61)	0.50 (4.93)	0.36 (5.16)
r	0.98	0.97	0.97	0.98
p-values	<0.01	<0.01	<0.01	<0.01

M difference: the average value of the difference between the manual and the automated thresholds (manual minus automated values); r: correlation coefficient.

previous studies that have shown strong benefits of automated pure-tone audiometry in screening for large-scale hearing modalities (Mahomed et al., 2013; Wasmann et al., 2022).

In this study, the correlation between automated and manual pure-tone audiometry was comparable in the groups with different hearing levels, demonstrating that automated audiometry can obtain reliable results in people with various degrees of hearing loss. In other studies of automated pure-tone audiometry, due to the difficulty of controlling ambient noise and the calibration of headphones, the application scenario is mainly for self-hearing healthcare monitoring at home, which is not available for screening mild hearing loss (Whitton et al., 2016; Sandström et al., 2020; Wasmann et al., 2022). There are a limited number of studies on automated pure-tone audiometry in participants with moderate and above hearing loss, where subjects are not grouped by degree of hearing loss. Brennan-Jones recruited 42 participants with different degrees of hearing loss to conduct a correlation study between automated and manual pure tone audiometry, with 86.5% of the thresholds differing within 10 dB, and 94.8% of the thresholds differing within 15 dB, which is similar to this study, but no subgroup analysis of the degree of hearing loss was performed (Brennan-Jones et al., 2016). Whitton (Whitton et al., 2016) performed automated pure tone audiometry on 19 subjects with varying degrees of hearing loss, finding higher thresholds at 250 Hz when collected at home, and attributing this to background noise in the home environment, but did not group the degrees of hearing loss to see if this phenomenon occurred only in subjects with specific levels of hearing loss. Tonder, Govender, and Bornman all performed automated pure-tone audiometry of participants with different degrees of hearing loss, but none of them performed detailed subgroup analyses based on the degree of hearing loss (Van Tonder et al., 2017; Bornman et al., 2018; Govender and Mars, 2018b).

In the three age groups, the thresholds correlated well between automated and manual tests, with correlation coefficients above 0.9, confirming that automated audiometry can be carried out in the elderly population. Margolis et al. performed automated pure-tone audiometry in a non-isolated environment on 28 older adults with a mean age of 65 years, and the hearing thresholds were slightly higher than those of manual pure-tone audiometry obtained in a sound-isolation room, but no statistical differences were observed (Margolis et al., 2016). In a similar study by Mosley, the mean hearing thresholds for four frequencies were correlated between automated and manual pure-tone audiometry in 112 older adults aged 60 years or older, as well as in different degrees of hearing loss (Mosley et al., 2019).

In addition to false-positive responses, which are the most common phenomenon affecting the reliability of test results, observing other indicators can help determine the test reliability. Margolis (Margolis et al., 2007) suggested a method for predicting the accuracy of automated audiometry thresholds (Qualind™), a multiple regression analysis of eight factors associated with test accuracy, including masked alarm rate, time per trial, false-positive rate, false-negative rate, mean test–retest variance, the number of air-bone gaps >50 dB, the number of air-bone gaps <−10 dB, and the mean air-bone gap, yielded a regression coefficient of 0.84. Not all of these eight factors were available in this study and therefore could not be cross-validated with Margolis' results. Therefore, more metrics with higher sensitivity and specificity still need to be explored for validation of individual quality control in automated pure-tone audiometry.

The response time of the participants to press the transponder button after hearing the sound was positively correlated with age, and the overall response time was 941.5 ± 279.3 ms. Significant differences were observed between the groups under 40 years and over 60 years, which may be explained by the gradual decline of brain function with age. The reaction time for all subjects in this study was set to 2000 ms or less; if the reaction time is set too long, a portion of the false positives may be included in the correct response, which will affect the ability to obtain accurate automated audiometric results. Samantha et al. (Govender and Mars, 2018a) set the reaction time to 1,500 ms in a group of children aged 6–12 years old for hearing screening, and the authors concluded that the reaction time may be insufficient for child subjects. The results of this study showed that the reaction time did not exceed 1,500 ms for all subjects. Still, the minimum age of the subjects in this study was 11 years old, which does not cover the

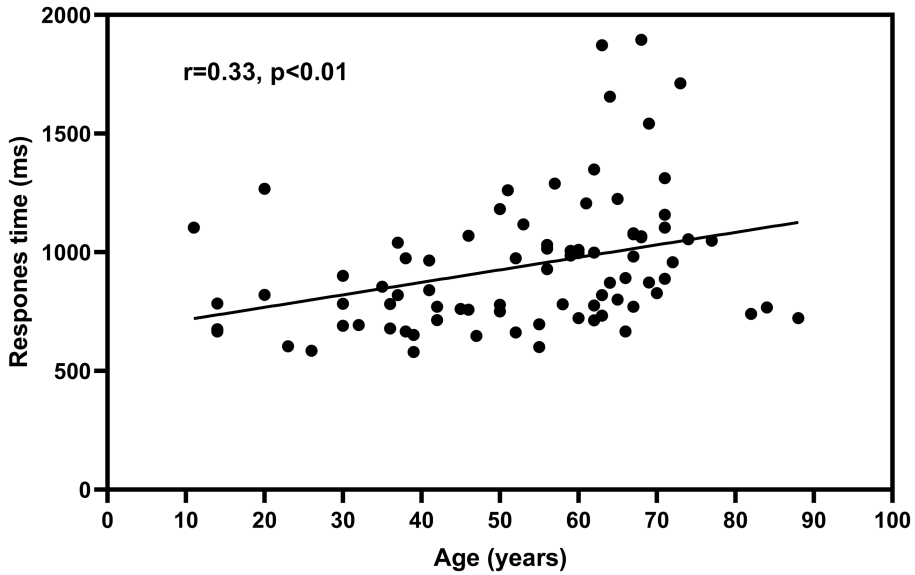


FIGURE 2
Relationship between participants' age and response time. The fitted curve in the figure is based on linear regression. Data were pooled across both test ears and all test frequencies.

subject population in the study of Governder and Mars. Perhaps a study that addresses a wider age range would provide more accurate information.

The maximum ambient noise monitored in this study was 66.2 ± 7.2 dB, which did not exceed the MPANLs specified in the instructions, for transient occurrences of high ambient noise, where KUDUwave pauses the test, allowing good correlation to be obtained between the results of the automated test performed in a general clinic room and the manual test in an acoustically insulated room. It has been shown that insert earphones, when used in combination with earmuffs, optimize ambient noise attenuation to a level where the total noise attenuation can exceed that of a single-walled sound-insulated room (Seluakumaran and Shaharudin, 2021). Because ambient noise can affect test results not only through air conduction, higher ambient noise can also affect results through bone conduction. Therefore, it is recommended that testing in a non-soundproofed environment be performed in a quiet room.

In our previous study, we performed a comparison of manual and automated tests under sound-isolation conditions and found that the reliability at 250 Hz and 8,000 Hz was worse than at other frequencies (Liu et al., 2022), however, this phenomenon did not occur in the present group of subjects, which may be related to the use of different headphones. In the previous study, insert headphones were used for automated audiometry and circumaural earphones were used for manual audiometry; in the present study, KUDUwave audiometer were used for both manual-and automated pure-tone audiometry, which eliminates the calibration differences that were introduced by two different devices, and could easily interpret some changes in hearing thresholds.

Study limitations.

One of the limitations of this study is that bone conduction threshold tests were not conducted on the reliability of automated pure tone audiometry. The relationship between bone and air conduction is an important basis for determining the presence or absence of conductive hearing loss, and a subsequent study will be conducted to investigate the clinical application of bone conduction for automated audiometry.

The ambient noise levels were recorded manually by a sound level meter, and were also continuously monitored by KUDUwave, however, the data from KUDUwave was not available. It would be useful to compare whether both the noise monitoring methods provided similar levels.

Although the sample size estimate indicated that the 83 initial subjects for this study met the requirements. However, the inclusion of a larger sample of subjects would have improved the credibility of this study. The insufficiently large sample size is a limitation of this study.

5 Conclusion

In this study, subjects were grouped according to age and hearing level, respectively. Automated pure-tone audiometry was performed in the general consultation room, and manual pure-tone audiometry was performed in the acoustic isolation room using KUDUwave audiometer. There was a good correlation between the automated and manual audiometric thresholds. Subjects' reaction times increased with age, and reaction time measurements provided a basis for a more accurate parameter setting of the automated tests. In the case of individual subjects with high variability of results, quality control of the automated test needs to be increased so that such subjects can be screened out and transferred to manual audiometry. In conclusion, automated air conduction pure-tone audiometry has great potential to play a greater role, especially in economically underdeveloped areas, or in mass hearing screening scenarios.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by The Medical Ethics Committee for Clinical Research of Beijing Tongren Hospital. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin.

Author contributions

HL: Writing – original draft, Writing – review & editing. XF: Writing – original draft, Writing – review & editing. ML: Writing – original draft. SW: Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was funded in part by grants from the National Key Research and Development Program of China (2023YFC3604200-project 2023YFC3604203), the National Natural Science Foundation of China (81870715), High-Level Public Health Technical Talent Training Plan (Discipline Backbone-02-42).

Acknowledgments

The authors thank all participants of this study for their valuable contributions. The authors also especially thank Robert H. Eikelboom from Ear Science Institute Australia, who supplied technical guidance in study design and statistical analysis.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2023.1292395/full#supplementary-material

References

Bornman, M., Swanepoel, D. W., Biagio De Jager, L., and Eikelboom, R. H. (2018). Extended high-frequency smartphone audiometry: validity and reliability. *J. Am. Acad. Audiol.* 30, 217–226. doi: 10.3766/jaaa.17111

Brennan-Jones, C. G., Eikelboom, R. H., Swanepoel De, W., Friedland, P. L., and Atlas, M. D. (2016). Clinical validation of automated audiometry with continuous noise-monitoring in a clinically heterogeneous population outside a sound-treated environment. *Int. J. Audiol.* 55, 507–513. doi: 10.1080/14992027.2016.1178858

Corry, M., Sanders, M., and Searchfield, G. D. (2017). The accuracy and reliability of an app-based audiometer using consumer headphones: pure tone audiometry in a normal hearing group. *Int. J. Audiol.* 56, 706–710. doi: 10.1080/14992027.2017.1321791

Eksteen, S., Launer, S., Kuper, H., Eikelboom, R. H., Bastawrous, A., and Swanepoel, W. (2019). Hearing and vision screening for preschool children using mobile technology, South Africa. *Bull. World Health Organ.* 97, 672–680. doi: 10.2471/BLT.18.227876

Govender, S. M., and Mars, M. (2018a). Assessing the efficacy of asynchronous telehealth-based hearing screening and diagnostic services using automated audiometry in a rural south African school. *S. Afr. J. Commun. Disord.* 65, e1–e9. doi: 10.4102/sajcd.v65i1.582

Govender, S. M., and Mars, M. (2018b). Validity of automated threshold audiometry in school aged children. *Int. J. Pediatr. Otorhinolaryngol.* 105, 97–102. doi: 10.1016/j.ijporl.2017.12.008

Kramer, S. E., Kapteyn, T. S., and Houtgast, T. (2006). Occupational performance: comparing normally-hearing and hearing-impaired employees using the Amsterdam checklist for hearing and work. *Int. J. Audiol.* 45, 503–512. doi: 10.1080/14992020600754583

Liu, H., Du, B., Liu, B., Fu, X., and Wang, Y. (2022). Clinical comparison of two automated audiometry procedures. *Front. Neurosci.* 16:1011016. doi: 10.3389/fnins.2022.1011016

Liu, H., Fu, X., Du, B., Wang, Y., and Zhong, Y. (2021). Accuracy and reliability of automatic pure tone audiometry. *Chin Arch Otolaryngol Head Neck Surg* 28, 348–351. doi: 10.16066/j.1672-7002.2021.06

Livingston, G., Sommerlad, A., Orgeta, V., Costafreda, S. G., Huntley, J., Ames, D., et al. (2017). Dementia prevention, intervention, and care. *Lancet* 390, 2673–2734. doi: 10.1016/S0140-6736(17)31363-6

Maclennan-Smith, F., Swanepoel De, W., and Hall, J. (2013). Validity of diagnostic pure-tone audiometry without a sound-treated environment in older adults. *Int. J. Audiol.* 52, 66–73. doi: 10.3109/14992027.2012.736692

Mahomed, F., Swanepoel De, W., Eikelboom, R. H., and Soer, M. (2013). Validity of automated threshold audiometry: a systematic review and meta-analysis. *Ear Hear.* 34, 745–752. doi: 10.1097/01.aud.0000436255.53747.a4

Margolis, R. H., Killion, M. C., Bratt, G. W., and Saly, G. L. (2016). Validation of the home hearing test. *J. Am. Acad. Audiol.* 27, 416–420. doi: 10.3766/jaaa.15102

Margolis, R. H., Saly, G. L., Le, C., and Laurence, J. (2007). Qualind: a method for assessing the accuracy of automated tests. *J. Am. Acad. Audiol.* 18, 78–89. doi: 10.3766/jaaa.18.1.7

Mosley, C. L., Langley, L. M., Davis, A., Mcmahon, C. M., and Tremblay, K. L. (2019). Reliability of the home hearing test: implications for public health. *J. Am. Acad. Audiol.* 30, 208–216. doi: 10.3766/jaaa.17092

Olusanya, B. O., Davis, A. C., and Hoffman, H. J. (2019). Hearing loss grades and the international classification of functioning, disability and health. *Bull. World Health Organ.* 97, 725–728. doi: 10.2471/BLT.19.230367

Rutjes, A. W., Reitsma, J. B., Di Nisio, M., Smidt, N., Van Rijn, J. C., and Bossuyt, P. M. (2006). Evidence of bias and variation in diagnostic accuracy studies. *CMAJ* 174, 469–476. doi: 10.1503/cmaj.050090

Sandström, J., Swanepoel, D., Laurent, C., Umefjord, G., and Lundberg, T. (2020). Accuracy and reliability of smartphone self-test audiometry in community clinics in low income settings: a comparative study. *Ann. Otol. Rhinol. Laryngol.* 129, 578–584. doi: 10.1177/0003489420902162

Seluakumaran, K., and Shaharudin, M. N. (2021). Calibration and initial validation of a low-cost computer-based screening audiometer coupled to consumer insert phone-earmuff combination for boothless audiometry. *Int. J. Audiol.* 61, 850–858. doi: 10.1080/14992027.2021.1969455

Shojaeemend, H., and Ayatollahi, H. (2018). Automated audiometry: a review of the implementation and evaluation methods. *Healthcare Informatics Res* 24, 263–275. doi: 10.4258/hir.2018.24.4.263

Storey, K. K., Munoz, K., Nelson, L., Larsen, J., and White, K. (2014). Ambient noise impact on accuracy of automated hearing assessment. *Int. J. Audiol.* 53, 730–736. doi: 10.3109/14992027.2014.920110

Sun, X., Yu, L., Qu, C., Liang, W., Wang, Q., and Wei, Z. (2008). An epidemiological study on the hearing-impaired population identified in China and proposed intervention strategies. *Chinese Scientific J Hearing and Speech Rehab* 2, 21–24.

Swanepoel De, W., Koekemoer, D., and Clark, J. (2010a). Intercontinental hearing assessment - a study in tele-audiology. *J. Telemed. Telecare* 16, 248–252. doi: 10.1258/jtt.2010.090906

Swanepoel De, W., Maclennan-Smith, F., and Hall, J. W. (2013). Diagnostic pure-tone audiometry in schools: mobile testing without a sound-treated environment. *J. Am. Acad. Audiol.* 24, 992–1000. doi: 10.3766/jaaa.24.10.10

Swanepoel De, W., Matthysen, C., Eikelboom, R. H., Clark, J. L., and Hall, J. W. (2015). Pure-tone audiometry outside a sound booth using earphone attentuation, integrated noise monitoring, and automation. *Int. J. Audiol.* 54, 777–785. doi: 10.3109/14992027.2015.1072647

Swanepoel De, W., Mngemane, S., Molemong, S., Mkwanazi, H., and Tutshini, S. (2010b). Hearing assessment-reliability, accuracy, and efficiency of automated audiometry. *Telemed. J. E Health* 16, 557–563. doi: 10.1089/tmj.2009.0143

Van Tonder, J., Swanepoel, D. W., Mahomed-Asmail, F., Myburgh, H., and Eikelboom, R. H. (2017). Automated smartphone threshold audiometry: validity and time efficiency. *J. Am. Acad. Audiol.* 28, 200–208. doi: 10.3766/jaaa.16002

Visagie, A., Swanepoel De, W., and Eikelboom, R. H. (2015). Accuracy of remote hearing assessment in a rural community. *Telemed. J. E Health* 21, 930–937. doi: 10.1089/tmj.2014.0243

Wasmann, J. W., Pragt, L., Eikelboom, R., and Swanepoel, W. (2022). Digital approaches to automated and machine learning assessments of hearing: scoping review. *J. Med. Internet Res.* 24:32581. doi: 10.2196/32581

Whitton, J. P., Hancock, K. E., Shannon, J. M., and Polley, D. B. (2016). Validation of a self-administered audiometry application: an equivalence study. *Laryngoscope* 126, 2382–2388. doi: 10.1002/lary.25988

World Health Organization (2021). *World report on hearing*. (Vol. Licence: CC BY-NC-SA 3.0 IGO.). Available at: https://www.who.int/publications/i/item/world-report-on-hearing.

frontiers | Frontiers in Audiology and Otology

TYPE Original Research
PUBLISHED 26 January 2024
DOI 10.3389/fauot.2024.1292949

Check for updates

OPEN ACCESS

EDITED BY
Faheema Mahomed-Asmail,
University of Pretoria, South Africa

REVIEWED BY
Inga Holube,
Jade University of Applied Sciences, Germany
Razan Alfakir,
Auburn University, United States

*CORRESPONDENCE
Arnaud Génin
✉ arnaud.g@sonup.fr

RECEIVED 12 September 2023
ACCEPTED 03 January 2024
PUBLISHED 26 January 2024

CITATION
Génin A, Courtial J, Balcon M, Puel J-L, Venail F and Ceccato J-C (2024) Development and validation of a French speech-in-noise self-test using synthetic voice in an adult population. *Front. Audiol. Otol.* 2:1292949. doi: 10.3389/fauot.2024.1292949

COPYRIGHT
© 2024 Génin, Courtial, Balcon, Puel, Venail and Ceccato. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Development and validation of a French speech-in-noise self-test using synthetic voice in an adult population

Arnaud Génin[1,2,3]*, Jérôme Courtial[1,2,4], Maxime Balcon[3], Jean-Luc Puel[2,4], Frédéric Venail[1,2,4] and Jean-Charles Ceccato[2,4]

[1]Département ORL&CMF- CHU, Université de Montpellier, Montpellier, France, [2]Institut des Neurosciences de Montpellier, INSERM U1298, Montpellier, France, [3]SONUP, Montpellier, France, [4]Audiocampus, Université de Montpellier, Montpellier, France

Background: Speech-in-noise (SIN) audiometry is a valuable part of audiological diagnostics and clinical measurements, providing information on an individual's ability to understand in noise. To date, such tests have been developed with natural voice presented monaurally or binaurally (via diotic and dichotic presentation). The time required to develop such tests is often long. The use of synthetic voice would simplify the test creation process and self-testing would improve accessibility.

Design: Measurements were performed using an Android tablet (Samsung Galaxy Tab A7) and calibrated Bluetooth headphones (Tilde Pro C, Orosound). Normative values were first defined using natural or synthetic voice on 69 normal-hearing participants. A total of 463 participants then undertook the SIN test comprising synthetic voice and dichotic antiphasic presentation. Of these, 399 also performed the SIN test with diotic presentation.

Results: No significant differences in the speech reception threshold (SRT) were found between natural and synthetic voices for diotic presentations ($p = 0.824$, paired Student t-test) with normative values of −10.7 dB SNR (SD = 1.5 dB) and −10.4 dB SNR (SD = 1.4 dB), respectively. For the SoNoise test with synthetic voice and dichotic antiphasic presentation, the results showed a normative value of −17.5 dB SNR (SD = 1.5 dB), and a strong correlation ($r = 0.797$, $p < 0.001$) with the four-frequency pure-tone average (4f-PTA). Receiver operating curves (ROC) were then calculated: for a 4f-PTA of 20 dB hearing level (HL), the SRT was −14.5 dB SNR with a sensitivity of 84% and specificity of 89%. For a 4f-PTA of 30 dB HL, the SRT was −13.7 dB SNR with a sensitivity of 89% and specificity of 91%. For a 4f-PTA of 35 dB HL, the SRT was −13.0 dB SNR with a sensitivity of 88% and specificity of 93%. The normative binaural intelligibility level difference (BILD) value was 8.6 dB (SD = 2.0 dB) with normal-hearing participants. The learning effect due to the task and interface was 1.7 dB (1st to 7th test) and test duration was 3 min.

Conclusion: The SoNoise test in its synthetic dichotic antiphasic presentation is a fast and reliable tool to diagnose hearing impairment at 20, 30, and 35 dB HL cut-offs.

KEYWORDS

hearing diagnosis, speech-in-noise test, natural and synthetic voices, dichotic antiphasic, tablet based

1 Introduction

While pure-tone audiometry is currently recognized as the gold-standard measurement for hearing loss assessment, compelling evidence suggests that speech-in-noise (SIN) audiometry should be systematically added to evaluate the functional impairment related to hearing loss (Plomp, 1986; Killion and Niquette, 2000; Smits et al., 2004; Smits and Houtgast, 2005; Jansen et al., 2012). Most SIN tests assess the signal-to-noise ratio (SNR) at which a participant can recognize 50% percent of words. This particular SNR is called the speech reception threshold (SRT). A high correlation (between 0.77 and 0.86) is generally observed between pure-tone average (PTA) and SRT in noise (Smits et al., 2004; Jansen et al., 2010, 2013; Koole et al., 2016; Potgieter et al., 2018a,b). However, discordance between PTA and SIN may be found in certain medical conditions, such as auditory neuropathy (Rance et al., 2012; Narne, 2013; Apeksha and Kumar, 2017; White-Schwoch et al., 2020, 2022) or central auditory processing disorders (Houtgast and Festen, 2008; Anderson et al., 2011; Bellis and Bellis, 2015; Füllgrabe et al., 2015). Use of SIN tests offers three main advantages: (1) they are more representative of the everyday discomfort and of the handicap experienced by hearing-impaired people (Carhart and Tillman, 1970; Kramer et al., 1998; Grant and Walden, 2013); (2) they are more sensitive to early events of age-related hearing impairment, detecting the loss of auditory neurons that cannot be detected by PTA or even by speech audiometry in quiet (Wu et al., 2020); and (3) the supra-threshold measurement of the SRT is less sensitive than pure-tone threshold measurements to minor calibration inaccuracies.

Several SIN tests have been developed in French over the past 20 years, showing a rising interest for this type of auditory evaluation. The most frequently used are currently the French version of the hearing in noise test (HINT; Vaillancourt et al., 2005), the French intelligibility sentence test (FIST; Luts et al., 2008), the French digit triplet test (FrDigit3; Jansen et al., 2010), the speech understanding in noise (SUN; Paglialonga et al., 2011), the FraMatrix (Jansen et al., 2012), the *vocale rapide dans le bruit* (VRB, fast speech in noise; Leclercq et al., 2018), the FrBio (Bergeron et al., 2019), and the antiphasic digit triplet test (Höra; Ceccato et al., 2021). The SRT can be determined directly with tests using adaptive methods, such as the digit triplet test, Matrix tests, FrDigit3, HINT, and Höra, in which the SNR evolves automatically according to the participants' answer at each presented item (Nilsson et al., 1994; Smits et al., 2004; Jansen et al., 2010; Kollmeier et al., 2015; Ceccato et al., 2021). The pace of SNR adaptation may vary throughout the test, according to the number of correct answers, and depends on the recognition score target. Other SIN tests, such as QuickSIN (Killion et al., 2004), SUN (Paglialonga et al., 2011), and VRB (Leclercq et al., 2018), use lists of items presented at fixed SNR, for which the SRT can be measured by fitting the obtained recognition score at each presented SNR with a psychometric function. The FrBio (Bergeron et al., 2019) aimed to provide a more ecological approach for the SIN paradigm by measuring the recognition score at fixed SNR in real-life sound situations. Among these tests, only the digit triplet test, the FrDigit3 and Höra are currently performed in self-test mode, and their use is limited to screening. On the other hand, none of the French SIN tests used in clinical assessment, such as the HINT (Vaillancourt et al., 2005), the FIST (Luts et al., 2008), the FrDigit3 test (Jansen et al., 2010), the FrBio (Bergeron et al., 2019), the SUN (Paglialonga et al., 2011), the FraMatrix (Jansen et al., 2012), and the VRB (Leclercq et al., 2018), are currently used in self-testing mode but mostly with the investigator recording the subject's answers.

SIN tests developed for clinical evaluations are mainly performed in free-field settings, which may entail practical difficulties linked to the need for space and maintenance of a reliable calibration (VRB, FIST, HINT, FraMatrix, and FrBio). For practical reasons, screening tests using SIN have been presented with headphones, either monaurally or binaurally (Smits et al., 2004; Jansen et al., 2012; Van den Borre et al., 2021). Some recent SIN tests performed binaurally, have used dichotic antiphasic presentation (De Sousa et al., 2020; Ceccato et al., 2021) of a speech signal, with a diotic presentation of the noise (i.e., SπN0). The phase shift allows the use of binaural mechanisms involving the comparison of the temporal clues between the two ears. It results in a binaural masking release that improves the perception of the target signal (Culling and Lavandier, 2021). Binaural masking level difference (BMLD) has been extensively explored for tonal stimulus (Hirsh, 1948; Webster, 1951; Wilson et al., 2003). In such studies, the BMLD of normal-hearing individuals mainly varies according to the frequency of the stimulus to be detected: between 10 and 15 dB of enhancement at 500 Hz, and 1–3 dB at 4,000 Hz. For the SIN test, use of the presentation mode BILD, corresponding to the difference between binaural diotic (S0N0) and antiphasic (SπN0) presentation, improves the sensitivity and specificity of the test for detecting asymmetric, unilateral, and conductive hearing loss when used as a screening tool (De Sousa et al., 2020, 2022; Ceccato et al., 2021). Concerning correlation between SRT and PTA, the SRT of antiphasic and binaural diotic tests, respectively correlate better to the PTA of the worst and better ear (De Sousa et al., 2020; Ceccato et al., 2021). The BILD could be of interest in a clinical assessment as it tests binaural auditory functions that cannot be observed with headphones in either monaural or binaural diotic presentation mode nor in most free-field configurations.

Speech material used for SIN tests is usually based on the studio recordings of a speaker's voice, which entails certain disadvantages such as the cost and the duration of test development (Dickerson et al., 2006). Considering the progress made in voice synthesis (Gong and Lai, 2003; King, 2014) and its current use in everyday life (telecommunications, information services, numeric applications), we wondered about the relevance of its application for speech audiometry. Some studies have assessed the comparability of natural and synthetic voice in speech audiometry (Koul, 2003; Cooke et al., 2013; Simantiraki et al., 2018; Schwarz et al., 2022). Most of the clinical SIN tests still use natural voices, but some use synthetic ones (Nuesse et al., 2019; Ibelings et al., 2022).

The objective of this study was to develop and normalize a SIN test that could be used both as a screening tool and for clinical evaluation of SIN. For this purpose, we recruited a normative population to: (1) evaluate if the use of a synthetic voice may induce a difference in SRT measurement relative to the use of a more classical natural voice recording; (2) determine normative values for diotic and antiphasic presentation of the test; and (3) determine normative values of the BILD. We then assessed the validity

of the test on a study population composed of normal-hearing and hearing-impaired participants presenting various audiometric profiles. We also evaluated the normative values for this test in screening and clinical assessment.

2 Materials and methods

2.1 Participants

The participants were recruited and tested in the ENT department of the university hospital in Montpellier (France). They were outpatients, accompanying persons, caregivers, students or hospital workers. Exclusion criteria were visual or motor impairments that prevented use of a tablet, self-reported cognitive functions disallowing understanding the principle of the tests, earwax, ear discharge, or ear malformation preventing the use of headphones. No exclusion criterion was based on PTA.

Normative values, duration and learning effect of each SoNoise test were calculated based on a first population of 69 normal-hearing French native speakers (4f-PTA 0.5/1/2/4 kHz ≤10 dB HL), 39 women and 30 men aged between 18 and 25 years, tested for the first time with SoNoise tests. Participants were chosen to be in agreement in age and hearing loss with the standard (ISO 8253-3). The SRT with natural or synthetic voice was compared in one group ($n = 43$) comprising 28 women and 15 men, with an average 4f-PTA of 5 dB HL (SD = 4.4 dB HL, median 5 dB HL, IC95 [3.7–6.3]) and mean age of 21.2 years (SD = 2.2 yrs, median 21 yrs, IC95 [20.2–22.2]). The SRT with diotic or antiphasic synthetic SIN was compared in the remaining group of normal-hearing participants ($n = 26$), comprising 12 women and 14 men of mean age 22.9 years (SD = 2.8 yrs, median 22.5 yrs, IC95 [21.8–24]), with an average 4f-PTA of 4.3 dB HL (SD = 2.7 dB HL, median 3.9 dB HL, IC95 [3.4–5.2]).

A second test population of 463 French native speakers, 230 women and 233 men all over 18 years of age (mean age 40 yrs, SD = 23 yrs, median 29 yrs, IC95 [38–42]), was used to assess the diagnostic performance of the SoNoise_SπN0_Syn test. Of these, for their best ear, 337 (72.8 %) were classified as having normal hearing, 51 (11.0 %) mild hearing loss, 55 (11.9%) moderate hearing loss, 19 (4.1%) moderately severe hearing loss, and 1 (0.2%) severe-profound hearing loss. For their worst ear, 328 (70.8%) were classified as having normal-hearing, 35 (7.7%) mild hearing loss, 57 (12.3%) moderate hearing loss, 27 (5.8%) moderately severe hearing loss, 15 (3.2%) severe-profound hearing loss, and 1 (0.2%) profound hearing loss. Degrees of hearing were based and categorized according to the World Health Organization grades of hearing impairment (World Health Organization, 2021) as follows: normal-hearing (PTA ≤20 dB HL), mild (PTA >20 ≤35 dB HL), moderate (PTA >35 ≤50 dB HL), moderately severe (PTA >50 ≤65 dB HL), severe-profound (PTA >65 ≤80 dB HL), profound (PTA >80 ≤95 dB HL), or complete hearing loss (PTA >95 dB HL).

Among the study population of 463 participants, a subset of 399 (188 women and 211 men) with a mean age of 36 years (SD = 21 yrs, median 27 yrs, IC95 [33–39]), received both SoNoise_S0N0_Syn and SoNoise_SπN0_Syn tests. Of these, 331 (83%) were classified as having normal hearing, 31 (7.7%) mild hearing loss, 31 (7.7%) moderate hearing loss, 5 (1.4%) moderately-severe hearing loss, and 1 (0.2%) severe-profound hearing loss. The audiometric profiles of both normative and test validation populations are displayed in Table 1.

2.2 Speech material

Praat software (Boersma and Weenink, 2013) was used to determine the main speech characteristics (duration, fundamental frequency, speech rate) of both natural and synthetic voice. Natural voice recordings were performed by a 38 years old French native female speaker, who is not a professional speaker. Synthetic words were generated with "neural voices" (powered by Acapela Group, version 2017.1).

Table 2 shows the main characteristics: fundamental frequency (Hz), word duration (ms) and speech rate (syllables/s). The fundamental frequency of the natural speech triplet was on average higher (210 Hz) than that of the synthetic speech (178 Hz). The average word duration of the natural voice was 661 ms, whereas that of the synthetic voice was 451 ms, leading to a speech rate for natural and synthetic voice of 1.6 and 2.3 syllables/s, respectively.

2.3 SoNoise tests

SoNoise tests are adaptive SIN self-tests aimed at automatically determining the SRT (dB SNR) in noise. They consisted here in the presentation of different triplets of words (digit -common noun—color) at different speech-to-noise ratios, as described elsewhere (Prang et al., 2021). Each word was randomly selected among 9, leading to $9^3 = 729$ triplet combinations, with equal odds of presentation. The participant had to select the word heard by pressing the corresponding icon representing that word on the screen of the tablet. On the first response screen, the participant had to pick the correct word among the 9 icons representing digits), then a second response screen was displayed with the next 9 icons (representing common nouns), and finally a third screen was displayed with the last 9 icons (representing colors). The participant was instructed to choose an icon, even if the word was not heard, making this a forced-choice test with a closed-set list of words (Smits et al., 2006; De Sousa et al., 2018).

SoNoise tests are adaptive SIN tests designed to be performed as self-tests. The SoNoise_S0N0 had a binaural diotic presentation, i.e., both earphones delivered the same sound stimuli (words + noise) to each ear. Words of the SoNoise_S0N0 were generated either with a natural (Na) or a synthetic (Syn) voice. The SoNoise_SπN0_Syn offered a dichotic antiphasic presentation of words while presenting the same level of noise to each ear. A phase shift of "π" was introduced with word presentation. For this test, words were generated with a synthetic voice (Syn).

The masking noise was a white noise with envelope shaped on the long-term spectrum of the test words, as described elsewhere (Plomp and Mimpen, 1979; Nilsson et al., 1994; Brand and Kollmeier, 2002; Smits et al., 2006; Soli and Wong, 2008; Jansen

TABLE 1 Demographic characteristics of both normative and test validation study populations.

Demographic characteristics		SoNoise_S0N0_Na	SoNoise_S0N0_Syn	SoNoise_Sπ N0_Syn	SoNoise_Sπ N0_Syn	SoNoise_S0N0_Syn
		Normative value	Normative value	Normative value	Study population	Study population
Number of participants (*n*)		43	69*	26	463	399**
Mean age (yrs)		21.7 (SD 2.5)	23.1 (SD 3.4)	24.8 (SD 3.7)	40 (SD 23)	36 (SD 21)
Female (*n*)		28 (65%)	39 (56%)	12 (46%)	230 (50%)	188 (47%)
Male (*n*)		15 (35%)	30 (44%)	14 (54%)	233 (50%)	211 (53%)
4f-PTA (dB HL)***						
	Normal (≤20)	43 (100%)	69 (100%)	26 (100%)	328 (70.8%)	331 (83%)
	Mild (>20, ≤35)				35 (7.7%)	31 (7.7%)
	Moderate (>35, ≤50)				57 (12.3%)	31 (7.7%)
	Moderately severe (>50, ≤65)				27 (5.8%)	5 (1.4%)
	Severe-profound (>65, ≤80)				15 (3.2%)	1 (0.2%)
	Profound (>80, ≤95)				1 (0.2%)	
	Complete (>95)					

*Sixty-nine is the sum of 26 and 43.
**Three hundred and ninety-nine is a sub-group of the 463 participants tested.
***We used 4f-PTA of the worst ear to define the population for the SoNoise_Sπ N0_Syn test, while we used 4f-PTA of the best ear for the SoNoise_S0N0_Syn test.

TABLE 2 Characteristics of the speech material for both natural and synthetic voices.

	Digit		Common noun		Color		Triplet	
	Na	Syn	Na	Syn	Na	Syn	Na	Syn
Fundamental frequency (Hz)	216 (SD 11)	178 (SD 14)	218 (SD 8)	189 (SD 8)	197 (SD 21)	169 (SD 5)	210 (SD 17)	178 (SD 13)
Word duration (ms)	582 (SD 100)	421 (SD 126)	769 (SD 105)	490 (SD 96)	633 (SD 48)	441 (SD 80)	661 (SD 117)	451 (SD 102)
Speech rate (syllables/s)	1.8 (SD 0.4)	2.6 (SD 0.9)	1.3 (SD 0.2)	2.1 (SD 0.4)	1.6 (SD 0.1)	2.3 (SD 0.4)	1.6 (SD 0.3)	2.3 (SD 0.6)

For digits, common nouns and colors, the mean values of the fundamental frequency, word duration and speech rates were calculated. Triplet values were obtained by averaging those of digit, common noun and color.

et al., 2012; Dillon et al., 2016; Potgieter et al., 2016; De Sousa et al., 2018). Shaping was made by: (1) obtaining the long-term spectrum of all the concatenated words; (2) using this spectrum to design a frequency sampling-based finite-impulse-response filter with the desired frequency shape and applying it on a white noise; (3) adjusting noise level with the concatenated words. Speech and noise power spectrum for both natural and synthetic tests are displayed in Figure 1.

The masking noise used depended on the SoNoise test performed, i.e., it was shaped on the long-term spectrum of natural voice words for the Na test, or synthetic words for the Syn tests. The noise started 500 ms before the first word and ended 500 ms after the third, as described elsewhere (Jansen et al., 2010; Smits et al., 2013; Kaandorp et al., 2015; Potgieter et al., 2016; Ceccato et al., 2021). A silence gap of 100 ms was inserted between words, to which was added a jitter (random extra delay) of 0–200 ms (Potgieter et al., 2016). The speech was presented at 75 dB SPL (sound pressure level, Leq measurements) and a SNR of 20 dB at the beginning of the test. The SNR level varied adaptively according to the number of words recognized correctly (0, 1, 2, or 3) as follows, respectively: +10, +5, −5, −10 until the first reversal, then +5, +2, −2, −5 between the first and the second reversal, and +3, +1, −1, −3 after that point.

When the SNR was positive, the noise level was modified. When the SNR level was negative, the speech level was modified. Twelve reversals were performed during the test, and the SRT was calculated by averaging the SNR results of the last eight reversals.

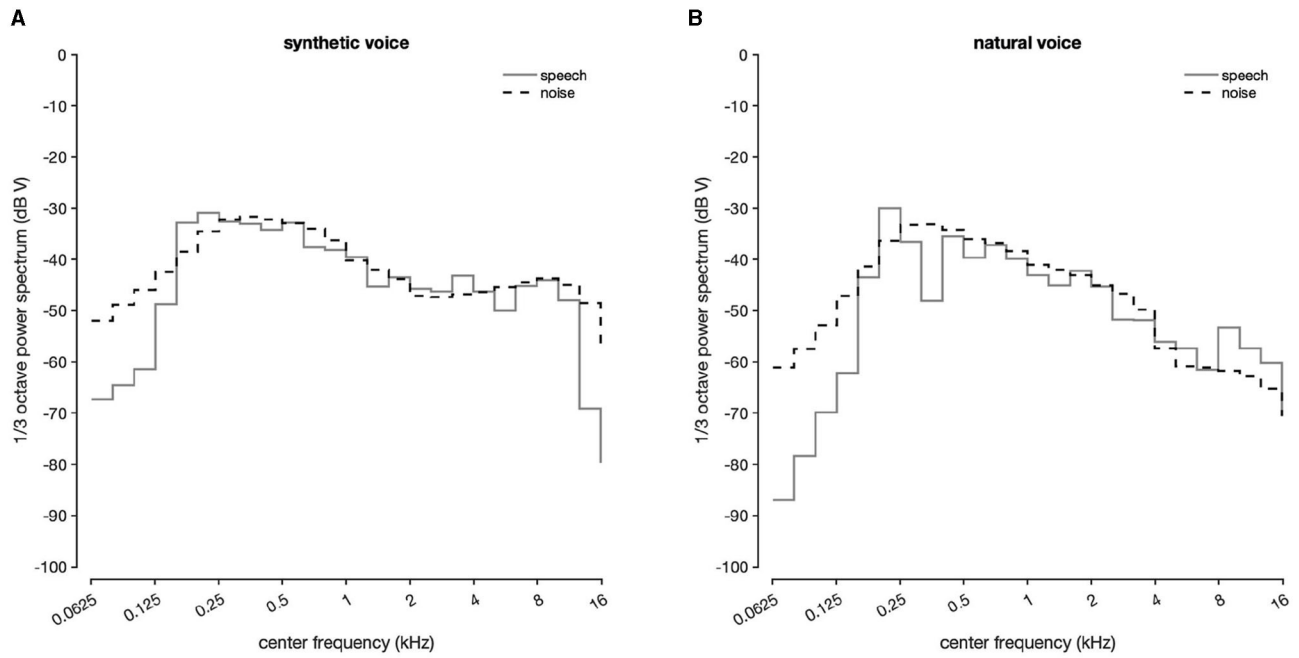


FIGURE 1
(A) Long-term average spectrum for synthetic speech and noise. **(B)** Long-term average spectrum for natural speech and noise.

2.4 Words recording and difficulty equalization

Speech material was composed of monosyllabic words for digits and colors, and disyllabic words for common nouns (except for one trisyllabic word: "sanglier"). All words used were common language and easy to represent as an image.

For both types of voice, the words were generated individually and then combined for testing, ensuring that the word triplet included no prosody or natural coarticulation. Prior to the study, equalization of the words was performed for all the SoNoise tests developed. It consisted in adjusting the level of presentation of each word to ensure that each one had a 50% chance of being recognized at the same SNR. A specific level adjustment has been performed for each one of the three tests developed (SoNoise_S0N0_Na, SoNoise_S0N0_Syn, SoNoise_SπN0_Syn). To do this, 76 normal-hearing participants [PTA 0.5/1/2/4 kHz $<$ 20 dB hearing level (HL)] aged from 18 to 40 years (mean age of 21.8, SD = 4.5 yrs, median 20 yrs, IC95 [20.8–22.8]) were tested 1–5 times each. In this population, 21 normal-hearing subjects didn't match the ISO 8253-3 recommendation in terms of age and/or PTA thresholds, but their SRT at the test didn't statistically differ from the other participants. This allowed a reasonable inclusion of their results as item difficulty equalization was made following an adaptation of both the method proposed Brand and Kollmeier in 2002 for Matrix tests (Brand and Kollmeier, 2002; Jansen et al., 2012) and recently by Masalski et al. (2021) for digits in noise tests. The principle was to use the final test procedure to perform an evaluation of each item difficulty that considered both their inner difficulty and their position in the triplet. A routine Matlab script was used to identify the words pronounced and answered at each SNR. The psychometric curves of recognition for each word were fitted with a logistic function to determine the SRT. The equalization values obtained for the SoNoise_SπN0_Syn test are displayed in Figure 2A. Each word's recording level was then adjusted using the difference between the SRT of each word and the average SRT of all words (−17.4 dB SNR) while the average SRT of the 76 subjects was −17.5 dB SNR (SD = 2.5). The ≪ before/after ≫ results of the equalization are presented Figure 2 and shows that for digits, common nouns, and colors, the average SRT were respectively −18 dB SNR (SD = 2.3), −18.3 dB SNR (SD = 1.4), −15.9 dB SNR (SD = 2.9) before equilibration, and respectively −18.5 dB SNR (SD = 1), −18.9 dB SNR (SD = 1), −18 dB SNR (SD = 1) after equilibration. Globally, the item standard deviation of SRT dropped from 2.5 to 1 dB.

2.5 Equipment and procedure

Conventional audiometry was carried out in a soundproof booth, with an AC33 audiometer (Interacoustics) calibrated with TDH-39 headphones.

For SIN testing, the SoNoise (SONUP, Montpellier, France) hearing application was used on an Android OS tablet (Samsung Galaxy Tab A7) connected via Bluetooth to circumaural headphones (Orosound Tilde Pro C). The SoNoise tests were performed in a quiet office. The KEMAR (Knowles Electronics Manikin for Acoustic Research, SET electronic GmbH, MK2-B, CE labeled), and its built-in prepolarized pressure microphones (GRAS 40AO ½), was chosen for the calibration of the tablet-headphones pair over an artificial ear for its superior acoustic coupling with the chosen headphones, that more closely resembles that of the adult participants tested in the clinic (Xie et al., 2009; Guo et al., 2021).

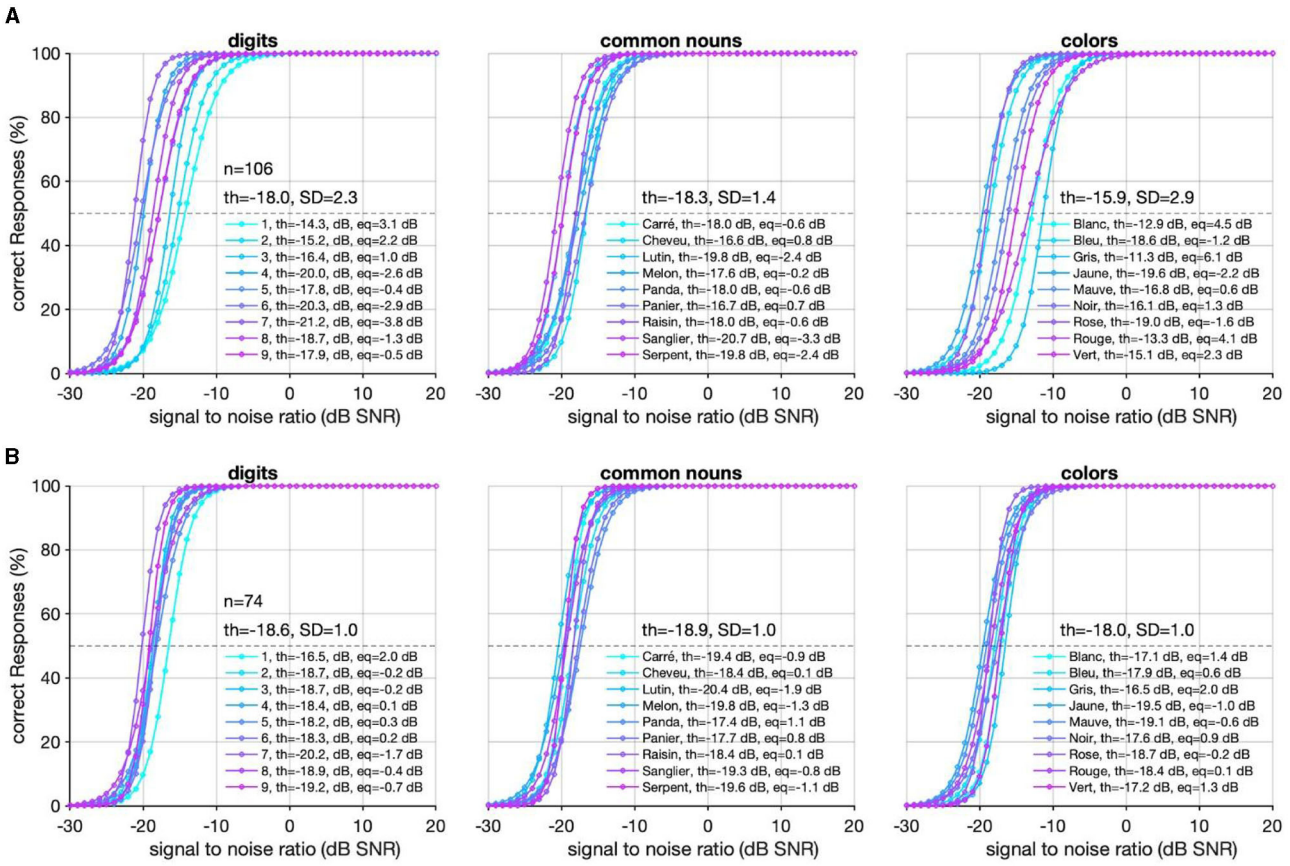


FIGURE 2
(A) Psychometric curves of the 27 synthetic words of the test before the first equalization. The mean SRT of each type of item is presented with SD. For each item, the SRT and the difference (eq) with the average SRT of all words is also presented. **(B)** Psychometric curves of these same 27 words after the first equalization. "*n*" corresponds to the number of SoNoise_SπN0_Syn tests performed.

The maximum output level was measured using the masking noise of the test, and used as a reference for the calibration level. Its value in dB SPL (sound pressure level, Leq measurements) was input into the application, which used it as a reference to send the desired level of signal.

This prospective study was conducted in the ENT department of the university hospital of Montpellier (France), and aimed at the development and the validation of the SoNoise SIN tests (SONUP). The study was approved by the local ethics committee (IRB-MTP_2021_09_202100889). All participants signed a consent form to participate in the study.

Audiometric thresholds (air and bone conduction) were determined for each subject (normative and study population) at 0.5, 1, 2, 4, and 8 kHz, using the modified Hughson-Westlake method (Carhart and Jerger, 1959) after bilateral otoscopy. The 4f-PTA was calculated by averaging the audiometric thresholds (0.5, 1, 2, and 4 kHz) measured during air conduction pure-tone audiometry.

To define normative values for both SoNoise_S0N0_Na and SoNoise_S0N0_Syn, participants (n = 43) performed a total of seven tests: a training test with the SoNoise_S0N0_Na, then alternating SoNoise_S0N0_Syn or SoNoise_S0N0_Na three times, followed by another sequence comprising the other test not presented in the first sequence (SoNoise_S0N0_Na or SoNoise_S0N0_Syn) again repeated three times, ensuring that both tests were passed three times each in an alternative manner.

To define normative values for both SoNoise_S0N0_Syn and SoNoise_SπN0_Syn, participants (n = 26) performed a total of seven tests as described earlier except that the training test was the SoNoise_SπN0_Syn and the six tests that followed alternated between SoNoise_S0N0_Syn and SoNoise_SπN0_Syn. All tests were achieved within the same session. The normative value of the BILD for the test was calculated by subtracting SπN0 SRT from S0N0 SRT. The tablet application measured the duration of each test from the start to the finish.

Of the 463 participants (normal or hearing-impaired) who underwent the SoNoise_SπN0_Syn test, 399 also underwent the SoNoise_S0N0_Syn test in a counterbalanced order.

2.6 Data and statistical analysis

The audiometric data, age and gender of the participants were stored by the SONUP application, uploaded to secure servers and retrieved via a secure dedicated website. They were then exported in xls format. Matlab R2021b software (MathWorks, Inc., USA)

was used for statistical analyses, with the significance level set to 5% (p-value < 0.05). ANOVA analysis was used to determine whether the number of trials (number of tests performed) and the voice type had a significant effect or not, followed by *post-hoc* t-tests for multiple comparisons. The diagnostic power (sensitivity and specificity) of the SoNoise_SπN0_Syn test to detect pure-tone average hearing loss was calculated for the study population. Different receiver operating characteristic (ROC) curves were then computed to determine the optimal SNR values to detect a pure-tone average hearing loss at different cut-off levels (20, 30, and 35 dB HL). The best sensitivity and specificity were determined for each of the three SRT values. The optimal sensitivity and specificity were achieved when the Youden index was the highest. Z-score was used to compare the results of the study population with the normative values, its value corresponding to the number of standard deviations separating a result from the normative value. For each individual tested with SoNoise_SπN0_Syn and SoNoise_S0N0_Syn, two different values for SRT (in dB SNR) were obtained. The difference between these values gave the BILD. ANOVA analysis was used to determine whether the hearing loss type had a significant effect or not. Wilcoxon *post-hoc* tests were performed for multiple comparisons, all samples not following a normal distribution (Jarque-Bera normality test).

3 Results

3.1 Normative values of the SoNoise tests

We firstly aimed to compare the use of natural and synthetic voices in a diotic presentation, before determining normative values for the binaural diotic tests, antiphasic test, and BILD. Figure 3A shows the average SRT values for the three tests (SoNoise_S0N0_Na, SoNoise_S0N0_Syn, SoNoise_SπN0_Syn) according to the different trials. The mean SRT for the SoNoise_S0N0_Na training test was −9.0 dB SNR (SD = 1.3) and was −10.0 dB SNR (SD = 1.4), −10.5 dB SNR (SD = 1.6), and −10.8 dB SNR (SD = 1.3) for the next three trials, respectively. The mean SRT for the three trials of SoNoise_S0N0_Syn that followed the SoNoise_S0N0_Na training test were −10.3 dB SNR (SD = 1.4), −10.9 dB SNR (SD = 1.6), and −11.1 dB SNR (SD = 1.6), respectively. A two factor ANOVA showed that trials number had a significant effect on SRT ($p < 0.001$) while the voice (natural and synthetic) had no significant effect ($p = 0.824$). The learning effect was significant for tests using the natural voice, with a difference of 1, 1.5, and 1.8 dB ($p < 0.001$ for all three) between the training and the 1st, 2nd, and 3rd tests, respectively. The 1st test also differed significantly ($p = 0.002$) from the last, with a difference of 0.8 dB. The learning effect was also significant for tests using the synthetic voice, with a difference of 1.3, 1.9, and 2.1 dB ($p = 0.016$, $p = 0.003$, and $p < 0.001$) between the training and the 1st, 2nd, and 3rd tests, respectively. When comparing binaural diotic and antiphasic presentation, a two factor ANOVA showed that both trials number and presentation had significant effects on SRT ($p < 0.001$). A SRT of −17.5 dB SNR (SD = 1.5) obtained in the first SoNoise_SπN0_Syn training test progressively increased to −18.9 dB SNR (SD = 1.3), −19.0 dB SNR (SD = 1.6), and −19.2 dB SNR (SD = 1.9) over the next three trials, respectively. The learning effect was significant with a difference of 1.4 dB between the training and the first test but no significant difference was found between the other consecutive tests. The value also appeared to stabilize around −19.2 dB SNR on the 3rd test, 1.7 dB better than the training test. We then calculated the BILD based on the difference between diotic and dichotic antiphasic SRT values for each individual. BILD values were 8.6 dB (SD = 2.0) for the 1st test, 8.1 dB (SD = 2.1) for the 2nd, and 8.1 dB (SD = 1.9) for the 3rd. One factor ANOVA showed no significant learning effect on the BILD ($p = 0.645$).

Figure 3B displays the distribution of SoNoise tests duration. The mean duration was 167 s (SD = 38 s, median 166 s, IC95 [164–170]). Figure 3C displays the tests duration according to the number of trials performed. The test duration lasted 189.2 s (SD = 41) for the training test, and seemed to stabilize at 158.9 s (SD = 36.8) after four trials.

3.2 Assessment of the diagnostic power of the tests

Participants in the groups of age [40–60] and 4f-PTA [20–40] were fewer in number compared to in the other age groups (Figure 4). All the normative values, sensitivity and specificity of the SoNoise tests are displayed in Table 3.

The threshold of 30 dB HL is the audiometric limit chosen to qualify for reimbursement of hearing aids in France (Journal Officiel de la République Française n°0265, 2018; Joly et al., 2022). The SRTs on the worst ear of the study population ($n = 463$) were plotted against the 4f-PTA (Figure 5A).

The correlation coefficient measured between SRTs of the worst ear and 4f-PTA was $r = 0.797$ ($p < 0.001$). The ROC curves allowed visualization of the sensitivity and specificity according to SNR values at different 4f-PTA cut-offs (20, 30, and 35 dB HL; Figure 5B). For a 4f-PTA of 20 dB HL (mild hearing loss), the optimal threshold value was −14.5 dB SNR and corresponded to a sensitivity of 84% and a specificity of 89%. For a 4f-PTA of 30 dB HL (clinical threshold defining deafness in France), the best threshold value was −13.7 dB SNR and corresponded to a sensitivity of 89% and a specificity of 91%. For a 4f-PTA of 35 dB HL (moderate hearing loss), the best threshold value was −13.0 dB SNR and corresponded to a sensitivity of 88% and a specificity of 93%.

The normative value (defined previously) of the SoNoise_SπN0_Syn training test was measured at −17.5 dB SNR, with a standard deviation of 1.5 dB. Theoretically therefore, 95% of the normal-hearing population would be expected to obtain a score lower (better) than −14.5 dB SNR (Z-score of 2). In this study population, 89.4% of the normal-hearing group, 0% of the unilateral hearing loss group, 17.8% of the symmetrical hearing loss group and 6.7% of the asymmetrical hearing loss group had a Z-Score < 2. The normative value of the SoNoise_S0N0_Syn test was measured at −10.3 dB SNR, with a standard deviation of 1.4 dB. Theoretically therefore, 95% of the normal-hearing population would be expected to obtain a score lower (better) than −7.5 dB SNR (Z-score of 2). In this study population, 95.6% of the normal-hearing group, 20% of the unilateral hearing loss group, 21.7% of the symmetrical hearing loss group and 22.2% of

10.3389/fauot.2024.1292949

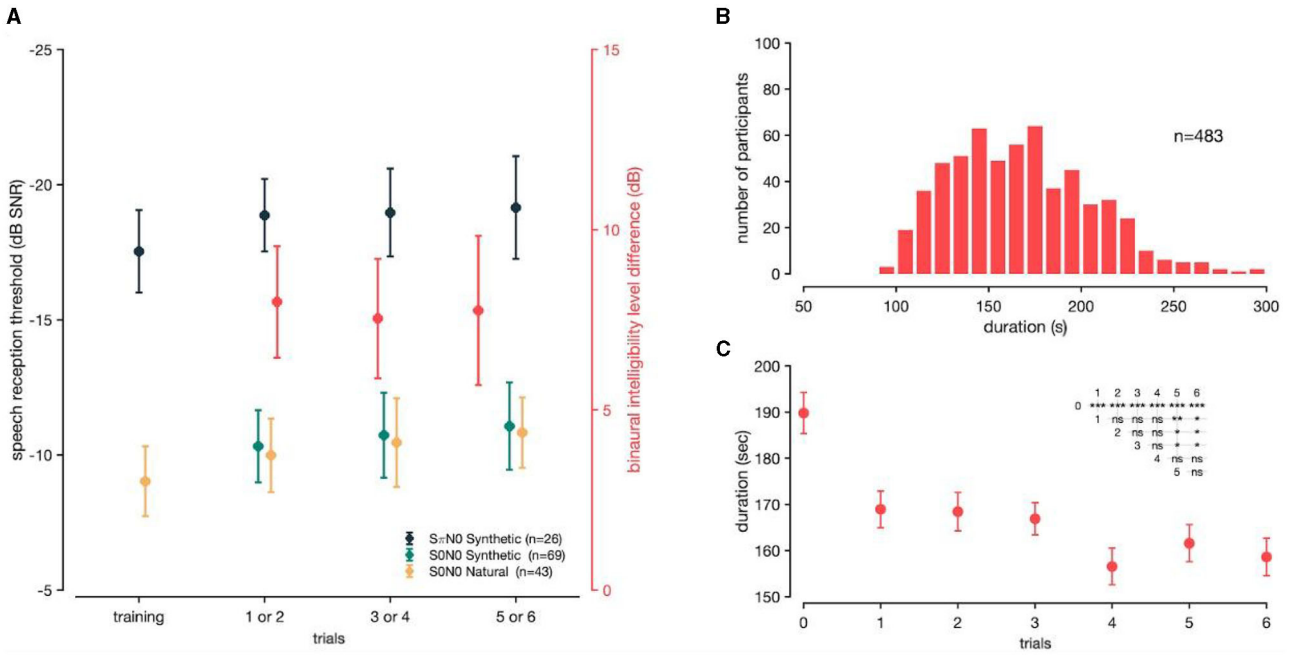


FIGURE 3
(A) Mean SRT and BILD values measured. SπN0 synthetic in dark blue (n = 26), S0N0 synthetic in green (n = 69), and S0N0 natural in yellow (n = 43). Binaural intelligibility level difference in red. The vertical bars represent the standard errors of the mean. **(B)** Histogram distribution of tests duration (n = 483) for SoNoise tests. **(C)** Tests duration according to the number of trials performed. The inserted triangular matrix represents the degree of significance of tests duration as a function of the number of trials (* if $p \leq 0.05$, ** if $p \leq 0.01$, *** if $p \leq 0.001$, "ns" meaning not significant). Trial "0" meaning performed first as a training test.

the asymmetrical hearing loss group had a Z-score < 2. Figure 6 represents the SRTs of the normative and study populations for both SoNoise_SπN0_Syn (A) and SoNoise_S0N0_Syn (B).

We measured the BILD using results of the 399 participants who performed both SoNoise_SπN0_Syn and SoNoise_S0N0_Syn tests. The distribution is given in Figure 7. The mean BILD for the 325 participants with normal hearing was 7.3 dB (SD = 2.1), and for those with hearing loss it was 1.3 dB (SD = 1.4) when unilateral (n = 6), 4.8 dB (SD = 5.4) when symmetrical (n=59), and 1.4 dB (SD=2.4) when asymmetrical (n = 9). ANOVA revealed significant differences in BILD across the hearing loss types ($p < 0.001$). Performing *post-hoc* tests (Wilcoxon test), normal-hearing group presented significantly better BILD than the other three (Sym: $p < 0.001$, Uni: $p < 0.001$, Asym: $p < 0.001$). Participants presenting symmetrical hearing loss had statistically better BILD than those presenting unilateral and asymmetric hearing loss (Uni: $p = 0.035$, Asym: $p < 0.001$). Unilateral and asymmetric hearing loss did not present statistical difference in BILD.

4 Discussion

There were several reasons for developing a new French SIN test. Firstly, accessing equipment for free-field testing currently represents a problem. The development of SIN tests on a tablet with Bluetooth calibrated headphones facilitates their distribution and accessibility. Finally, no French SIN test currently exists enabling the assessment of BILD effect using headphones. It is important to assess the function of both ears working together. BILD measurement thus permitted would provide useful information to categorize hearing loss types with reasonable accuracy, any unilateral or asymmetric hearing loss being revealed by minimal unmasking.

The first part of this study was to determine normative values of SoNoise tests, based on the number of times they are performed. The normative value of the SoNoise_S0N0_Na was −10.0 dB SNR (SD = 1.4), meaning that 95% of the normal-hearing population obtained a score lower (better) than −7.2 dB SNR (Z-score of 2). For the SoNoise_S0N0_Syn, the normative value measured was −10.3 dB SNR (SD = 1.4), meaning that 95% of the normal-hearing population obtained a score lower (better) than −7.5 dB SNR (Z-score of 2). These results can only be compared against SIN tests using binaural diotic presentation. De Sousa et al. (2020) and Prang et al. (2021), respectively found −11.1 dB SNR (SD = 0.8) and −7.1 dB SNR (SD = 1.4) with triplets of words (digits, and digit—common noun—color, respectively) in South African English and English languages. Difference in the results between both tests may be either due to types of words, presentation mode (binaural or monaural) or noise used. They respectively tested 26 and 20 normal-hearing participants (PTA < 15 dB HL). In another study on 202 normal-hearing participants with 4f-PTA better than 25 dB HL, De Sousa et al. (2022) reported a mean SRT of −10.3 dB SNR (SD = 1.3). Matrix tests measured SRT between −10.1 dB SNR (SD = 0.7) and −6 dB SNR (SD = 0.8) for 14 different languages, but with 5 words (Kollmeier et al., 2015). SRT obtained with Matrix tests are higher (worse) than SoNoise and digit in noise (DIN) tests, probably because they use open lists while we used

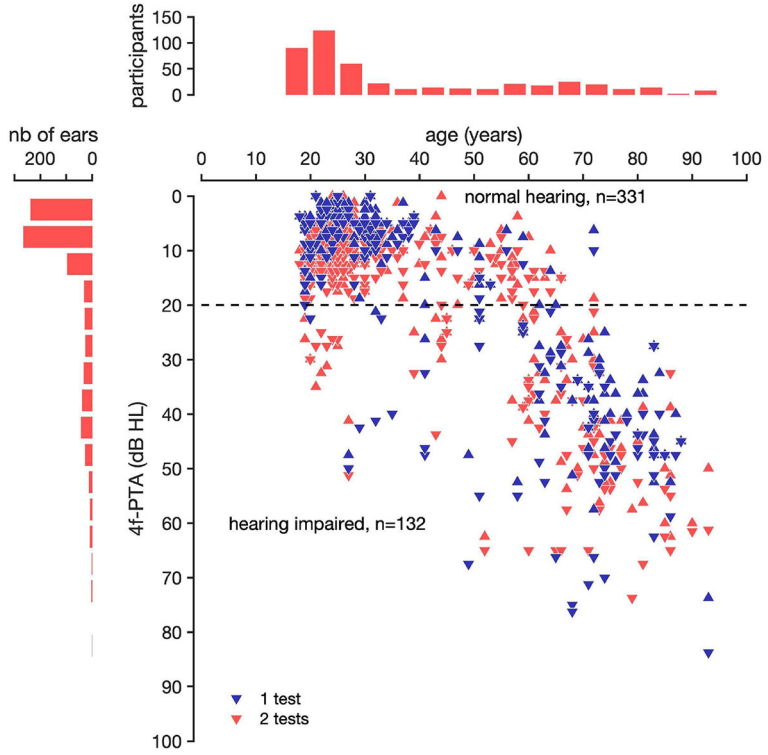


FIGURE 4
Best and worst ear 4f-PTA as a function of age of the 463 participants tested. The horizontal line at 20 dB HL represents the threshold for hearing impairment as defined by the WHO (World Health Organization). Participants were considered as hearing-impaired over 20 dB HL. The results of participants who performed only 1 test (SoNoise_SπN0_Syn) are represented as a red triangle and as a blue triangle if they performed 2 tests (SoNoise_SπN0_Syn + SoNoise_S0N0_Syn). The 4f-PTA was calculated by averaging the hearing thresholds at 0.5, 1, 2, and 4 kHz. Triangles with the point facing up represent the 4f-PTA of the best ears, and triangles with the point facing down represent the 4f-PTA of the worst ears.

closed ones. The normative value of the SoNoise_SπN0_Syn test was −17.5 dB SNR (SD = 1.5), meaning that 95% of the normal-hearing population obtained a score lower (better) than −14.5 dB SNR (*Z*-score of 2). These results can only be compared against SIN tests using dichotic antiphasic presentation. In their study, Smits et al. (2016) tested 16 normal-hearing participants (14 women and 2 men) aged between 19 and 25 years (average 22 yrs) with a DIN test in Dutch and American English languages. They respectively found standards of −15.3 dB SNR (SD = 0.9) and −17.1 dB SNR (SD = 0.9). In their studies, De Sousa et al. (2020, 2022) tested 26 and 243 normal-hearing participants with a DIN test in South African English language, and found standards of −18.4 dB SNR (SD = 1.4) and −17.2 dB SNR (SD = 2.4), respectively. The normative values of the present study are similar to both publications previously cited, but a little closer to that of De Sousa.

Synthetic voice has made enormous progress and is now frequently used in everyday life. Previous studies have evaluated the intelligibility of synthetic voice compared to natural voice. Some found no significant difference in intelligibility between the two (Mirenda and Beukelman, 1987; Paris et al., 1995; Koul, 2003; Nuesse et al., 2019; Ibelings et al., 2022; Schwarz et al., 2022). In their study, Nuesse et al. (2019) measured an SRT of −9.1 dB SNR for the natural voice, and −8.6 dB SNR for the synthetic voice. Tests were performed with the German Matrix test (Wagener et al., 1999). Some others suggested a dependence on synthetic voice quality (Clark, 1983; Pisoni et al., 1985; Greene et al., 1986; Mirenda and Beukelman, 1987, 1990; Kangas and Allen, 1990; Humes et al., 1991; Wolters et al., 2007; Papadopoulos et al., 2009; Cooke et al., 2013; Aoki et al., 2022). Finally, some authors found that natural voice had significantly higher intelligibility than synthetic voice (Koul, 2003; Venkatagiri, 2003; Simantiraki et al., 2018). While intelligibility was found to strongly depend on different features such as speech synthesizer quality, listening conditions and experience (Koul, 2003), the way the words are recorded and the speech material were reported to have only a minimal impact on the results (Van den Borre et al., 2021). We compared the SRT scores obtained in SoNoise_S0N0_Na and SoNoise_S0N0_Syn tests. The results show no significant difference between the average SRT measured for the three trials with respective *p*-values of 0.913, 0.691, and 0.754.

Learning effect is a key element and needs to be considered when performing SIN tests. For natural voice, the learning effect was significant with differences of 1, 1.5, and 1.8 dB ($p < 0.001$ for all three) between the training and the first, second and third tests respectively. For synthetic voice, the learning effect was also significant with a difference of 1.3, 1.9, and 2.1 dB ($p = 0.016$, $p = 0.003$, and $p < 0.001$) between the training and the first, second and third tests, respectively. These results are similar to those found in the literature, which show about 1 dB improvement between the first two tests, and about 2 dB between the first and

TABLE 3 Diotic and dichotic antiphasic SRT for the best and the worst ear respectively, normative and study populations, and both natural and synthetic voices.

		Diotic		Dichotic antiphasic		
		Natural voice ($n = 43$)	Synthetic voice ($n = 69$)	Synthetic voice ($n = 26$)		
		SRT (SD)	SRT (SD)	SRT (SD)	Se (%)	Sp (%)
Normal-hearing	Training	−9.0 (1.3)		−17.5 (1.5)		
	Trial 1	−10.0 (1.4)	−10.3 (1.4)	−18.9 (1.3)		
	Trial 2	−10.5 (1.6)	−10.9 (1.6)	−19.0 (1.6)		
	Trial 3	−10.8 (1.3)	−11.1 (1.6)	−19.2 (1.9)		
				Synthetic voice ($n = 463$)		
Normal- and hearing-impaired	PTA 20 dB HL			−14.5	84	89
	PTA 30 dB HL			−13.7	89	91
	PTA 35 dB HL			−13.0	88	93

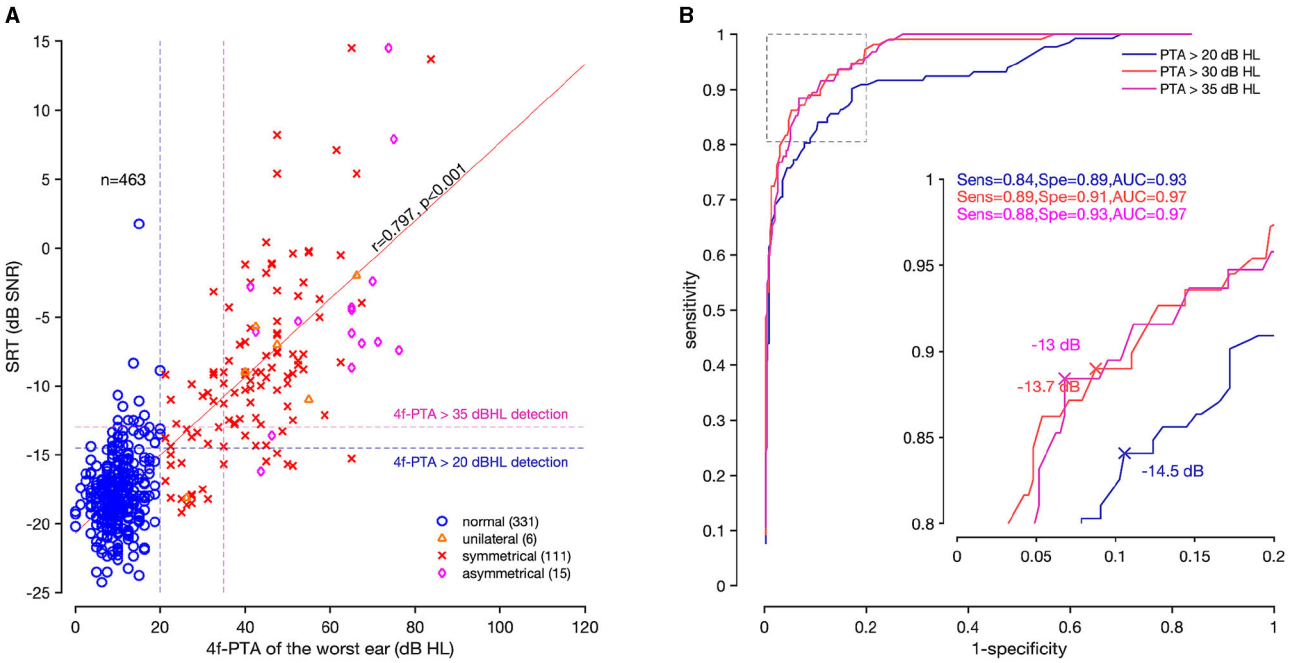


FIGURE 5
Correlation and ROC curves of the SoNoise_Sπ N0_Syn test. **(A)** Correlation coefficient ($r = 0.797$, $p < 0.001$) between SRT and 4f-PTA of the worst ear ($n = 463$). Normal-hearing is represented as a blue circle, unilateral hearing impairment as an orange triangle, symmetrical hearing impairment as a red cross, and asymmetrical deafness as a pink diamond. Two vertical dashed lines are drawn at 20 dB HL (blue) and 35 dB HL (pink). The horizontal dashed lines represent the best SRT value predictive of a 20 dB HL and 35 dB HL 4f-PTA. The top right corner represents participants with abnormal hearing results for both tests (true-positive). The top left corner shows discordant results between 4f-PTA (normal) and SoNoise_Sπ N0_Syn (abnormal; false-positive). The bottom right corner shows discordant results between 4f-PTA (abnormal) and SoNoise_Sπ N0_Syn (normal; false-negative). The bottom left corner represents participants with normal-hearing results for both tests (true-negative). **(B)** ROC curves corresponding to 4f-PTA of 20 dB HL (blue), 30 dB HL (red), and 35 dB HL (pink). The inserted graph enlarges the dotted square.

fourth (Brand and Kollmeier, 2002; Jansen et al., 2012; Kollmeier et al., 2015; Schlueter et al., 2016; Nuesse et al., 2019). However, some studies measured a negligible learning effect after the first training (Hagerman and Kinnefors, 1995; Rhebergen et al., 2008; Paglialonga et al., 2014; Kaandorp et al., 2015; Sheikh Rashid et al., 2017). These findings highlight the need to compare participants' results to the appropriate normative value, according to the number of times the test is performed: i.e., a normative value for screening, and others for diagnosis or follow-up.

To assess the accuracy of the test and thus diagnostic power, we assessed the sensitivity and specificity of the SoNoise_Sπ N0_Syn test at different 4f-PTA cut-off levels. The main characteristic for an accurate SIN test is its ability to detect almost all cases of hearing loss without identifying individuals with normal hearing. For a

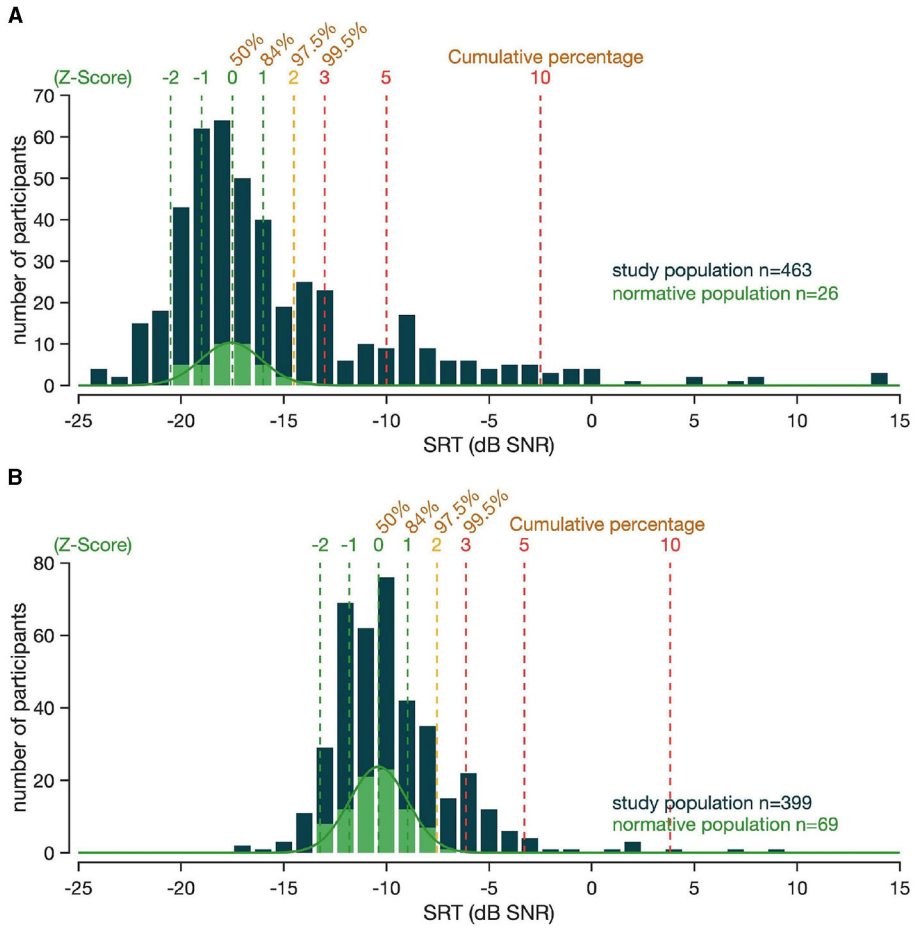


FIGURE 6
(A) SRT distribution of the SoNoise_Sπ N0_Syn. The SRT of the normative population is represented in light green ($n = 26$). The SRT of the study population that participated in the diagnostic power assessment study ($n = 463$) are shown in dark blue. The vertical dashed lines represent the Z-score, with associated percentile values. **(B)** SRT distribution of the SoNoise_S0N0_Syn. The SRT of the population is represented in light green ($n = 69$). The SRT of the study population ($n = 399$) are shown in dark blue. The vertical dashed lines represent the Z-score, with associated percentile values.

4f-PTA of 20 dB HL, the threshold value was −14.5 dB SNR with a sensitivity of 84% and specificity of 89%. For a 4f-PTA of 30 dB HL, the threshold value was −13.7 dB SNR with a sensitivity of 89% and a specificity of 91%. Finally, for a 4f-PTA of 35 dB HL, the threshold value was −13.0 dB SNR with a sensitivity of 88% and specificity of 93%. The French DIN test measured the correlation between DIN SRT and 4f-PTA (0.5/1/2/4 kHz; Ceccato et al., 2021). They tested 167 participants (77 women and 90 men), aged from 19 to 90 years (average 56, SD = 22). Among the 167 participants tested, 66 were classified as having normal hearing, 75 symmetric sensorineural hearing loss, 19 unilateral or asymmetric hearing loss, and seven mixed hearing loss. For a 20 dB HL cut-off value, the best sensitivity and specificity were respectively 96 and 93% with −12.9 dB SNR. For a 40 dB HL cut-off value, the best sensitivity and specificity found were respectively 99 and 83% with −10.3 dB SNR. De Sousa et al. (2020) compared the DIN results of the worst ear to the 4f-PTA (0.5/1/2/4 kHz). For a 40 dB HL cut-off value, the best sensitivity and specificity were respectively 87 and 91% with −14.2 dB SNR. For a 25 dB HL cut-off value, the best sensitivity and specificity were respectively 90 and 84% with −15.7 dB SNR (De Sousa et al., 2022). In their study, 489 participants were tested: 243 were classified as having normal hearing, 172 symmetric sensorineural hearing loss, 42 unilateral or asymmetric hearing loss, and 32 conductive hearing loss. Unfortunately, in their study, Smits et al. (2016) did not report sensitivity or specificity with the PTA.

In the present study, the correlation coefficient found between 4f-PTA and SRT of the SoNoise_Sπ N0_Syn test was $r = 0.797$ ($p < 0.001$). This is consistent with the two studies using dichotic antiphasic presentation which both found $r = 0.82$ (De Sousa et al., 2020; Ceccato et al., 2021). Unfortunately, again, there was no mention of correlation coefficient with the PTA in the study published by Smits et al. (2013).

In this study and those presented, age and hearing-impairment repartition may have a non-negligible effect on the performance of speech reception threshold in noise to predict hearing-impairment based on 4f-PTA. For example, more people with 4f-PTA under 20 dB HL and poor understanding in noise would raise false-positive number, lowering specificity, meanwhile more people with 4f-PTA mild hearing loss but good understanding in noise would raise false-negative number, lowering sensitivity.

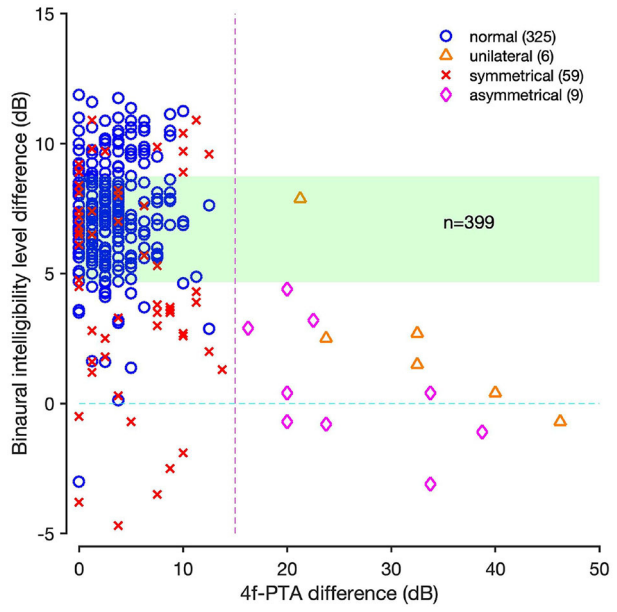


FIGURE 7
BILD according to 4f-PTA difference for the participants who performed both SoNoise_SπN0_Syn and SoNoise_S0N0_Syn tests. Normal ears (worst ear≤20 dB HL) are represented as blue circles, unilateral (best ear<=20 dB HL and worst ear>20 dB HL) deafness as orange triangles, symmetrical (best ear>20 dB HL and diffPTA≤15 dB) deafness as red crosses, and asymmetrical (worst ear>20 dB HL and best ear>20 dB HL and diffPTA>15 dB) deafness as pink diamonds. The light green area represents the normative values of the BILD.

The ROC analysis values must be taken with perspective with the profile of these people misclassified according to their problems. On Figure 5A, participants in the top left corner demonstrated problems understanding in noise, however this would likely be left unmanaged today due to the good 4f-PTA results, considered the main criteria for the management of hearing-impaired people. These people often have real complaints about their capacity to communicate in difficult daily situations and more and more countries consider that poor understanding in noise alone merits treatment. On the opposite, the bottom right corner shows discordant results between 4f-PTA (abnormal) and SoNoise_SπN0_Syn (normal). These participants have mostly good low frequency but poor high frequency thresholds, and can be explained by dichotic antiphasic presentation improving the speech understanding for individuals with symmetric hearing loss but with well-preserved low frequency thresholds (Culling and Lavandier, 2021). These people often have no real complaint in noisy situations but are often eligible for hearing loss management, not for milder hearing loss but for 4f-PTA above 30 dB HL. In this study, their detection is difficult due to the antiphasic presentation, and more of these profiles would have lowered the obtained sensitivity. Studies have shown that diotic presentation fails to detect unilateral and asymmetrical hearing loss due to the dominance of the better ear. Also, this presentation is mostly unaffected by bone conduction hearing loss when presented at suprathreshold levels (De Sousa et al., 2020, 2022; Ceccato et al., 2021). Dichotic antiphasic presentation correlates well with the results of the worst ear, but misses symmetric hearing loss with well-preserved low frequency thresholds.

Each test type having its advantages and disadvantages demonstrates the need for a combined test with both diotic and dichotic antiphasic assessment allowing the discrimination between different types of hearing loss. Another option could be to combine a screening audiometric test and a questionary evaluating the hearing impairment in daily life like the HHIE-S (Ventry and Weinstein, 1982; Duchêne et al., 2022) to assess the problematic of subjects that perform well but still experienced hearing difficulties. Those profiles sometimes get a real benefit from hearing care, even with under-clinical audiometric requirements.

4.1 Binaural intelligibility level difference

The BILD were 8.6 dB (SD = 2.0), 8.1 dB (SD = 2.1), and 8.1 dB (SD = 1.9) for the three SoNoise_S0N0_Syn and SoNoise_SπN0_Syn comparisons, respectively. ANOVA analysis showed no significant learning effect on the BILD (p = 0.645). These results are consistent with those published by De Sousa et al. (2020, 2022) who found a difference between diotic and dichotic presentation of the DIN test between 6 and 8 dB. They reported a SRT of −11.1 and −18.4 dB SNR respectively for diotic and dichotic presentation for the first study, and −10.3 dB SNR and −17.2 dB SNR respectively for the second one. Smits et al. (2016) meanwhile found a smaller BILD (called binaural masking level difference BMLD) with the DIN test in Dutch and US English giving 5.7 and 5.6 dB, respectively. In an additional study (unpublished), Ceccato et al. (2021) tested 19 normal-hearing young adults. They found a SRT of −10.7 dB SNR (SD = 1.3) with the diotic presentation and −15.4 dB SNR (SD = 1.3) with the dichotic antiphasic, both of which are lower than what was measured using diotic and dichotic SoNoise tests with synthetic voice. The differences in BILD may be explained by the specific equalization for binaural diotic and antiphasic presentation for SoNoise test, and by the fact that SoNoise tests do not only use digits.

4.2 Time duration

Concerning test duration, the mean duration of the SoNoise tests was 167 s (SD = 38 s, median 166 s, IC95 [164–170]), and took no longer than other screening SIN tests. Indeed, this duration is consistent with the 3 min measured with the DIN triplet test (Smits et al., 2004; Smits and Houtgast, 2005; Koole et al., 2016; De Sousa et al., 2020). In their studies, they respectively tested: 3,327 adults aged above 50 years (mean = 65 yrs), 38 normal-hearing and hearing-impaired participants (76 ears) among which 22 normal ears and 54 impaired ears, and 39,968 participants during a telephone mass screening study (75% older than 44 yrs of age). The DIN was reported (Smits et al., 2013) to have a 2-minute test duration, although the way this was measured was not mentioned. In their study on 19 normal-hearing and 21 hearing-impaired participants, Jansen et al. (2010) reported a 5-minute duration for the FrDigit3 test, longer than other triplet tests. Where the SoNoise_SπN0_Syn test saves a lot of time is in hearing

screening. SoNoise tests for use in screening and in diagnosis have the advantage of being fast in both cases and thus fulfill a key requirement for performing SIN tests.

4.3 Speech material

The development and validation of SIN tests requires several steps. The equalization phase ensures that each word has a 50% chance of being recognized correctly at the same SNR. Indeed, the generation of the words does not certify an identical difficulty between them. One equalization was done for the SoNoise_SπN0_Syn test, and one for the SoNoise_S0N0_Syn test. We decided to equalize each word separately, without prosody and coarticulation. Normal-hearing participants were tested with 1 to 5 tests, to present all words equally at different SNR levels. Psychometric curves were plotted for each of the words, and an intensity level correction then performed so that each word had a 50% chance of being recognized at the same SNR. The Dutch DIN equalized the whole triplet. Only some have been selected—those with a certain slope and SNR value—and normal-hearing subjects performed the DIN with an adaptive method (Smits et al., 2004). In their study, Jansen et al. (2010) used digits pronounced with natural intonation, but without coarticulation from one digit to another. They selected digits with steep slopes and with SRTs near the average SRT, before having them equalized by normal-hearing subjects with a fixed SNR method. In other studies, different procedures were used (Jansen et al., 2010; Potgieter et al., 2016; Ceccato et al., 2021). For Matrix tests, equalization protocol was more complicated and used coarticulation between the five words of the tests in different languages.

Then, level adjustments were applied: words with high intelligibility were reduced and words with low intelligibility were increased (Kollmeier et al., 2015). In this protocol, the final test was used to perform equalization, as proposed in Brand and Kollmeier (2002), Jansen et al. (2012) and recently by Masalski et al. (2021) for digits. This allows to get information of the item inner difficulties as well as the difficulties induced by their position in the group of words. We showed that the common nouns in the middle were the easiest, followed by the digits presented in first position and the colors in last position that is coherent with studies on the Matrix tests (Brand and Kollmeier, 2002; Jansen et al., 2012; Nuesse et al., 2019). We re-tested the difficulty of items after equalization and found that the item SRT variability dropped from 2.5 to 1 dB of standard deviation. While variations in SRT remained between digits, common nouns and colors, they were slightly reduced as well. Moreover, according to the second evaluation further equilibration in difficulty may be done, and even later by getting the results of the future test that are done. In this study, a different equalization has been performed for each one of the dichotic antiphasic and diotic SoNoise tests, unlike for the DIN tests where the words used for the dichotic antiphasic presentation were equalized using a diotic presentation of the words. The question could be raised as to the impact on a dichotic test of an equalization with diotic word presentation.

The word "sanglier" is trisyllabic. When compared to other words, length appears not to be an issue. The jitter allows a random variation in duration between words, and the equalizations ensure that words are equally complicated to recognize. In their study, Potgieter et al. (2016) separated words with 200 ms of silence and 100 ms of jitter. In the DIN (Smits et al., 2013), the silent interval was 150 ms between digits, and was enlarged or reduced with a random $\pm$50 ms. This reduces the rhythm of the test, and limits whether patients can understand a word only due to its duration. In their studies, Lyzenga and Smits (2011) and Smits et al. (2016) detected no significant difference in the SRT measured between triplets of digits pronounced with and without prosody and coarticulation, using male and female voices, respectively. It appears that coarticulatory cues are no longer available at a SNR of 10 dB (Fernandes et al., 2007). All combinations of triplet words were kept and available for testing, as described elsewhere (Prang et al., 2021).

The masking noise used in these tests was a white noise fitted to the long-term spectrum of the test words (Plomp and Mimpen, 1979; Nilsson et al., 1994; Brand and Kollmeier, 2002; Soli and Wong, 2008; Jansen et al., 2012; Dillon et al., 2016; Potgieter et al., 2016), ensuring that the chosen noise depended on the SoNoise test performed. Indeed, the long-term spectrum of the diotic test words differs from the dichotic antiphasic one, both having different shapes due to the inverted temporal envelopes. This approach therefore differs to that of Ceccato et al. (2021) who used the same masking noise for both diotic and dichotic presentations. In their study, De Sousa et al. (2022) did not specify whether the same masking noise was used for both diotic and dichotic antiphasic tests. In the SoNoise tests, there are differences between speech and noise power spectrum levels below 150 Hz for both natural and synthetic tests (Figure 1). The finite-impulse-response filter used tends to increase the noise spectrum shape, keeping more low frequencies for the noise. This is due to the algorithm's difficulty in generating a filter that follows rapid spectral changes of the speech.

Characteristics of the speech material are displayed in Table 2. The fundamental frequency was higher for natural voice (210 Hz) than for synthetic voice (178 Hz). In the US DIN test, the fundamental frequency was 208.9 Hz for the female voice used (Smits et al., 2016). In the present study, the tests also revealed a longer word duration for natural voice (661 ms) compared to synthetic voice (451 ms), leading to a slower speech rate with 1.6 and 2.3 syllables/s, respectively. In their study, Nuesse et al. (2019) also found a higher fundamental frequency for the natural speech, and a slower speech rate for the natural female OLSA (167 syllables/min) compared to the synthetic female voice (175 syllables/min). The FraMatrix and the French Intelligibility Sentence Test (FIST) measured higher speech rate with respectively 4.2 and 3.6 syllables/s (Jansen et al., 2012; Luts et al., 2008). The FrDigit3 shows a lower speech rate with 1.9 syllables/s (Jansen et al., 2010).

Most of the cited tests are designed for free-field presentation and the only ones specifically designed for headphones presentation are the screening tests. While free-field presentation may be more ecological and very useful for prosthetic evaluation (hearing aids, cochlear implants), a clinical setting is often not suited for a good free-field installation that requires space and stability to ensure reliable calibration. A clinical SIN test administered with headphones may therefore be of interest, especially if the test also measures some of the main binaural functions of the hearing.

4.4 Limitations

The results of this study were obtained using an Android OS tablet (Galaxy Tab A7) and calibrated Orosound Bluetooth headphones, chosen for their capacity for automatic calibration with the high accuracy required in a professional hearing application.

SRT measured in normal-hearing participants with natural and synthetic voices showed no significant differences. It would now be interesting to compare SRT of both voices in hearing-impaired subjects. In this study, the learning effect calculation is not completely reliable. Normal-hearing participants tested to define the normative values performed the tests in a counterbalanced manner, as described elsewhere (Culling et al., 2005; McArdle et al., 2005; Potgieter et al., 2016). If each SoNoise test had been performed in succession by the participants, the normative values would probably have been different. This choice was made in order to carry out intra-individual analyses of the SRT on participants on different tests (natural vs. synthetic voices).

5 Conclusion

SoNoise tests are adaptive SIN self-tests, performed with Bluetooth headphones and a tablet. These SIN tests are the unique French auditory tests operating on the binaural system. Its dichotic antiphasic presentation enables its accurate measurement of SIN abilities. With a duration of only 3 min, it can be used for screening and diagnosis, with the corresponding normative values.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by Institutional Review Board of Montpellier University Hospital (IRB-MTP_2021_09_202100889). The patients/participants provided their written informed consent to participate in this study.

Author contributions

AG: Conceptualization, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing, Data curation. JC: Investigation, Writing – original draft. MB: Funding acquisition, Project administration, Software, Supervision, Writing – review & editing. J-LP: Resources, Validation, Writing – review & editing. FV: Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Validation, Writing – review & editing. J-CC: Conceptualization, Data curation, Formal analysis, Methodology, Project administration, Software, Supervision, Visualization, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

AG is the Chief Scientific Officer of SONUP. MB is the Chief Executive Officer of SONUP. FV and J-CC have relationships with SONUP, including equity, consulting, and potential royalties. FV has been awarded the Early Career Scientific Prize from the Fondation pour l'Audition.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Anderson, S., Parbery-Clark, A., Yi, H. G., and Kraus, N. (2011). A neural basis of speech-in-noise perception in older adults. *Ear Hear.* 32:750. doi: 10.1097/AUD.0b013e31822229d3

Aoki, N. B., Cohn, M., and Zellou, G. (2022). The clear speech intelligibility benefit for text-to-speech voices: effects of speaking style and visual guise. *JASA Expr. Lett.* 2:10274. doi: 10.1121/10.0010274

Apeksha, K., and Kumar, A. U. (2017). Speech perception in quiet and in noise condition in individuals with auditory neuropathy spectrum disorder. *J. Int. Adv. Otol.* 13, 83–87. doi: 10.5152/iao.2017.3172

Bellis, T. J., and Bellis, J. D. (2015). Central auditory processing disorders in children and adults. *Handb. Clin. Neurol.* 129, 537–556. doi: 10.1016/B978-0-444-62630-1.00030-5

Bergeron, F., Berland, A., Fitzpatrick, E. M., Vincent, C., Giasson, A., Leung Kam, K., et al. (2019). Development and validation of the FrBio, an international French adaptation of the AzBio sentence lists. *Int. J. Audiol.* 58, 510–515. doi: 10.1080/14992027.2019.1581950

Boersma, P., and Weenink, D. (2013). Praat: Doing Phonetics by Computer [Computer Program]. *Version 6.1.53*.

Brand, T., and Kollmeier, B. (2002). Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests. *J. Acoust. Soc. A*. 111, 2801–2810. doi: 10.1121/1.1479152

Carhart, R., and Jerger, J. F. (1959). Preferred method for clinical determination of pure-tone thresholds. *J. Speech Hear. Disord*. 24, 330–345. doi: 10.1044/jshd.2404.330

Carhart, R., and Tillman, T. W. (1970). Interaction of competing speech signals with hearing losses. *Archiv. Otolaryngol*. 91, 273–279. doi: 10.1001/archotol.1970.00770040379010

Ceccato, J. C., Duran, M. J., Swanepoel, D. W., Smits, C., De Sousa, K. C., Gledhill, L., et al. (2021). French version of the antiphasic digits-in-noise test for smartphone hearing screening. *Front. Publ. Health*. 9, 725080. doi: 10.3389/fpubh.2021.725080

Clark, J. E. (1983). Intelligibility comparisons for two synthetic and one naturalspeech source. *J. Phonet*. 11, 37–49. doi: 10.1016/S0095-4470(19)30775-2

Cooke, M., Mayo, C., and Valentini-Botinhao, C. (2013). "Intelligibility-enhancingspeech modifications: the hurricane challenge," in *Interspeech* (International Speech Communication Association), 3552–3556. doi: 10.21437/Interspeech.2013-764

Culling, J. F., and Lavandier, M. (2021). "Binaural unmasking and spatial release from masking," in *Binaural Hearing, Vol. 73*, eds. R. Y. Litovsky, M. J. Goupell, R. R. Fay, and A. N. Popper (Berlin: Springer International Publishing), 209–241.

Culling, J. F., Zhao, F., and Stephens, D. (2005). The viability of speech-in-noise audiometric screening using domestic audio equipment : La viabilidad del tamizaje audiométrico con lenguaje en ruido utilizando equipo doméstico de audio. *Int. J. Audiol*. 44, 691–700. doi: 10.1080/14992020500267017

De Sousa, K. C., Smits, C., Moore, D. R., Myburgh, H. C., and Swanepoel, D. W. (2022). Diotic and antiphasic digits-in-noise testing as a hearing screening and triage tool to classify type of hearing loss. *Ear Hear*. 43, 1037–1048. doi: 10.1097/AUD.0000000000001160

De Sousa, K. C., Swanepoel, D. W., Moore, D. R., Myburgh, H. C., and Smits, C. (2020). Improving sensitivity of the digits-in-noise test using antiphasic stimuli. *Ear Hear*. 41, 442–450. doi: 10.1097/AUD.0000000000000775

De Sousa, K. C., Swanepoel, D. W., Moore, D. R., and Smits, C. (2018). A smartphone national hearing test : performance and characteristics of users. *Am. J. Audiol*. 27, 448–454. doi: 10.1044/2018_AJA-IMIA3-18-0016

Dickerson, R., Johnsen, K., Raij, A., Lok, B., Stevens, A., Bernard, T., et al. (2006). Virtual patients: assessment of synthesized versus recorded speech. *Stud. Health Technol. Inform*. 119, 114–119.

Dillon, H., Beach, E. F., Seymour, J., Carter, L., and Golding, M. (2016). Developmentof Telscreen: a telephone-based speech-in-noise hearing screening test with a novel masking noise and scoring procedure. *Int. J. Audiol*. 55, 463–471. doi: 10.3109/14992027.2016.1172268

Duchêne, J., Billiet, L., Franco, V., and Bonnard, D. (2022). Validation of the French version of HHIE-S (Hearing Handicap Inventory for the Elderly - Screening) questionnaire in French over-60 year-olds. *Eur. Ann. Otorhinolaryngol. Head Neck Dis*. 139, 198–201. doi: 10.1016/j.anorl.2021.11.003

Fernandes, T., Ventura, P., and Kolinsky, R. (2007). Statistical information and coarticulation as cues to word boundaries: a matter of signal quality. *Percept Psychophys*. 69, 856–864. doi: 10.3758/BF03193922

Füllgrabe, C., Moore, B. C., and Stone, M. A. (2015). Age-group differences in speechidentification despite matched audiometrically normal-hearing: contributions from auditory temporal processing and cognition. *Front. Aging Neurosci*. 6, 347. doi: 10.3389/fnagi.2014.00347

Gong, L., and Lai, J. (2003). To mix or not to mix synthetic speech and human speech?Contrasting impact on judge-rated task performance versus self-rated performance and attitudinal responses. *Int. J. Speech Technol*. 6, 123–131. doi: 10.1023/A:1022382413579

Grant, K. W., and Walden, T. C. (2013). Understanding excessive SNR loss in hearing-impaired listeners. *J. Am. Acad. Audiol*. 24, 258–273. doi: 10.3766/jaaa.24.4.3

Greene, B. G., Logan, J. S., and Pisoni, D. B. (1986). Perception of synthetic speech produced automatically by rule: intelligibility of eight text-to-speech systems. *Behav. Res. Methods Instr. Comput*. 18, 100–107. doi: 10.3758/BF03201008

Guo, Z., Yu, G., Zhou, H., Wang, X., Lu, Y., and Meng, Q. (2021). Utilizing true wireless stereo earbuds in automated pure-tone audiometry. *Trends Hear*. 25, 23312165211057367. doi: 10.1177/23312165211057367

Hagerman, B., and Kinnefors, C. (1995). Efficient adaptive methods for measuring speech reception threshold in quiet and in noise. *Scand. Audiol*. 24, 71–77. doi: 10.3109/01050399509042213

Hirsh, I. J. (1948). The influence of interaural phase on interaural summation and inhibition. *J. Acoust. Soc. Audiol. Am*. 20, 536–544. doi: 10.1121/1.1906407

Houtgast, T., and Festen, J. M. (2008). On the auditory and cognitive functions that may explain an individual's elevation of the speech reception threshold in noise. *Int. J. Audiol*. 47, 287–295. doi: 10.1080/14992020802127109

Humes, L. E., Nelson, K. J., and Pisoni, D. B. (1991). Recognition of synthetic speech by hearing-impaired elderly listeners. *J. Speech Hear. Res*. 34, 1180–1184. doi: 10.1044/jshr.3405.1180

Ibelings, S., Brand, T., and Holube, I. (2022). Speech recognition and listening effortof meaningful sentences using synthetic speech. *Trends Hear*. 26, 23312165221130656. doi: 10.1177/23312165221130656

Jansen, S., Luts, H., Dejonckere, P., van Wieringen, A., and Wouters, J. (2013). Efficient hearing screening in noise-exposed listeners using the digit triplet test. *Ear Hear*. 34, 773–778. doi: 10.1097/AUD.0b013e318297920b

Jansen, S., Luts, H., Wagener, K. C., Frachet, B., and Wouters, J. (2010). The French digit triplet test : a hearing screening tool for speech intelligibility in noise. *Int. J. Audiol*. 49, 378–387. doi: 10.3109/14992020903431272

Jansen, S., Luts, H., Wagener, K. C., Kollmeier, B., Del Rio, M., Dauman, R., et al. (2012). Comparison of three types of French speech-in-noise tests: a multi-center study. *Int. J. Audiol*. 51, 164–173. doi: 10.3109/14992027.2011.633568

Joly, C. A., Reynard, P., Mezzi, K., Bakhos, D., Bergeron, F., Bonnard, D., et al. (2022). Recommandations de la Société française d'ORL et de chirurgie de laface et du cou (SFORL) et de la Société française d'audiologie (SFA) pour la pratiquede l'audiométrie vocale dans le bruit chez l'adulte. *Ann. françaises d'Oto-rhino-laryngologie et de Pathologie Cervico-faciale*. 139, 20–27. doi: 10.1016/j.aforl.2021.03.004

Journal Officiel de la République Française n°0265 (2018). *Texte n°7—Portant modification des modalités de prise en charge des aides auditives et prestations associées au chapitre 3 du titre II de la liste des produits et prestations prévue à l'article L. 165-1 du code de la sécurité sociale*.

Kaandorp, M. W., Smits, C., Merkus, P., Goverts, S. T., and Festen, J. M. (2015). Assessing speech recognition abilities with digits in noise in cochlear implant and hearing aid users. *Int. J. Audiol*. 54, 48–57. doi: 10.3109/14992027.2014.945623

Kangas, K. A., and Allen, G. D. (1990). Intelligibility of synthetic speech for normal-hearing and hearing-impaired listeners. *J. Speech Hear. Disord*. 55, 751–755. doi: 10.1044/jshd.5504.751

Killion, M. C., and Niquette, P. A. (2000). What can the pure-tone audiogram tell us about a patient's SNR loss? *Hear. J*. 53, 46–48. doi: 10.1097/00025572-200003000-00006

Killion, M. C., Niquette, P. A., Gudmundsen, G. I., Revit, L. J., and Banerjee, S. (2004). Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. *J. Acoust. Soc. A*. 116, 2395–2405. doi: 10.1121/1.1784440

King, S. (2014). Measuring a decade of progress in Text-to-Speech. *Loquens* 1, e006. doi: 10.3989/loquens.2014.006

Kollmeier, B., Warzybok, A., Hochmuth, S., Zokoll, M. A., Uslar, V., Brand, T., et al. (2015). The multilingual matrix test: principles, applications, and comparison across languages: a review. *Int. J. Audiol*. 54, 3–16. doi: 10.3109/14992027.2015.1020971

Koole, A., Nagtegaal, A. P., Homans, N. C., Hofman, A., Baatenburg de Jong, R. J., and Goedegebure, A. (2016). Using the digits-in-noise test to estimate age-related hearing loss. *Ear Hear*. 37, 508–513. doi: 10.1097/AUD.0000000000000282

Koul, R. (2003). Synthetic speech perception in individuals with and without disabilities. *Augment. Alternat. Commun*. 19, 49–58. doi: 10.1080/0743461031000073092

Kramer, S. E., Kapteyn, T. S., and Festen, J. M. (1998). The self-reported handicapping effect of hearing disabilities. *Audiology* 37, 302–312. doi: 10.3109/00206099809072984

Leclercq, F., Renard, C., and Vincent, C. (2018). Speech audiometry in noise: development of the French-language VRB (vocale rapide dans le bruit) test. *Eur. Ann. Otorhinolaryngol Head Neck Dis*. 135, 315–319. doi: 10.1016/j.anorl.2018.07.002

Luts, H., Boon, E., Wable, J., and Wouters, J. (2008). FIST: a French sentence test forspeech intelligibility in noise. *Int. J. Audiol*. 47, 373–374. doi: 10.1080/14992020801887786

Lyzenga, J., and Smits, C. (2011). Effects of coarticulation, prosody, and noise freshness on the intelligibility of digit triplets in noise. *J. Am. Acad. Audiol*. 22, 215–221. doi: 10.3766/jaaa.22.4.4

Masalski, M., Adamczyk, M., and Morawski, K. (2021). Optimization of the speech test material in a group of hearing impaired subjects: a feasibility study for multilingual digit triplet test development. *Audiol. Res*. 11, 342–356. doi: 10.3390/audiolres11030032

McArdle, R. A., Wilson, R. H., and Burks, C. A. (2005). Speech recognition in multitalker babble using digits, words, and sentences. *J. Am. Acad. Audiol*. 16, 726–739. doi: 10.3766/jaaa.16.9.9

Mirenda, P., and Beukelman, D. (1987). A comparison of speech synthesis intelligibility with listeners from three age groups. *Augment. Alternat. Commun*. 3, 120–128. doi: 10.1080/07434618712331274399

Mirenda, P., and Beukelman, D. (1990). A comparison of intelligibility among natural speech and seven speech synthesizers with listeners from three age groups. *Augment. Alternat. Commun*. 6, 61–68. doi: 10.1080/07434619012331275324

Narne, V. K. (2013). Temporal processing and speech perception in noise by listenerswith auditory neuropathy. *PLoS ONE*. 8, e55995. doi: 10.1371/journal.pone.0055995

Nilsson, M., Soli, S. D., and Sullivan, J. A. (1994). Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *J. Acoust. Soc. Am*. 95, 1085–1099. doi: 10.1121/1.408469

Nuesse, T., Wiercinski, B., Brand, T., and Holube, I. (2019). Measuring speech recognition with a matrix test using synthetic speech. *Trends Hear.* 23, 233121651986298. doi: 10.1177/2331216519862982

Paglialonga, A., Tognola, G., and Grandori, F. (2011). SUN-test (Speech Understanding in Noise): a method for hearing disability screening. *Audiol. Res.* 1, e13. doi: 10.4081/audiores.2011.e13

Paglialonga, A., Tognola, G., and Grandori, F. (2014). A user-operated test of suprathreshold acuity in noise for adult hearing screening: the SUN (SPEECH UNDERSTANDING IN NOISE) test. *Comput. Biol. Med.* 52, 66–72. doi: 10.1016/j.compbiomed.2014.06.012

Papadopoulos, K., Koutsoklenis, A., Katemidou, E., and Okalidou, A. (2009). Perception of synthetic and natural speech by adults with visual impairments. *J. Vis. Impairment Blindness* 103, 403–414. doi: 10.1177/0145482X0910300704

Paris, C. R., Gilson, R. D., Thomas, M. H., and Silver, N. C. (1995). Effect of syntheticvoice intelligibility on speech comprehension. *Hum. Fact.* 37, 335–340. doi: 10.1518/001872095779064609

Pisoni, D. B., Nusbaum, H. C., and Greene, B. G. (1985). Perception of syntheticspeech generated by rule. *Proc. IEEE* 73, 1665–1676. doi: 10.1109/PROC.1985.13346

Plomp, R. (1986). A signal-to-noise ratio model for the speech-reception threshold ofthe hearing-impaired. *J. Speech Lang. Hear. Res.* 29, 146–154. doi: 10.1044/jshr.2902.146

Plomp, R., and Mimpen, A. M. (1979). Improving the reliability of testing the speechareception threshold for sentences. *Audiology* 18, 43–52. doi: 10.3109/00206097909072618

Potgieter, J.-M., Swanepoel, D. W., Myburgh, H. C., Hopper, T. C., and Smits, C. (2016). Development and validation of a smartphone-based digits-in-noise hearing test in South African English. *Int. J. Audiol.* 55, 405–411. doi: 10.3109/14992027.2016.1172269

Potgieter, J.-M., Swanepoel, D. W., Myburgh, H. C., and Smits, C. (2018b). The South African english smartphone digits-in-noise hearing test : effect of age, hearing loss, and speaking competence. *Ear Hear.* 39, 656–663. doi: 10.1097/AUD.0000000000000522

Potgieter, J.-M., Swanepoel, D. W., and Smits, C. (2018a). Evaluating a smartphone digits-in-noise test as part of the audiometric test battery. *South Afri. J. Commun. Disord.* 65, 574. doi: 10.4102/sajcd.v65i1.574

Prang, I., Parodi, M., Coudert, C., Legoff, S., Exter, M., Buschermöhle, M., et al. (2021). The simplified French Matrix. A tool for evaluation of speech intelligibility in noise. *Eur. Ann. Otorhinolaryngol. Head Neck Dis.* 138, 253–256. doi: 10.1016/j.anorl.2020.12.003

Rance, G., Ryan, M. M., Carew, P., Corben, L. A., Yiu, E., Tan, J., et al. (2012). Binaural speech processing in individuals with auditory neuropathy. *Neuroscience* 226, 227–235. doi: 10.1016/j.neuroscience.2012.08.054

Rhebergen, K. S., Versfeld, N. J., and Dreschler, W. A. (2008). Learning effect observed for the speech reception threshold in interrupted noise with normal-hearing listeners. *Int. J. Audiol.* 47, 185–188. doi: 10.1080/14992020701883224

Schlueter, A., Lemke, U., Kollmeier, B., and Holube, I. (2016). Normal and time-compressed speech : how does learning affect speech recognition thresholds in noise? *Trends Hear.* 20, 233121651666988. doi: 10.1177/2331216516669889

Schwarz, T., Frenz, M., Bockelmann, A., and Husstedt, H. (2022). *Untersuchung einesynthetischen Stimme für den Freiburger Einsilbertest. GMS Z Audiol (Audiol Acoust). 4:Doc04.*

Sheikh Rashid, M., Dreschler, W. A., and de Laat, J. A. P. M. (2017). Evaluation of an internet-based speech-in-noise screening test for school-age children. *Int. J. Audiol.* 56, 967–975. doi: 10.1080/14992027.2017.1378932

Simantiraki, O., Cooke, M., and King, S. (2018). "Impact of different speech types on listening effort," in *Interspeech* (International Speech Communication Association), 2267–2271. doi: 10.21437/Interspeech.2018-1358

Smits, C., and Houtgast, T. (2005). Results from the Dutch speech-in-noise screening test by telephone. *Ear Hear.* 26, 89–95. doi: 10.1097/00003446-200502000-00008

Smits, C., Kapteyn, T. S., and Houtgast, T. (2004). Development and validation of an automatic speech-in-noise screening test by telephone. *Int. J. Audiol.* 43, 15–28. doi: 10.1080/14992020400050004

Smits, C., Kramer, S. E., and Houtgast, T. (2006). Speech reception thresholds in noise and self-reported hearing disability in a general adult population. *Ear Hear.* 27, 538–549. doi: 10.1097/01.aud.0000233917.72551.cf

Smits, C., Theo Goverts, S., and Festen, J. M. (2013). The digits-in-noise test: assessing auditory speech recognition abilities in noise. *J. Acoust. Soc. A.* 133, 1693–1706. doi: 10.1121/1.4789933

Smits, C., Watson, C. S., Kidd, G. R., Moore, D. R., and Goverts, S. T. (2016). A comparison between the Dutch and American-English digits-in-noise (DIN) tests in normal-hearing listeners. *Int. J. Audiol.* 55, 358–365. doi: 10.3109/14992027.2015.1137362

Soli, S. D., and Wong, L. L. (2008). Assessment of speech intelligibility in noise with the Hearing in Noise Test. *Int. J. Audiol.* 47, 356–361. doi: 10.1080/14992020801895136

Vaillancourt, V., Laroche, C., Mayer, C., Basque, C., Nali, M., Eriks-Brophy, A., et al. (2005). Adaptation of the hint (hearing in noise test) for adult canadianfrancophone populations. *Int. J. Audiol.* 44, 358–361. doi: 10.1080/14992020500060875

Van den Borre, E., Denys, S., van Wieringen, A., and Wouters, J. (2021). The digit triplettest: a scoping review. *Int. J. Audiol.* 60, 946–963. doi: 10.1080/14992027.2021.1902579

Venkatagiri, H. S. (2003). Segmental intelligibility of four currently used text-to-speechsynthesis methods. *J. Acoust. Soc. A.* 113, 2095–2104. doi: 10.1121/1.1558356

Ventry, I. M., and Weinstein, B. E. (1982). The hearing handicap inventory for the elderly:a new tool. *Ear Hear.* 3, 128–134. doi: 10.1097/00003446-198205000-00006

Wagener, K., Kühnel, V., and Kollmeier, B. (1999). Entwicklung und evaluation einessatztests in deutscher sprache i: design des oldenburger satztests. *Z Audiol.* 38, 4–115.

Webster, W. R. (1951). The influence of interaural phase on masked thresholds I. The role of interaural time deviation. *J. Acoust. Soc. Am.* 23, 452–462. doi: 10.1121/1.1906787

White-Schwoch, T., Anderson, S., and Kraus, N. (2020). Long-term follow-up of a patient with auditory neuropathy and normal hearing thresholds. *J. Am. Med. Assoc. Otolaryngol. Head Neck Surg.* 146, 499–501. doi: 10.1001/jamaoto.2019.4314

White-Schwoch, T., Anderson, S., Krizman, J., Bonacina, S., Nicol, T., Bradlow, A. R., et al. (2022). Multiple cases of auditory neuropathy illuminate the importance of subcortical neural synchrony for speech-in-noise recognition and the frequency-following response. *Ear Hear.* 43, 605–619. doi: 10.1097/AUD.0000000000001122

Wilson, R. H., Moncrieff, D. W., Townsend, E. A., and Pillion, A. L. (2003). Development of a 500-Hz masking- leveldifference protocol for clinic use. *J. Am. Acad. Audiol.* 14, 1–8. doi: 10.3766/jaaa.14.1.2

Wolters, M., Campbell, P., DePlacido, C., Liddell, A., and Owens, D. (2007). "The effect ofhearing loss on the intelligibility of synthetic speech," in *Proceedings of the 16th International Congress of the ICPhS* (University of Edinburgh, Edinburgh Research Explorer).

World Health Organization (2021). *World Report on Hearing*. Geneva: World Health Organization. Available online at: https://apps.who.int/iris/handle/10665/339913 (accessed August 5, 2023).

Wu, P. Z., O'Malley, J. T., de Gruttola, V., and Liberman, M. C. (2020). Age-related hearing loss is dominated by damage to inner ear sensory cells, not the cellular batterthat powers them. *J. Neurosci.* 40, 6357–6366. doi: 10.1523/JNEUROSCI.0937-20.2020

Xie, Z., Shen, J., Liu, Y., and Rao, D. (2009). "Calibration of headphones and earphone with KEMAR," in *2009 2nd International Congress on Image and Signal Processing* (Institute of Electrical and Electronics Engineers - IEEE), 1–4. doi: 10.1109/CISP.2009.5304268

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 09 February 2024
DOI 10.3389/fnins.2024.1293120

Check for updates

OPEN ACCESS

EDITED BY
Karina De Sousa,
University of Pretoria, South Africa

REVIEWED BY
Sam Denys,
KU Leuven, Belgium
Jean-Charles Ceccato,
Université de Montpellier, Audiocampus, France

*CORRESPONDENCE
Luke Meyer
✉ l.meyer@rug.nl

RECEIVED 12 September 2023
ACCEPTED 15 January 2024
PUBLISHED 09 February 2024

CITATION
Meyer L, Araiza-Illan G, Rachman L, Gaudrain E and Başkent D (2024) Evaluating speech-in-speech perception via a humanoid robot.
Front. Neurosci. 18:1293120.
doi: 10.3389/fnins.2024.1293120

COPYRIGHT
© 2024 Meyer, Araiza-Illan, Rachman, Gaudrain and Başkent. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Evaluating speech-in-speech perception via a humanoid robot

Luke Meyer[1,2]*, Gloria Araiza-Illan[1,2], Laura Rachman[1,2,3], Etienne Gaudrain[4] and Deniz Başkent[1,2]

[1]Department of Otorhinolaryngology/Head and Neck Surgery, University Medical Center Groningen, University of Groningen, Groningen, Netherlands, [2]University Medical Center Groningen, W.J. Kolff Institute for Biomedical Engineering and Materials Science, University of Groningen, Groningen, Netherlands, [3]Pento Audiology Centre, Zwolle, Netherlands, [4]Lyon Neuroscience Research Center, CNRS UMR 5292, INSERM UMRS 1028, Université Claude Bernard Lyon 1, Université de Lyon, Lyon, France

Introduction: Underlying mechanisms of speech perception masked by background speakers, a common daily listening condition, are often investigated using various and lengthy psychophysical tests. The presence of a social agent, such as an interactive humanoid NAO robot, may help maintain engagement and attention. However, such robots potentially have limited sound quality or processing speed.

Methods: As a first step toward the use of NAO in psychophysical testing of speech- in-speech perception, we compared normal-hearing young adults' performance when using the standard computer interface to that when using a NAO robot to introduce the test and present all corresponding stimuli. Target sentences were presented with colour and number keywords in the presence of competing masker speech at varying target-to-masker ratios. Sentences were produced by the same speaker, but voice differences between the target and masker were introduced using speech synthesis methods. To assess test performance, speech intelligibility and data collection duration were compared between the computer and NAO setups. Human-robot interaction was assessed using the Negative Attitude Toward Robot Scale (NARS) and quantification of behavioural cues (backchannels).

Results: Speech intelligibility results showed functional similarity between the computer and NAO setups. Data collection durations were longer when using NAO. NARS results showed participants had a relatively positive attitude toward "situations of interactions" with robots prior to the experiment, but otherwise showed neutral attitudes toward the "social influence" of and "emotions in interaction" with robots. The presence of more positive backchannels when using NAO suggest higher engagement with the robot in comparison to the computer.

Discussion: Overall, the study presents the potential of the NAO for presenting speech materials and collecting psychophysical measurements for speech-in-speech perception.

KEYWORDS

speech perception, psychophysics testing, speech masking, NAO robot, human robot interaction

1 Introduction

Daily life often presents us with situations in which sounds with overlapping properties originating from different sources compete for our attention. Perception of speech in background noise requires segregating target speech and interfering masker signals. Further, in the case of competing background speech (speech masking), listeners need to suppress the information provided by the masking speech, oftentimes resulting in informational/perceptual masking (Carhart et al., 1969; Pollack, 1975; Mattys et al., 2009). Speakers' voice characteristics facilitate segregating target speech from masking speech (Abercrombie, 1982; Bregman, 1990). Fundamental frequency (F0), related to the pitch of a speaker's voice (e.g., Fitch and Giedd, 1999), and vocal-tract length (VTL), related to the size and height of a speaker (e.g., Smith and Patterson, 2005), are two such speaker voice characteristics often used in differentiating voices and speakers (Skuk and Schweinberger, 2014; Gaudrain and Başkent, 2018). Normal-hearing listeners have been shown to be sensitive to small differences in F0 and VTL cues (Gaudrain and Başkent, 2018; El Boghdady et al., 2019; Nagels et al., 2020a; Koelewijn et al., 2021), and can make effective use of these differences to differentiate between target and masker speech (Darwin et al., 2003; Drullman and Bronkhorst, 2004; Vestergaard et al., 2009; Başkent and Gaudrain, 2016; El Boghdady et al., 2019; Nagels et al., 2021). In contrast, hard-of-hearing individuals who hear via the electric stimulation of a cochlear implant (CI), a sensorineural prosthesis for the hearing-impaired, struggle in such situations, and show less sensitivity to F0 and VTL cues (El Boghdady et al., 2019). This challenge could be due to the inherent spectrotemporal degradation of electric hearing [see Başkent et al. (2016) for more information on the workings of CIs] and thus, difficulty in perceiving various speaker voice cues (Gaudrain and Başkent, 2018; El Boghdady et al., 2019). Therefore, the investigation of these vocal cues through psychophysical testing, both in clinical and research settings, is important. On the other hand, evaluation of speech-in-speech perception requires the use of long and repetitive auditory psychophysical tests to ensure data reliability (Mühl et al., 2018; Smith et al., 2018; Humble et al., 2023). This can be a challenge for individuals being tested, especially for those with short or limited attention spans, such as young children (Hartley et al., 2000; Bess et al., 2020; Cervantes et al., 2023), or those with hearing loss, such as the elderly (Alhanbali et al., 2017). Therefore, any interface or setup that can improve engagement and focus may be helpful in collecting such data.

The use of a computer auditory psychophysics testing has led to methods that allow for better controlled experiments, more complex test designs, and the varying of more test parameters (Laneau et al., 2005). This has subsequently led to the use of desktop or laptop computers as typical test interfaces for auditory psychophysical tests (Marin-Campos et al., 2021; Zhao et al., 2022). When used as the test interfaces, the computer presents stimuli and collects responses. These capabilities have also expanded the potential use of computers for psychophysics testing outside of clinical or highly controlled environments (Gallun et al., 2018). Sometimes interfaces are modified to resemble a game-like format, especially for children (Moore et al., 2008; Kopelovich et al., 2010; Nagels et al., 2021; Harding et al., 2023). However, in a previous study by Looije et al. (2012), it was shown that during learning tasks, the use of a robot was better able to hold the attention of children in comparison to a computer interface. Furthermore, literature has shown that the physical presence of a social actor, both human and human-like, has a greater effect on engagement (Lee et al., 2006), in comparison to its virtual counterpart (Kidd and Breazeal, 2004; Kontogiorgos et al., 2021). This can be leveraged to motivate users to exert more effort during a given task (Bond, 1982; Song et al., 2021). This was also reported by Marge et al. (2022), who comment that a robot can be advantageous in motivating and engaging users. Therefore, it could be that the inclusion of an interactive robot, such as the NAO humanoid robot, could be used to further retain one's attention, especially for psychophysical tests of speech-in-speech perception.

Over the last two decades humanoid robots have gained presence in a wide range of areas, including: high-risk environments (Sulistijono et al., 2010; Kaneko et al., 2019), entertainment (Fujita et al., 2003), home (Asfour et al., 2006), and healthcare (Ting et al., 2014; Choudhury et al., 2018; Saeedvand et al., 2019), to name only a few. Joseph et al. (2018) details more specifically how humanoid robots have been involved in healthcare applications, such as assisting tasks through social interactions (McGinn et al., 2014), telehealthcare (Douissard et al., 2019), and nurse assistive tasks (Hu et al., 2011). The use of social robotics has steadily increased in recent years to the point where they are no longer only being used as research tools, but being implemented in day-to-day life (Henschel et al., 2021). The robots from Aldebaran Robotics (NAO and Pepper) are the two most frequently recurring robots in the field of social robotics. Moreover, the use of both the NAO and Pepper robots has been suggested in the literature as a facilitating interface in testing procedures for hearing research. Uluer et al. (2023), for example, have explored using a Pepper robot to increase motivation during auditory tests with CI children. The NAO has frequently been used in healthcare contexts, as shown in a scoping review by Dawe et al. (2019). Due to the robot's small size and its friendly and human-like appearance, the NAO has been used often in the investigation of child-robot interactions (Amirova et al., 2021). Polycarpou et al. (2016) used a NAO robot with seven CI children between the ages of 5–15 years to assess their speaking and listening skills through play. Although there have been other audiological studies utilising robot interactions, to the best of our knowledge, the evaluation or analysis of the human-robot interaction (HRI) has been limited and has predominantly focussed on task performance.

User engagement (Kont and Alimardani, 2020) is one of the most frequently used metrics in human-robot interaction (HRI) analysis as it provides a measure of interaction quality, and thus one's perception toward an interface. One's own perception toward a robot is often performed using self-assessments, such as the Negative Attitude toward Robots Scale [NARS; (Nomura et al., 2004)]. The NARS is used to determine the attitudes one has toward communication with robots in daily life and is divided into three components: subordinate scale 1 (S1), negative attitudes toward situations and interactions with robots; S2, negative attitudes toward social influence of robots; and S3, negative attitudes toward emotions in interactions with robots. In addition to self-assessments, much can be gleaned regarding the perception toward a robot as well as user engagement through the analysis of behavioural cues using video recordings. Verbal or gestural

behavioural cues, known as backchannels and defined as cues directed back to a conversation initiator to convey understanding or comprehension, and a desire for the interaction to continue (Rich et al., 2010), have also been suggested as measures to evaluate user engagement (Türker et al., 2017).

In this study, we aim to expand the use of a NAO robot in psychophysical evaluations of speech-in-speech perception. Combined with its speech-based mode of communication, the NAO robot could be a relatively low-cost tool for auditory perception evaluation. In both research and clinical contexts, such an implementation could potentially provide participants with an interactive testing interface, possibly helping with engagement and enjoyment during experiments and diagnostic measurements (Henkemans et al., 2017). On the other hand, a number of factors related to the hardware and software of the robot could potentially affect auditory testing. For example, the internal soundcard and speaker combination may not be able to produce sound stimuli of sufficient quality for all psychophysical measurements (Okuno et al., 2002), such as stimuli measured close to hearing thresholds. Non-experimental artefacts such as the noise of the fans or actuators in the robot could add unintentional background noise to the stimuli (Frid et al., 2018). Although the robot could potentially offer beneficial engagement during psychophysical tests, the different test setup with the NAO may impact the quality of the test results. Therefore, we first need to investigate how comparable the results are when conducting a psychophysics test using a robot to those when using the standard computer setup, while also evaluating the engagement factor via HRI analysis.

2 Materials and methods

The present experiment is part of a large project, Perception of Indexical Cues in Kids and Adults (PICKA), and expands on previous work conducted using the same NAO robot for other psychophysical tests (Meyer et al., 2023). The purpose of the PICKA project is to investigate the perception of voice and speech in varying populations, such as normal-hearing and hard-of-hearing adults and children with varying degrees and types of hearing loss and hearing devices, and in varying languages, such as English, Dutch, and Turkish.

In the present study, the PICKA speech-in-speech perception test was used. The speech-in-speech perception test evaluates speech intelligibility of sentences presented in competing speech, using an adapted version of the coordinate response measure (CRM, Bolia et al., 2000; Brungart, 2001; Hazan et al., 2009; Welch et al., 2015). The test was performed via the computer [identical to that reported in Nagels et al. (2021)] as well as with a NAO humanoid robot named "Sam," chosen to represent a gender-neutral name. The computer and Sam versions of the test differ slightly in their implementation, much of which was done intentionally. The implementation differences are further explained in the sections below.

To compare the test performance with the robot to both the standard computer setup and to previous relevant work, we have collected both auditory speech intelligibility scores and data collection duration. To quantify the human-robot interaction (HRI), we have collected data in the form of a questionnaire, the Negative Attitude Toward Robots Scale (NARS), a common HRI metric (Nomura et al., 2004), and behavioural cues exhibited during the experiment to explore engagement related factors.

2.1 Participants

Twenty-nine (aged 19–36; 23.46 ± 4.40 years) individuals took part in the study. Two participants did not meet the inclusion criteria for normal hearing, and therefore data for the speech-in-speech perception test was analysed from 27 participants (aged 19–36; 23.23 ± 4.43 years). However, all 29 participants were included in the analysis of the HRI as there was no inclusion criteria for this component of the study. Sample size was determined based on a rule of thumb for human-robot interaction studies in which it is recommended that a minimum of 25 participants are included per tested condition (Bartneck, 2020), and an extra four participants to account for potential drop-outs. All participants reported English as either native or additional language and completed at least high school education. A pure-tone audiogram was conducted to confirm normal hearing (NH). Hearing thresholds >20 dB HL (Hearing Level) at any of the audiometric octave frequencies (between 250 Hz and 8 kHz) qualified for exclusion. The study was conducted according to the guidelines of the Declaration of Helsinki, and the PICKA project protocol was approved by the Medical Ethical Committee (METc) at UMCG (METc 2018/427, ABR nr NL66549.042.18). Written informed consent was obtained prior to the start of the experiment. The participants were compensated €8/hr for their participation.

2.2 Stimuli for speech-in-speech test

The CRM sentence stimuli used were in English, introduced by Hazan et al. (2009), Messaoud-Galusi et al. (2011), and Welch et al. (2015), and similar in structure to the Dutch sentences used by Nagels et al. (2021). The 48 English sentences contained a carrier phrase with a call sign ("dog" or "cat"), one colour keyword (selected from six colours: red, green, pink, white, black, and blue, all monosyllabic), and one number keyword (selected from eight numbers between 1 and 9, excluding disyllabic seven); e.g., *Show the dog where the pink* (colour) *five* (number) *is*. The same 48 sentences were used to create all stimuli for the present test. Each of the stimuli sets (Dutch and English) of the PICKA test battery were generated by a female speaker with a reference F0 of 242 Hz.

Target and masker sentences were originally produced by the same speaker. Speech-in-speech conditions were implemented by combining target and masker speech with two manipulations: (1) the target-to-masker ratios (TMRs) were varied, and (2) the voice cues F0 and VTL of the masker speech varied to introduce a voice difference between the target and masker speech [see El Boghdady et al. (2019) and Nagels et al. (2021) for a detailed explanation on the influence of TMR and voice cues on speech-in-speech perception]. For TMRs, expressed in dB, three conditions were used (−6 dB, 0 dB, + 6 dB). F0 and VTL voice cues were expressed in semitones (st.), an intuitive frequency increment unit used in music and expressed as 1/12th of an octave. Four different voice conditions were used: (1) the same voice parameters as

the target speech, but with resynthesis to account for synthesis artefacts (ΔF0: 0 st., ΔVTL: 0.0 st.); (2 and 3) a difference of either −12 st. in F0 or + 3.8 st in VTL (ΔF0: −12 st., ΔVTL: 0.0 st.; ΔF0: 0 st., ΔVTL: + 3.8 st.); and (4) a difference of −12 st in F0 and + 3.8 st. in VTL (ΔF0: −12 st., ΔVTL: + 3.8 st.). This resulted in 12 experimental conditions (three TMRs x four voice conditions). An additional condition with no manipulations (no TMR, no voice condition) was included as a baseline condition for a check of the experiment paradigm, but not included in data analyses. Each condition was tested with 7 trials (i.e., 7 target sentences), resulting in a total of 84 experimental trials + 7 baseline trials = 91 trials in the experimental corpus, all tested within one block.

For familiarisation of the test, a small corpus of training stimuli was created with nine F0 and VTL combinations: ΔF0 = −12 st., ΔVTL = 0.0 st.; ΔF0 = −12 st., ΔVTL = + 1.9 st.; ΔF0 = −12 st., ΔVTL = + 3.8 st.; ΔF0 = −6 st., ΔVTL = 0.0 st.; ΔF0 = −6 st., ΔVTL = + 1.9 st.; ΔF0 = −6 st., ΔVTL = + 3.8 st.; ΔF0 = 0 st., ΔVTL = 0.0 st.; ΔF0 = 0 st., ΔVTL = + 1.9 st.; ΔF0 = 0 st., ΔVTL = + 3.8 st. The first two trials had a TMR of 0 dB and the remaining trials a TMR of + 6 dB. Of the nine training stimuli, four were randomly selected for the training phase of the test.

For each trial, a target sentence was randomly selected from the 48 sentences with the "dog" call sign, and the masker speech was prepared from 48 sentences with the "cat" call sign. For the masker speech, random sentences were selected while avoiding sentences with the same number and colour keywords as the target sentence. From these sentences, 150–300 ms segments were randomly selected, applying 50 ms raised cosine ramps to prevent spectral splatter, and concatenating these segments to produce the masker speech. The masker speech started 750 ms before the target sentence onset and continued for 250 ms after the target sentence offset.

2.3 Human-robot interaction evaluation

The HRI was evaluated via the NARS questionnaire and behavioural data captured in video recordings of the experiment. The NARS is presented as a five-point Likert scale (1: strongly disagree—5: strongly agree), used to grade each item, and the higher the score, the more negative an attitude one has toward robots. Total scores for each of the NARS subscales are obtained by totalling the grades of each subscale (S1, S2, S3). Therefore, minimum and maximum scores are 6 and 30 for S1, 5 and 25 for S2, and 3 and 15 for S3. For the video recordings, we analysed behaviours that could be used to indicate engagement (backchannels). "Smiling" and "laughing" (Türker et al., 2017) are two behaviours which can be considered positive backchannels and therefore positive engagement. To characterise negative backchannels, "frowning," and "grimacing" were used as opposites to smiling and laughing.

2.4 Setup

As mentioned previously, the paradigm of the speech-in-speech perception test is based on the CRM, which has been used extensively in the literature (Hazan et al., 2009; Welch et al., 2015; Semeraro et al., 2017; Nagels et al., 2021). In the standard version of the test, to log responses, participants make use of a coloured and numbered matrix representing all possible response combinations (**Figure 1**). Although other tests of the PICKA battery have been modified to resemble game-like interfaces (Nagels et al., 2020a,b; Meyer et al., 2023), the speech-in-speech perception test has not been similarly modified to remain consistent with literature and allow for comparison to previously reported data.

2.4.1 Computer setup

The speech-in-speech perception test was run using MATLAB 2019b (MATLAB, 2019) on an HP Notebook (Intel Core i5 7th gen) running Ubuntu 16.04. The user interface with the standard numbered matrix (**Figure 1**) was used, similar to Nagels et al. (2021). There are two deviations from the aforementioned study: English vs. Dutch stimuli, and use of high-quality headphones vs. internal soundcard and stereo speakers. We made use of the computer's loudspeakers in this study to present a more comparable test setup with the NAO, on which there is no audio connection for headphones.

2.4.2 Robot setup

A NAO V5 H25 humanoid robot developed by Aldebaran Robotics (Sam) was used as an auditory interface to introduce the speech-in-speech perception test and present all corresponding stimuli. The PICKA Matlab scripts were rewritten into Python, which allowed all tests and stimuli to be stored and run directly on Sam. Housed in Sam is an Atom Z530 1.6 GHz CPU processor, 1 Gb RAM, and a total of 11 tactile sensors (three on the head, three on each hand, one bumper on each foot), two cameras and four ultrasound sensors (**Figure 2A**). The software locally installed on the NAO robot is the NAOqi OS, an operating system based on Gentoo Linux specifically created for NAO by the developers. A cross-platform NAOqi SDK (software development kit) framework is installed onto a computer, which can then be used to control and communicate with the robot. The NAO SDKs available are Python (Van Rossum and Drake, 2009), C + + (Stroustrup, 2000), and Java (Arnold et al., 2005). NAO has 25 degrees of freedom and is able to perform movements and actions resembling that of a human.

To improve the useability of running the PICKA tests through Sam, a simple website was designed for the researcher conducting any of the PICKA tests and hosted on Sam. Through this website, displayed on a Samsung Galaxy Tablet A, relevant participant information (e.g., participant ID and language) could be entered and the relevant PICKA auditory test could be initiated (**Figure 2B**). Stimuli were presented through the onboard soundcard, and the internal stereo loudspeakers located in Sam's head. The same tablet depicted a scaled down (approximately by a factor of 1.8) version of the aforementioned standard computer matrix for participants to log their responses (**Figure 1**). Henceforth, the robot and tablet are referred to as the "robot setup" and "auditory interface" refers to the robot only, as the tablet is considered a response logging interface.

2.4.3 Auditory interface calibration

The computer and the NAO inherently differ in their abilities to reproduce sounds due to the different hardware. To measure the

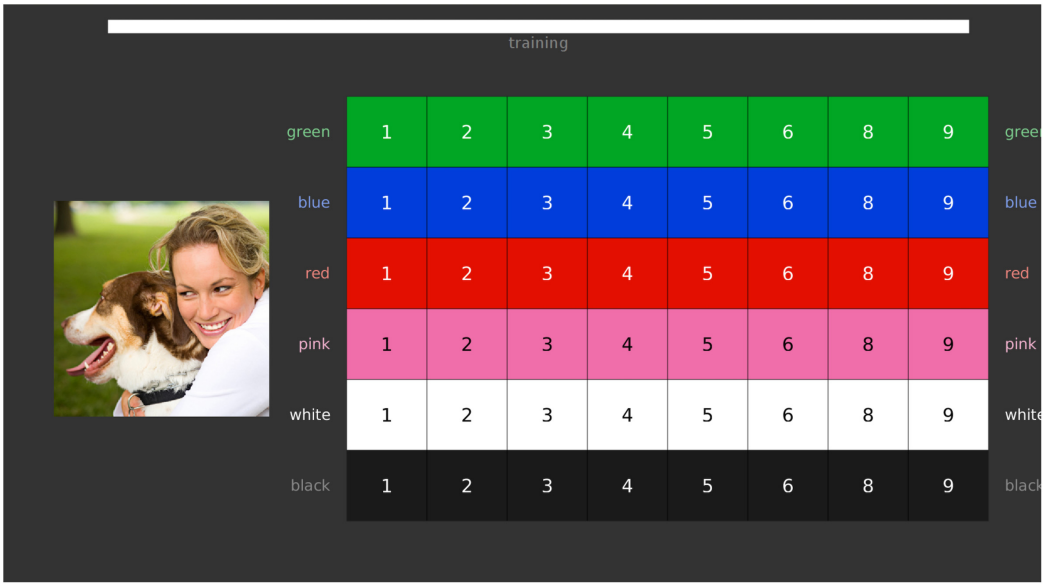


FIGURE 1
The standard computer user interface, showing the speech-in-speech perception coordinate response measure (CRM) test matrix as presented on the screen. Each item in the matrix represents a possible response option, corresponding to the target sentence. Bar at the top of the image depicts progress indicating how many stimuli are remaining in either the training or data collection phases. The matrix and image are published under the CC BY 4.0 licence (https://creativecommons.org/licenses/by/4.0/).

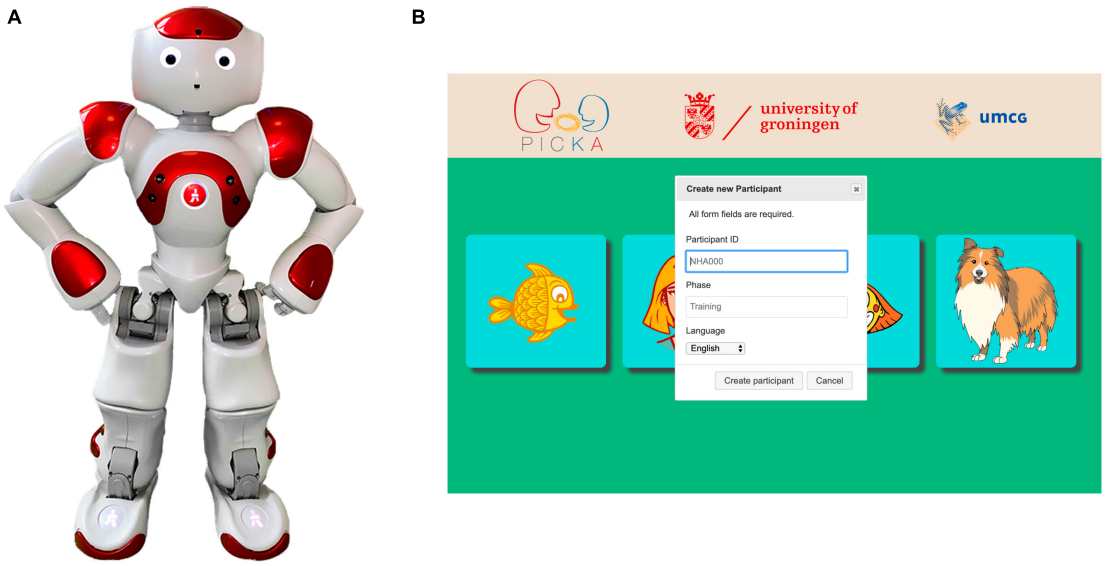


FIGURE 2
Panel **(A)** The robot auditory interface, NAO V5 H25 humanoid robot from Aldebaran Robotics. Panel **(B)** Webpage displayed on the Samsung Tablet to input participant details and begin one of the four PICKA psychophysics tests. Participant details included: participant ID, the phase of the test (either training or data collection), and the language of the test (either English or Dutch). Test buttons from left to right are for starting the different PICKA tests: voice cue sensitivity, voice gender categorization, voice emotion identification, and speech-in-speech perception (the focus of the present experiment), respectively. The cartoon illustrations were made by Jop Luberti for the purpose of the PICKA project. This image is published under the CC BY 4.0 licence (https://creativecommons.org/licenses/by/4.0/).

output of the speakers, a noise signal that was spectrally shaped to match the averaged spectrum of the test stimuli was used. On both the computer and Sam, the noise was presented and measurements were recorded in third-octaves using a Knowles Electronics Mannequin for Acoustic Research (KEMAR, GRAS, Holte, Denmark) head assembly and a Svantek sound level metre (Svan 979). Measurements were conducted in a sound treated room, identical to that used for experimentation. The KEMAR was placed approximately one metre away from the auditory interface, similar to how a participant would be seated during the experiment. Replicating the experimental setup, the sounds were played on both interfaces at the calibrated level of 65 dB SPL (Figure 3). To

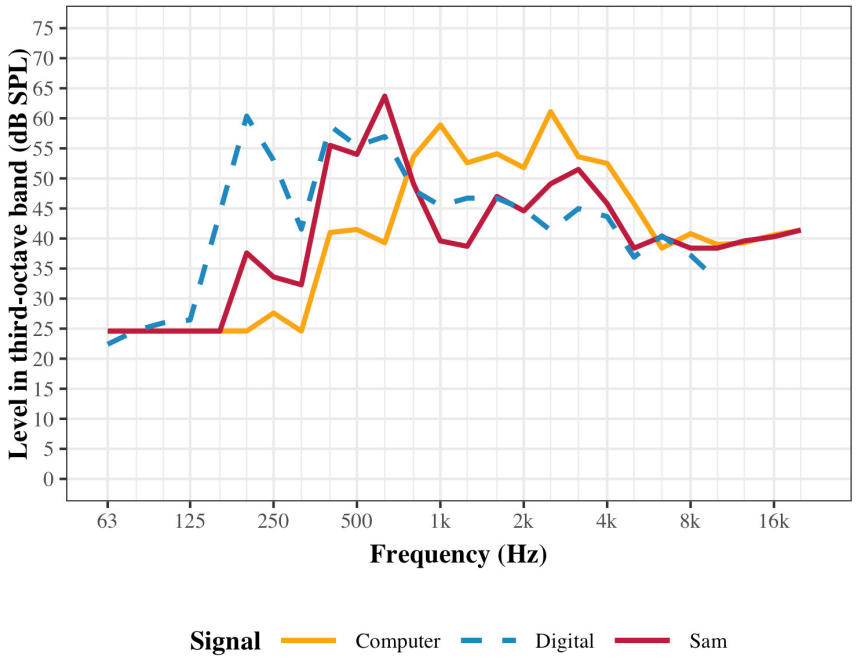


FIGURE 3
Comparison of auditory interface speaker comparison. Yellow and red lines show the levels of the noise signal when presented at the calibrated 65 dB SPL for the computer and Sam, respectively. Each point represents the total power within a third-octave band. The blue line is the digitally extracted levels from the noise signal and shifted to the ideal presentation of 65 dB SPL.

further compare these signals, the digitally extracted levels from the original noise signal used for calibration have also been included in Figure 3 (blue line) to depict its spectral shape.

Figure 3 shows that both the computer and Sam have relatively low level outputs below 250 Hz, compared to the original sound. Furthermore, the computer shows lower levels than Sam at frequencies below 800 Hz. To maintain the overall level of 65 dB SPL, this lack of low frequencies is then compensated above 800 Hz in the computer. These level differences will affect the perceived loudness and timbre of the sounds, and could also potentially affect audibility of lower harmonics in the speech stimuli.

2.4.4 General setup

Participants were seated at a desk with either the computer or Sam and the tablet placed in front of them on the desk in an unoccupied and quiet room. Participants were seated approximately one metre from the auditory interface; however, this varied as participants moved to interact with Sam or the computer. The unused setup was removed from the desk and placed outside the participants' line of sight. To capture the behavioural HRI data, two video cameras were placed to the side and in front of the participant to capture their body positioning and facial expressions, respectively.

2.5 Procedure

Prior to their experimental session, participants were requested to complete the NARS questionnaire online. The order of the setups with which participants started the test was randomised. The speech-in-speech perception test consisted of two phases: a training phase and a data collection phase. The task was the same for both training and data collection. Participants were instructed that they would hear a coherent target sentence with the call sign "dog" that contained both a colour and a number (such as "*Show the dog where the red four is.*") in the presence of a speech masker to replicate a speech-in-speech listening scenario. Participants were also told that the speech masker might be louder, quieter, or have the same volume as the target, or be absent. Participants were instructed to log the heard colour and number combination on the provided colour-number matrix either by clicking with the connected mouse when using the computer or by touching the tablet screen.

Once the participant started the training phase and prior to the presentation of the first training trial, all stimuli for both the training and experimental corpora were processed with all TMR and voice conditions, and the splicing and resynthesis of speech maskers were randomised per participant. The training phase presented participants with four trials to familiarise themselves with the procedure of the test, but the participant responses were not taken into account for scoring purposes. Once confirmed by the researcher that the participant understood the test, the data collection phase started, consisting of a single block of all 91 trials (84 experimental + 7 baseline) with all sentences presented in a random order. Each logged response was then recorded as either correct or incorrect. Responses were only considered as correct when both the colour and number combination were correct. Participants performed the speech-in-speech perception test twice, once on each auditory interface with a break in-between, in a single session lasting approximately 40 min. Following the completion of the first iteration of the test on either the computer or Sam, participants were offered a break by the researcher

before being seated again at the same desk with the next setup placed upon the desk.

When using the computer, participants were presented with the start screen of the test. Once "start" was clicked, the test immediately began with the training phase. Once completed, participants would again be presented with the start screen, which would initiate the data collection phase. No positive feedback was presented to participants; however, negative feedback was presented in the form of the correct colour-number pair briefly being outlined in green before continuing with the next trial. During the data collection phase, at predefined points, breaks would be offered to participants. A pop-up window would inform participants that they could take a break should they wish, and the test would resume when the pop-up window was dismissed.

When using Sam, the robot first introduced itself to the participant before explaining how the test would be carried out. Similar to the computer, first a training phase was presented to participants to familiarise themselves with the robot and the test procedure. Sam informed participants when the training phase was completed and waited for the participant to touch the top of its head to continue to the data collection phase. To maintain motivation and encouragement during the test, both positive (head nod) and negative (head shake) feedback were presented to participants throughout the training and data collection phases, as well as visual feedback to signal when a response could be logged (eyes turning green), and when the response was successfully logged (eyes return to default white). During the data collection phase, at the same predefined points as with the computer, a break was offered to participants. Sam would verbally ask the participant if they wanted to take a break, to which the participant could then verbally reply with either "yes" or "no," If the participant decided to take a break, Sam would ask a follow-up question if they would like to stand up and join in a stretch routine. Again, the participants could respond verbally with either "yes" or "no," If answered with "yes," Sam would stand and perform a short stretch routine. If answered with "no," Sam would stay in a seated position for 10 s before asking if the participant was ready to continue, again awaiting a verbal response. If "yes," Sam continued the experiment. If answered with "no," Sam would allow for another 10 s break before continuing the test. Once all trials were completed, Sam informed participants that they reached the end of the test and thanked them for their participation.

2.6 Data analysis

2.6.1 Test performance

Test performance was quantified by speech intelligibility scores (percentage correct) and data collection duration (minutes) with the computer and Sam setups. Intelligibility scores were calculated by averaging the recorded correct responses across all presented test trials per TMR and voice condition per participant. Data collection durations were calculated from when the first trial was presented until the response of the last trial was logged. Therefore, neither the interactions with Sam in the beginning and end of the test were taken into account, nor the duration of the training phases.

A classical repeated-measures ANOVA (RMANOVA) with three-repeated factors was performed for the intelligibility: the auditory interface with which the test was performed (computer or Sam), the four voice conditions applied to the masker voice (ΔF0: 0 st., ΔVTL: 0 st.; ΔF0: −12 st., ΔVTL: 0 st.; ΔF0: 0 st., ΔVTL: + 3.8 st.; ΔF0: −12 st., ΔVTL: + 3.8 st.), and the three TMR conditions (−6 dB, 0 dB, + 6 dB), resulting in a 2 × 4 × 3 repeated-measures design. When RMANOVA tests violated sphericity, Greenhouse-Geisser corrections were applied (p_{gg}). Evaluation of data collection phase duration was performed using paired *t*-tests.

As the purpose of this study is to present a potential alternative auditory interface to the computer, we aim to look for evidence that both setups (using the computer and Sam) are comparable in their data collection. Therefore, for robustness, we also conducted a Bayesian RMANOVA using the same three-repeated factors as a conclusion of similarity cannot be obtained with classical (frequentist) inference. Bayesian inferential methods focus solely on the observed data, and not on hypothetical datasets as with classical methods. Therefore, they can provide an alternative interpretation of the data, the amount of evidence, based on the observed data, that can be attributed to the presence or absence of an effect [for more detailed explanations see (Wagenmakers et al., 2018)]. The output of Bayesian inferential methods is the Bayes factor (BF) and can be denoted in one of two ways: BF_{01} where $0 < BF < 1$ shows increasing evidence for the null hypothesis as the BF approaches 0, and $BF > 1$ shows increasing evidence for the alternative hypothesis as the BF approaches infinity; and BF_{10}, which is the inverse of BF_{01}; i.e., $0 < BF < 1$ shows evidence for the alternative hypothesis, and $BF > 1$ shows evidence for the null hypothesis. The two notation methods can be used interchangeably for easier interpretation depending on the inference to be made. Since the intended focus of the inference of this study is evidence for the null hypothesis, the BF_{10} notation is used. The degree of evidence is given by different thresholds of the BF: anecdotal, $0.33 < BF < 1$ or $1 < BF < 3$; medium, $0.1 < BF < 0.33$ or $3 < BF < 10$; strong, $0.03 < BF < 0.1$ or $10 < BF < 30$.

2.6.2 Human-robot interaction

Analysis of the NARS was performed using one sample *t*-tests were performed for each subscale to determine if the results were significantly different from the expected means (18, 15, and 9 for S1, S2, and S3, respectively), which would indicate neutrality toward interactions with robots, and thus an unbiased sample.

To analyse the behavioural data from the video recordings, two independent coders viewed the recordings and logged the frequency of displayed behaviours using the behavioural analysis software BORIS (Friard and Gamba, 2016). Total duration of raw video footage was approximately 23 h 57 min. To reduce the workload of coders, video recordings were post-processed and segments of different phases of the test were extracted. Segments were pseudo randomised and concatenated, resulting in approximately 8 h 23 min of footage to be coded. Due to the repetitive nature of the test, these segments would provide "snapshots" during the different phases. Segments were created as follows: 35 s from the introduction when using Sam (introduction in its entirety); 30 s from the training phase for both the computer and Sam; 2 min from the beginning, 1 min from the middle, and 2 min from the end of the data collection phase for both the computer and Sam; 7 s from the break during the data collection phase in the case where the total duration was less than 10, or 45 s if the break was up to a minute. Engagement was assessed using the frequency of backchannels recorded by the two coders

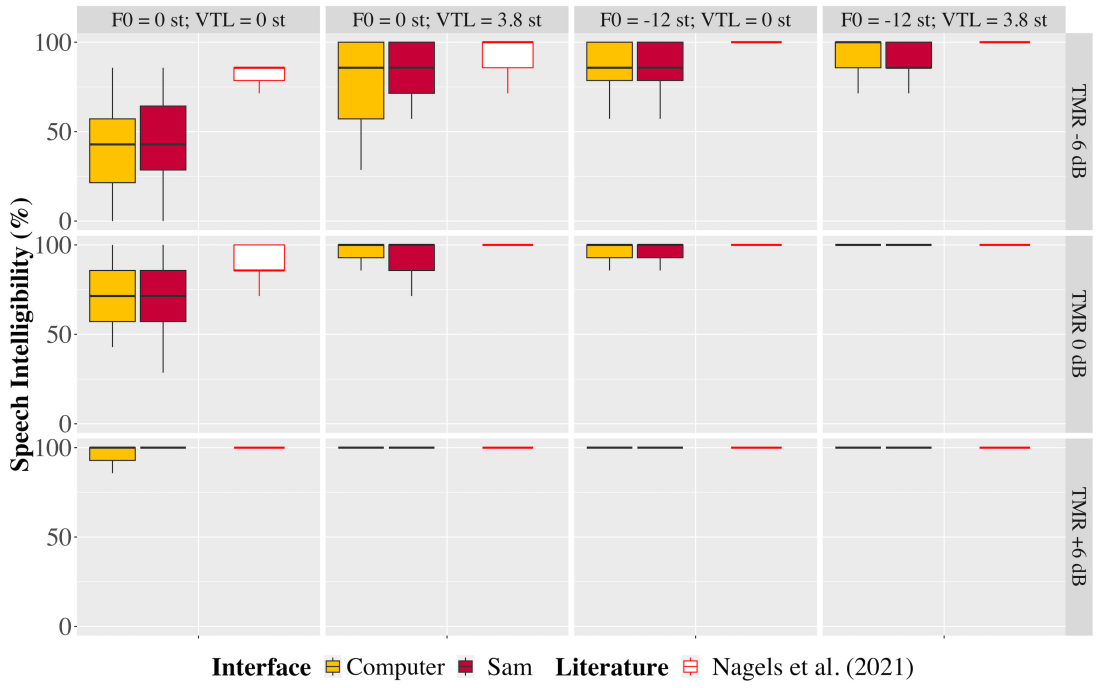


FIGURE 4
Boxplots depicting the range, quartiles, and median percent correct scores of the speech-in-speech perception test, shown for each talker-to-masker ratio (TMR, rows from top to bottom) and voice condition (columns from left to right) for the computer and Sam setups (yellow and red filled boxes, respectively), and in comparison to data reported by Nagels et al. (2021; empty boxes).

and were compared both within and between coders. Within coder comparisons were performed using Student *t*-tests. Reliability between coders was evaluated using intraclass correlation [ICC; (Bartko, 1966)] based on the frequency of exhibited behaviours during each of the concatenated video segments. An ICC analysis is often used for ordinal, interval, or ratio data (Hallgren, 2012). Because the frequency of behaviours is analysed per interval of the full video recording, as well as all subjects are observed by multiple coders, this makes an ICC appropriate.

3 Results

3.1 Test performance

3.1.1 Speech intelligibility scores

The baseline speech intelligibility scores with no speech masker showed good consistency of the experimental paradigm: 99.0% on average when using the computer, and 99.5% on average when using Sam. Figure 4 shows the intelligibility scores per TMR and voice condition across all participants. Table 1 shows the results of both the classical and Bayesian RMANOVAs performed across both setups, three TMRs and four voice conditions. Results of the classical RMANOVA showed no significant difference between participants' intelligibility scores when using the computer or Sam [$F_{(1,36)} = 1.090$, $p = 0.306$, $n_p^2 = 0.040$], no significant interaction between the auditory interface and the TMR [$F_{gg(1.490,38.746)} = 0.065$, $p_{gg} = 0.888$, $n_p^2 = 0.003$], no significant interaction between the auditory interface and the voice condition [$F_{gg(2.353,61.182)} = 0.673$, $p = 0.537$, $n_p^2 = 0.025$], and no significant interaction between all three factors [$F_{(3.643,94.730)} = 0.587$, $p = 0.657$, $n_p^2 = 0.022$].

Bayesian RMANOVA showed moderate evidence that the auditory interface on which the test was performed did not affect the results of the speech-in-speech perception test ($BF_{10} = 0.185$), strong evidence of no interaction between the auditory interface and the TMR ($BF_{10} = 0.081$), strong evidence of no interaction between the auditory interface and the voice condition ($BF_{10} = 0.060$), and strong evidence of no interaction between all three factors ($BF_{10} = 0.039$).

3.1.2 Data collection duration

Figure 5 shows the duration of the speech-in-speech perception test when performed using each auditory interface, and in comparison, to previous data reported by Nagels et al. (2021). The average duration of the data collection phase was 9 ± 1 min on the computer and 15 ± 5.1 min on Sam. However, we observed that three outlier participants took substantially longer to complete the data collection phase when using Sam. Removing these outliers resulted in an average duration of 13 ± 1 min. The removal of the outliers showed that they had a significant effect on the total duration of the data collection phase [$t(45) = -12.22$, $p < 0.001$].

3.2 Human-robot interaction

Average scores for the subscales were 14.8 ± 3.74, 15.8 ± 2.17, and 8.5 ± 1.91 out of possible totals 30, 25, and 15 for S1, S2, and S3, respectively. One sample *t*-tests for each subscale showed only a statistically significant difference to the expected mean for

TABLE 1 Results of the classical and Bayesian RMANOVAs.

		Classical RMANOVA			Bayesian RMANOVA
	Case	Sphericity Correction	*F*, *p*	η_p^2	B_{10}
Main factors	Primary test interface	None	$F_{(1,36)} = 1.090$, $p = 0.306$	0.04	0.185
	TMR	Greenhouse-Geisser	$F_{(1.186,30.826)} = 147.980$, $p < 0.001$	0.851	5.54E + 18
	Condition	Greenhouse-Geisser	$F_{(1.982,51.526)} = 131.767$, $p < 0.001$	0.835	3.50E + 26
Interactions	Primary test interface × TMR	Greenhouse-Geisser	$F_{(1.490,38.746)} = 0.065$, $p = 0.888$	0.003	0.081
	Primary test interface × Condition	Greenhouse-Geisser	$F_{(2.353,61.182)} = 0.673$, $p = 0.537$	0.025	0.06
	Condition × TMR	Greenhouse-Geisser	$F_{(4.002,104.461)} = 50.928$, $p < 0.001$	0.662	2.99E + 30
	Primary test interface × TMR × Condition	Greenhouse-Geisser	$F_{(3.643,94.730)} = 0.587$, $p = 0.657$	0.022	0.039

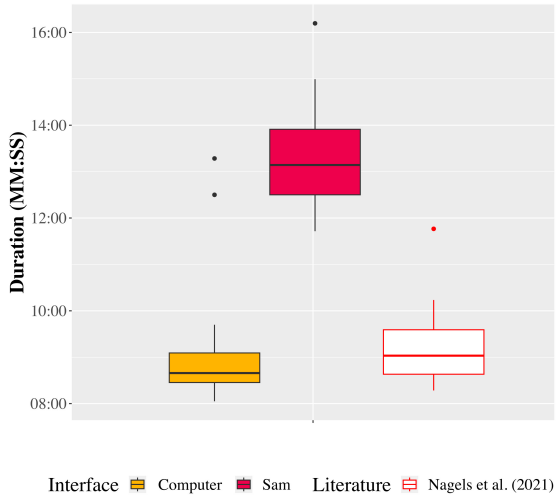


FIGURE 5
Duration to complete the data collection phase of the speech-in-speech perception test on the computer and Sam setups (following the removal of three outliers), and in comparison to data reported by Nagels et al. (2021).

S1 [t(19) = −3.83, $p < 0.01$], and non-significant differences for S2 and S3. The results are summarised in Table 2 below.

Behavioural coding results (Figure 6) showed on average (after pooling all backchannels) more frequent "frowning" when using the computer, although not statistically significant [$t(1.493) = 0.721$, $p > 0.05$], and significantly more frequent "smiling" when using Sam [$t(1) = -13$, $p < 0.05$]. "Grimacing" and "laughing" showed near identical frequencies between the two auditory interfaces. Intraclass correlation showed poor absolute agreement between coders for the behaviours "frowning" [ICC(*2, k*) = 0.175] and "laughing" [ICC(*2,k*) = −0.375], and high correlation for the behaviours "grimacing" [ICC(*2, k*) = 0.671] and "smiling" [ICC(*2, k*) = 0.697].

4 Discussion

The aim of the present study was to evaluate Sam as an alternative auditory interface for the testing of speech-in-speech perception. To explore this, we compared the test performance data (both percent correct scores of intelligibility and data collection phase duration) obtained from normal-hearing young adults for the speech-in-speech perception test when using the proposed robot setup, to data when using the standard computer setup, as well as to previous studies using similar methods. Due to the inherent repetition of the speech-in-speech perception test, we propose Sam to offer an engaging experience for participants when conducting such a psychophysical test. Although there have been other studies in which psychophysical tests have been gamified to offer more engagement (Moore et al., 2008; Nagels et al., 2021; Harding et al., 2023), there may be certain tests for which gamification may not be appropriate, either to be consistent with literature, or gamification may result in an overcomplication (e.g., Hanus and Fox, 2015) of the test, having instead the opposite effect. In such cases, it may be beneficial to incorporate a social agent to facilitate engagement, not only due to its presence, but also playing an active role in the procedure. To explore this, we have also evaluated engagement with the two setups using an HRI questionnaire and analyses of behavioural data from video recordings.

4.1 Test performance

4.1.1 Speech intelligibility scores

Results of the classical RMANOVA showed no significant difference between the percent correct scores obtained when using the computer or Sam. In addition, there was no significant interaction between the auditory interface and TMR, auditory interface and voice condition, or a combination of auditory interface, TMR and voice condition. Results of the Bayesian RMANOVA reflected the results of the classical RMANOVA,

TABLE 2 Results of the one sample *t*-tests comparing each of the NARS subscales to their respective expected means (indicating neutrality).

Questions	Subscale	Expected mean	Mean	SD	95% CI	t-score	*p*
I would feel uneasy if I was given a job where I had to use robots.	S1	18	14.8	3.73	[13.05, 16.55]	−3.83	<0.01
The word "robot" means nothing to me.							
I would feel nervous operating a robot in front of other people.							
I would hate the idea that robots or artificial intelligences were making judgements about things.							
I would feel very nervous just standing in front of a robot.							
I would feel paranoid talking with a robot.							
I would feel uneasy if robots really had emotions.	S2	15	15.8	2.17	[14.79, 16.81]	−1.65	N.S.
Something bad might happen if robots developed into living beings.							
I feel that if I depend on robots too much, something bad might happen.							
I am concerned that robots would be a bad influence on children.							
I feel that in the future society will be dominated by robots.							
I would feel relaxed talking with robots*	S3	9	8.5	1.91	[7.61, 9.39]	−1.17	N.S.
If robots had emotions I would be able to make friends with them.*							
I feel comforted being with robots that have emotions.*							

The * symbol indicates inverse items.

showing strong evidence in support of the two auditory interfaces being functionally identical. Visual inspection of **Figure 4** also shows that the spread of the data between the TMRs and voice conditions are identical between the two auditory interfaces, and in comparison, to data reported by Nagels et al. (2021). It is also illustrated that most incorrect answers were given when the TMR was −6 dB, and a clear ceiling effect was observed at the TMR of + 6 dB. The relatively higher percent correct scores for all conditions in the data reported by Nagels et al. (2021) could be due to several reasons. One possibility is that in their study the participants used high-quality headphones instead of the built-in loudspeakers of the computer. In addition, their stimuli were Dutch, whereas the stimuli presented to participants in the present study were English. Although Nagels et al. (2021) used Dutch stimuli, their population consisted of native Dutch-speaking participants. In the present study, participants reported English as either their native or an additional language. Therefore, the lower intelligibility scores seen in the present study in comparison to those reported by Nagels et al. (2021) may be due to a non-native effect. It is not expected that the structure of the CRM sentences would affect the intelligibility of the sentences since the paradigm of the sentence structure is intended to work across languages, as suggested by Brungart (2001). However, the English stimuli were presented by a British English speaker. This may have also affected the intelligibility of the target sentences in the presence of the masker sentences, especially in the −6 dB TMR condition, for the non-native English-speaking participants who may be more acquainted with US English, for example.

While we attempted to replicate the test procedure of Nagels et al. (2021) as closely as possible, as has been detailed above, there were some differences in the implementation of the test between the computer and Sam. Some of these implementation differences were necessary to perform a fairer comparison between the computer and Sam, but others were related to the interaction between the participant and Sam. These differences may have inadvertently introduced differences in the overall percent correct scores, resulting in the lower intelligibility scores.

In the present study, between the computer and Sam, stimuli, language, and target and masker speaker were kept consistent. Several factors were postulated to potentially limit the usability of the robot, such as the soundcard, speaker quality, processing speed, and non-experimental artefacts. An analysis of the speaker quality of the two auditory interfaces showed that there was a reduced quality of the computer in comparison to Sam, especially at lower frequency ranges. However, the consistent scores of the speech-in-speech perception test show that despite these limitations and the implementation differences between the computer and Sam, both setups were capable of presenting and collecting comparable test data. In addition, both the computer and Sam showed similar patterns in test results for the different TMR and voice conditions to those reported in literature. Therefore, the

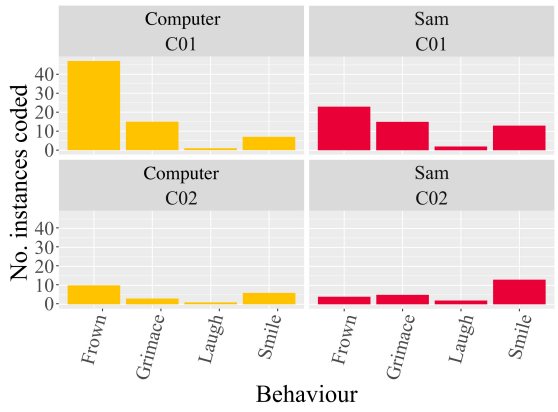


FIGURE 6
Coded backchannel frequencies by coder 1 (C01) and coder 2 (C02) when using both the computer and Sam.

comparable results between the computer and Sam, and previously reported data, indicate that Sam can be used as an effective auditory interface for the speech-in-speech perception test with a normal-hearing population.

4.1.2 Data collection duration

The duration to complete the data collection phase of the speech-in-speech perception test was longer when using Sam in comparison to when using the computer; however, this increased duration did not seem to affect the performance of Sam's setup for collecting comparable intelligibility scores. The three outliers removed from the data collection duration were the first three participants with whom this test was conducted. During the experimental procedure with these participants, it was discovered that the pauses between stimuli were increasing. This was determined to be due to how response data was saved during the test; with each response given, the size of the save file increased, resulting in a longer duration to open and write to the file. Upon discovering this response saving issue, the test code was amended to save the results to a smaller file format during the test and subsequently saving the full results once the test was completed, thus rectifying the duration problem.

However, it can still be seen in **Figure 5** that, even without the outliers, the duration of the data collection phase when using Sam was much longer than that when using the computer. We have considered several factors that could contribute to this difference. Potential delays due to online stimulus preparation were ruled out, since the stimulus corpus was created prior to the training phase. Further investigation into the data collection phase durations per participant showed that on average, there were 6 s between the logging of one response and the logging of the next response when using the computer. With Sam, however, this was on average 9 s. Closer analysis of this 3 s difference showed that this occurs due to the feedback presented to participants following their response logging (head nod or head shake). Subsequent to the completion of data collection, separate measurements were taken by timing the duration of the head movements of Sam. On average, when a correct response was given, timings showed that it took 2.5 s for Sam to nod its head and then present the next stimulus. When an incorrect response was given, this time was on average 3.2 s for Sam to shake its head before presenting the next stimulus. Bootstrap simulations using the mean accuracy as the probability of a correct or incorrect response (and thus a head nod or head shake) for the various tested conditions showed that on average, the movement of Sam's head added 3.9 min ± 2 s over the 91 trials. No positive feedback and brief negative feedback (outlining of the correct response) was presented to participants when using the computer. The inclusion of positive and negative feedback when using Sam, although different to the computer implementation, was done to increase the social presence of the robot (Akalin et al., 2019).

4.2 Human-robot interaction

As mentioned previously, engagement during repetitive auditory tasks can be challenging, especially for certain populations, and to address this challenge we propose the use of a humanoid NAO robot. The use of such an interface for these tasks, at its core, relies on interactions, consisting of both social and physical components, between humans and the robot. The NARS questionnaire we used was developed as a measure of one's attitudes toward communication robots in daily life (Nomura et al., 2004). The NARS is further broken down into three subscales to identify the attitudes of individuals toward social interactions with robots where the higher the score, the more an individual has negative attitudes toward those situations. The subscales are: S1, negative attitudes toward situations and interactions with robots; S2, negative attitudes toward social influence of robots; and S3, negative attitudes toward emotions in interactions with robots. Performing such a questionnaire prior to any interaction involving a robot allows it to be used as a cross-reference to explain any potential skewing of subsequently collected HRI data following the interaction. Results of the NARS subscales showed that only S1 was statistically different from the expected mean. The non-significant results of subscales S2 and S3 indicate that participants had neutral attitudes towards the social influence of robots and emotions in interactions with robots, respectively. However, the lower average S1 score indicates that participants had overall a relatively positive attitude toward situations of interactions with robots prior to their interaction with Sam. This is also reflected in the behavioural backchannels, coded from the video recordings. These showed more frequent smiling when using Sam in comparison to the computer, indicating both a state of comfort and engagement with Sam. This is contrasted by the more frequent frowning (although not significant, can be seen visually in **Figure 6**) when using the computer, which could indicate either a state of confusion (Rozin and Cohen, 2003) or contemplation (Keltner and Cordaro, 2017). Due to the nature of the speech-in-speech perception test and its fluctuating difficulty (especially when the TMR is −6 dB and where the target and masker speech did not differ in voice cues, the most difficult listening conditions tested), the more likely interpretation of the frowning is contemplation as participants focus harder in the more difficult voice conditions. Although this appears to be more frequent with the computer, this is not necessarily to say that the computer requires more focus. With both setups, this directed focus may subsequently lead to mental fatigue during

the task (Boksem and Tops, 2008). However, the results of the speech-in-speech perception test show that this increased directed focus does not affect the outcome of speech intelligibility between the computer and Sam.

4.3 General remarks

Our overall results show that the NAO robot shows promise to be used as an auditory interface for speech-in-speech testing. This finding is in line with and adds to our previous work (Meyer et al., 2023), which evaluated the test performance from two other PICKA tests (voice cue sensitivity and voice gender categorization). Voice cue sensitivity test measures the smallest difference between two voice cues a listener can hear. The linguistic content seems to have little effect on the voice cue perception (Koelewijn et al., 2021), and the perceived voice could be biassed by the perceived gender of the robot (Seaborn et al., 2022). Speech-in-speech perception relies not only on processing voice and speech cues, but also on modulating attention and inhibition to separate target speech from masker speech, and further use of cognitive and linguistic mechanisms to decode the lexical content. It is not clear if a voice bias due to perceived robot gender would affect the speech intelligibility scores (Ellis et al., 1996). Despite such differing natures of these tests, our findings were consistent, and both showed comparable test performance with both setups.

4.4 Future directions

In comparing the test performance between the two setups, the only significant difference between the computer and Sam was the increased duration of the speech-in-speech perception test when using Sam. Although this is predominantly due to the presentation of positive and negative feedback to participants following the logged responses, we believe that it is an important component in establishing and maintaining the social presence of Sam. Therefore, instead of attempting to decrease the overall duration of the speech-in-speech perception test on Sam by removing the visual feedback, the social interaction with Sam could be improved. This way, we accept the longer duration with the inclusion of the feedback but provide the participant with a more natural interaction when performing the test. One such way this can be accomplished is by removing the use of the Samsung Galaxy tablet, which pulls the attention away from Sam with every response and replacing it with speech recognition on Sam. This would maintain the interaction with Sam both by not forcibly moving the participants' attention between Sam and the tablet, but also by engaging in more natural speech communication with Sam. The use of automatic speech recognition (ASR) for response logging has been explored in another study from our lab by Araiza-Illan et al. (in press) with the use of Kaldi (Povey et al., 2011), an open-source speech recognition toolkit. The ASR was used to automatically score participant's spoken responses during a speech audiometry test. Their results show the robustness of the ASR when decoding speech from normal-hearing adults, offering a natural alternative for participants to give their responses throughout the test. Therefore, an ASR system, such as Kaldi, could be coupled with Sam, enhancing its social presence and overall interface functionality.

Literature has shown that the gamification of tests can also have beneficial effects on attention and engagement (Moore et al., 2008; Kopelovich et al., 2010; Harding et al., 2023). Although the speech-in-speech perception test has been suggested above to not be appropriate for gamification, it may indeed be interesting to explore how an intentional gamification of the test compares to the data collected here. This applies both to how speech intelligibility may be affected by such a gamification, but also how engagement may differ in comparison to Sam, especially after the implementation of speech recognition and removal of the tablet.

Both the present study and our previous work show the potential use of a NAO humanoid for speech-in-speech perception (present study) and voice manipulation perception (previous work) assessments by taking advantage of the robot's speech-related features. Furthermore, since current technical limitations are expected to be improved in the future, the proposed setup with the NAO provides exciting application possibilities in research and clinical applications.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://doi.org/10.34894/IAGXVF.

Ethics statement

The studies involving humans were approved by METc (Medical Ethical Committee) at UMCG (METc 2018/427, ABR nr NL66549.042.18). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

LM: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Software, Visualization, Writing – original draft, Writing – review and editing. GA-I: Conceptualization, Investigation, Methodology, Project administration, Supervision, Visualization, Writing – review and editing. LR: Conceptualization, Methodology, Supervision, Visualization, Writing – review and editing. EG: Investigation, Resources, Software, Visualization, Writing – review and editing. DB: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – review and editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was supported by the VICI grant 918-17-603 from Netherlands Organization for Scientific Research (NWO) and Netherlands Organization for Health Research and Development (ZonMw). Further support was provided by the WJ Kolff Institute for Biomedical Engineering and Material Sciences, University of Groningen, the Heinsius Houbolt Foundation and the Rosalind Franklin Fellowship from University Medical Center Groningen, University of Groningen.

Acknowledgments

We thank Josephine Marriage and Debi Vickers for sharing the English CRM stimuli, Paolo Toffanin, Iris van Bommel, Evelien Birza, Jacqueline Libert and Jop Luberti for their contribution to the development of the PICKA test battery, as well as Tord Helliesen and Conor Durkin for coding the video footage. The study was conducted in the framework of the LabEx CeLyA (ANR-10-LABX-0060/ANR-11-IDEX-0007) operated by the French ANR and is also part of the research program of the UMCG Otorhinolaryngology Department: Healthy Aging and Communication.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abercrombie, D. (1982). *Elements of general phonetics*. Scotland: Edinburgh University Press.

Akalin, N., Kristoffersson, A., and Loutfi, A. (2019). The influence of feedback type in robot-assisted training. *Mult. Technol. Interact.* 3:67. doi: 10.3390/mti3040067

Alhanbali, S., Dawes, P., Lloyd, S., and Munro, K. J. (2017). Self-reported listening-related effort and fatigue in hearing-impaired adults. *Ear Hear.* 38, e39–e48. doi: 10.1097/AUD.0000000000000361

Amirova, A., Rakhymbayeva, N., Yadollahi, E., Sandygulova, A., and Johal, W. (2021). 10 years of human-NAO interaction research: A scoping review. *Front. Robot. AI* 8:744526. doi: 10.3389/frobt.2021.744526

Araiza-Illan, G., Meyer, L., Truong, K., and Baskent, D. (in press). Automated speech audiometry: Can it work using open-source pre-trained Kaldi-NL automatic speech recognition?

Arnold, K., Gosling, J., and Holmes, D. (2005). *The java programming language*. Boston, MA: Addison-Wesley Professional.

Asfour, T., Regenstein, K., Azad, P., Schroder, J., Bierbaum, A., Vahrenkamp, N., et al. (2006). "ARMAR-III: An integrated humanoid platform for sensory-motor control," in *6th IEEE-RAS international conference on humanoid robots*, Genova, 169–175. doi: 10.1109/ICHR.2006.321380

Bartko, J. J. (1966). The intraclass correlation coefficient as a measure of reliability. *Psychol. Rep.* 19, 3–11. doi: 10.2466/pr0.1966.19.1.3

Bartneck, C. (2020). *Human–robot interaction: An introduction*. Cambridge, MA: Cambridge University Press, doi: 10.1017/9781108676649

Başkent, D., and Gaudrain, E. (2016). Musician advantage for speech-on-speech perception. *J. Acoust. Soc. Am.* 139, EL51–EL56. doi: 10.1121/1.4942628

Başkent, D., Gaudrain, E., Tamati, T., and Wagner, A. (2016). "Perception and psychoacoustics of speech in cochlear implant users," in *Scientific foundations of audiology: perspectives from physics, biology, modeling, and medicine*, eds. A. T. Cacace, E. de Kleine, A. G. Holt, and P. van Dijk (San Diego, CA: Plural Publishing, Inc).

Bess, F. H., Davis, H., Camarata, S., and Hornsby, B. W. Y. (2020). Listening-related fatigue in children with unilateral hearing loss. *Lang. Speech Hear. Serv. Sch.* 51, 84–97. doi: 10.1044/2019_LSHSS-OCHL-19-0017

Boksem, M. A. S., and Tops, M. (2008). Mental fatigue: Costs and benefits. *Brain Res. Rev.* 59, 125–139. doi: 10.1016/j.brainresrev.2008.07.001

Bolia, R. S., Nelson, W. T., Ericson, M. A., and Simpson, B. D. (2000). A speech corpus for multitalker communications research. *J. Acoust. Soc. Am.* 107, 1065–1066. doi: 10.1121/1.428288

Bond, C. F. (1982). Social facilitation: A self-presentational view. *J. Person. Soc. Psychol.* 42, 1042–1050. doi: 10.1037/0022-3514.42.6.1042

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of Sound*. Cambridge, MA: The MIT Press, doi: 10.7551/mitpress/1486.001.0001

Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *J. Acoust. Soc. Am.* 109, 1101–1109. doi: 10.1121/1.1345696

Carhart, R., Tillman, T. W., and Greetis, E. S. (1969). Perceptual masking in multiple sound backgrounds. *J. Acoust. Soc. Am.* 45, 694–703. doi: 10.1121/1.1911445

Cervantes, J.-A., López, S., Cervantes, S., Hernández, A., and Duarte, H. (2023). Social robots and brain–computer interface video games for dealing with attention deficit hyperactivity disorder: A systematic review. *Brain Sci.* 13:1172. doi: 10.3390/brainsci13081172

Choudhury, A., Li, H., Greene, C., and Perumalla, S. (2018). Humanoid robot-application and influence. *Arch. Clin. Biomed. Res.* 2, 198–227. doi: 10.26502/acbr.50170059

Darwin, C. J., Brungart, D. S., and Simpson, B. D. (2003). Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers. *J. Acoust. Soc. Am.* 114, 2913–2922. doi: 10.1121/1.1616924

Dawe, J., Sutherland, C., Barco, A., and Broadbent, E. (2019). Can social robots help children in healthcare contexts? A scoping review. *BMJ Paediatr. Open* 3:e000371. doi: 10.1136/bmjpo-2018-000371

Douissard, J., Hagen, M. E., and Morel, P. (2019). "The da Vinci Surgical System," in *Bariatric robotic surgery: A comprehensive guide*, eds C. E. Domene, K. C. Kim, R. V. Puy, and P. Volpe (Berlin: Springer International Publishing), 13–27. doi: 10.1007/978-3-030-17223-7_3

Drullman, R., and Bronkhorst, A. W. (2004). Speech perception and talker segregation: Effects of level, pitch, and tactile support with multiple simultaneous talkers. *J. Acoust. Soc. Am.* 116, 3090–3098. doi: 10.1121/1.1802535

El Boghdady, N., Gaudrain, E., and Başkent, D. (2019). Does good perception of vocal characteristics relate to better speech-on-speech intelligibility for cochlear implant users? *J. Acoust. Soc. Am.* 145, 417–439. doi: 10.1121/1.5087693

Ellis, L., Reynolds, L., Fucci, D., and Benjamin, B. (1996). Effects of gender on listeners' judgments of speech intelligibility. *Percept. Mot. Skills* 83(3 Pt 1), 771–775. doi: 10.2466/pms.1996.83.3.771

Fitch, W. T., and Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *J. Acoust. Soc. Am.* 106(3 Pt 1), 1511–1522. doi: 10.1121/1.427148

Friard, O., and Gamba, M. (2016). BORIS: A Free, versatile open-source event-logging software for video/audio coding and live observations. *Methods Ecol. Evolut.* 7, 1325–1330. doi: 10.1111/2041-210X.12584

Frid, E., Bresin, R., and Alexanderson, S. (2018). Perception of Mechanical Sounds Inherent to Expressive Gestures of a NAO Robot - Implications for Movement Sonification of Humanoids. *Proceedings of the Sound and Music Computing Conference 2018 (SMC2018)*, Limassol. doi: 10.5281/zenodo.1422499

Fujita, M., Kuroki, Y., Ishida, T., and Doi, T. T. (2003). "A small humanoid robot SDR-4X for entertainment applications," in *Proceedings IEEE/ASME international conference on advanced intelligent mechatronics (AIM 2003)*, Vol. 2, Kobe, 938–943. doi: 10.1109/AIM.2003.1225468

Gallun, F. J., Seitz, A., Eddins, D. A., Molis, M. R., Stavropoulos, T., Jakien, K. M., et al. (2018). "Development and validation of portable automated rapid testing (PART) measures for auditory research," in *Proceedings of meetings on acoustics. Acoustical society of America*, Vol. 33, Boston, MA, 050002. doi: 10.1121/2.0000878

Gaudrain, E., and Başkent, D. (2018). Discrimination of voice pitch and vocal-tract length in cochlear implant users. *Ear Hear.* 39, 226–237. doi: 10.1097/AUD.0000000000000480

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials Quant. Methods Psychol.* 8, 23–34. doi: 10.20982/tqmp.08.1.p023

Hanus, M. D., and Fox, J. (2015). Assessing the effects of gamification in the classroom: A longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance. *Comput. Educ.* 80, 152–161. doi: 10.1016/j.compedu.2014.08.019

Harding, E. E., Gaudrain, E., Hrycyk, I. J., Harris, R. L., Tillmann, B., Maat, B., et al. (2023). Musical emotion categorization with vocoders of varying temporal and spectral content. *Trends Hear.* 27:23312165221141142. doi: 10.1177/23312165221141142

Hartley, D. E., Wright, B. A., Hogan, S. C., and Moore, D. R. (2000). Age-related improvements in auditory backward and simultaneous masking in 6- to 10-year-old children. *J. Speech Lang. Hear. Res.* 43, 1402–1415. doi: 10.1044/jslhr.4306.1402

Hazan, V., Messaoud-Galusi, S., Rosen, S., Nouwens, S., and Shakespeare, B. (2009). Speech perception abilities of adults with dyslexia: Is there any evidence for a true deficit? *J. Speech Langu. Hear. Res.* 52, 1510–1529. doi: 10.1044/1092-4388(2009/08-0220)

Henkemans, O. A. B., Bierman, B. P. B., Janssen, J., Looije, R., Neerincx, M. A., van Dooren, M. M. M., et al. (2017). Design and evaluation of a personal robot playing a self-management education game with children with diabetes type. *Int. J. Hum. Comput. Stud.* 106, 63–76. doi: 10.1016/j.ijhcs.2017.06.001

Henschel, A., Laban, G., and Cross, E. S. (2021). What makes a robot social? A review of social robots from science fiction to a home or hospital near you. *Curr. Robot. Rep.* 2, 9–19. doi: 10.1007/s43154-020-00035-0

Hu, J., Edsinger, A., Lim, Y.-J., Donaldson, N., Solano, M., Solochek, A., et al. (2011). "An advanced medical robotic system augmenting healthcare capabilities – Robotic nursing assistant," in *IEEE International conference on robotics and automation*, Shanghai, 6264–6269. doi: 10.1109/ICRA.2011.5980213

Humble, D., Schweinberger, S. R., Mayer, A., Jesgarzewsky, T. L., Dobel, C., and Zäske, R. (2023). The Jena Voice Learning and memory test (JVLMT): A standardized tool for assessing the ability to learn and recognize voices. *Behav. Res. Methods* 55, 1352–1371. doi: 10.3758/s13428-022-01818-3

Joseph, A., Christian, B., Abiodun, A. A., and Oyawale, F. (2018). A review on humanoid robotics in healthcare. *MATEC Web Confer.* 153:02004. doi: 10.1051/matecconf/201815302004

Kaneko, K., Kaminaga, H., Sakaguchi, T., Kajita, S., Morisawa, M., Kumagai, I., et al. (2019). Humanoid robot HRP-5P: An electrically actuated humanoid robot with high-power and wide-range joints. *IEEE Robot. Automat. Lett.* 4, 1431–1438. doi: 10.1109/LRA.2019.2896465

Keltner, D., and Cordaro, D. T. (2017). "Understanding multimodal emotional expressions: Recent advances in basic emotion theory," in *The science of facial expression*, eds J. A. Russell and J. M. F. Dols (Oxford: Oxford University Press), doi: 10.1093/acprof:oso/9780190613501.003.0004

Kidd, C. D., and Breazeal, C. (2004). "Effect of a robot on user perceptions," in *Conference on intelligent robots and systems (IROS)*, Vol. 4, ed. R. S. J. International Piscataway, NJ: IEEE Publications, 3559–3564. doi: 10.1109/IROS.2004.1389967

Koelewijn, T., Gaudrain, E., Tamati, T., and Baskent, D. (2021). The effects of lexical content, acoustic and linguistic variability, and vocoding on voice cue perception. *J. Acoust. Soc. Am.* 150, 1620–1634. doi: 10.1121/10.0005938

Kont, M., and Alimardani, M. (2020). "Engagement and mind perception within human–robot interaction: A comparison between elderly and young adults," in *Social robotics. Lecture notes in computer science*, Vol. 12483, ed. A. R. Wagner (Berlin: Springer International Publishing), 344–356. doi: 10.1007/978-3-030-62056-1_29

Kontogiorgos, D., Pereira, A., and Gustafson, J. (2021). Grounding behaviours with conversational interfaces: Effects of embodiment and failures. *J. Mult. User Interf.* 15, 239–254. doi: 10.1007/s12193-021-00366-y

Kopelovich, J. C., Eisen, M. D., and Franck, K. H. (2010). Frequency and electrode discrimination in children with cochlear implants. *Hear. Res.* 268, 105–113. doi: 10.1016/j.heares.2010.05.006

Laneau, J., Boets, B., Moonen, M., van Wieringen, A., and Wouters, J. (2005). A flexible auditory research platform using acoustic or electric stimuli for adults and young children. *J. Neurosci. Methods* 142, 131–136. doi: 10.1016/j.jneumeth.2004.08.015

Lee, K. M., Peng, W., Jin, S.-A., and Yan, C. (2006). Can robots manifest personality: An empirical test of personality recognition, social responses, and social presence in human–robot interaction. *J. Commun.* 56, 754–772. doi: 10.1111/j.1460-2466.2006.00318.x

Looije, R., van der Zalm, A., Neerincx, M. A., and Beun, R.-J. (2012). "Help, I Need Some Body the effect of embodiment on playful Learning. In IEEE RO-MAN," in *The 21st IEEE international symposium on robot and human interactive communication, 2012*, Piscataway, NJ: IEEE Publications, 718–724. doi: 10.1109/ROMAN.2012.6343836

Marge, M., Espy-Wilson, C., Ward, N. G., Alwan, A., Artzi, Y., Bansal, M., et al. (2022). Spoken language interaction with robots: Recommendations for future research. *Comput. Speech Lang.* 71:101255. doi: 10.1016/j.csl.2021.101255

Marin-Campos, R., Dalmau, J., Compte, A., and Linares, D. (2021). StimuliApp: Psychophysical tests on mobile devices. *Behav. Res. Methods* 53, 1301–1307. doi: 10.3758/s13428-020-01491-4

MATLAB (2019). *version 9.7.0.1190202, R2019b*. Natick, MA: MathWorks, Incorp.

Mattys, S. L., Brooks, J., and Cooke, M. (2009). Recognizing speech under a processing load: Dissociating energetic from informational factors. *Cogn. Psychol.* 59, 203–243. doi: 10.1016/j.cogpsych.2009.04.001

McGinn, C., Cullinan, M., Holland, D., and Kelly, K. (2014). "Towards the design of a new humanoid robot for domestic applications," in *IEEE international conference on technologies for practical robot applications (TePRA)*, Woburn, MA, 1–6. doi: 10.1109/TePRA.2014.6869155

Messaoud-Galusi, S., Hazan, V., and Rosen, S. (2011). Investigating speech perception in children with dyslexia: Is there evidence of a consistent deficit in individuals? *J. Speech Lang. Hear. Res.* 54, 1682–1701. doi: 10.1044/1092-4388(2011/09-0261)

Meyer, L., Rachman, L., Araiza-Illan, G., Gaudrain, E., and Başkent, D. (2023). Use of a humanoid robot for auditory psychophysical testing. *PLoS One* 18:e0294328. doi: 10.1371/journal.pone.0294328

Moore, D. R., Ferguson, M. A., Halliday, L. F., and Riley, A. (2008). Frequency discrimination in children: Perception. Learning and attention. *Hear. Res.* 238, 147–154. doi: 10.1016/j.heares.2007.11.013

Mühl, C., Sheil, O., Jarutytė, L., and Bestelmeyer, P. E. G. (2018). The Bangor voice matching test: A standardized test for the assessment of voice perception ability. *Behav. Res. Methods* 50, 2184–2192. doi: 10.3758/s13428-017-0985-4

Nagels, L., Gaudrain, E., Vickers, D., Hendriks, P., and Başkent, D. (2020a). Development of voice perception is dissociated across gender cues in school-age children. *Sci. Rep.* 10:5074. doi: 10.1038/s41598-020-61732-6

Nagels, L., Gaudrain, E., Vickers, D., Matos Lopes, M. M., Hendriks, P., and Başkent, D. (2020b). Development of vocal emotion recognition in school-age children: The EmoHI test for hearing-impaired populations. *PeerJ* 8:e8773. doi: 10.7717/peerj.8773

Nagels, L., Gaudrain, E., Vickers, D., Hendriks, P., and Baskent, D. (2021). School-age children benefit from voice gender cue differences for the perception of speech in competing speech. *J. Acoust. Soc. Am.* 149, 3328–3344. doi: 10.1121/10.0004791

Nomura, T., Kanda, T., Suzuki, T., and Kato, K. (2004). "Psychology in human–robot communication: An attempt through investigation of negative attitudes and anxiety toward robots," in *RO-MAN 2004. 13th IEEE international workshop on robot and human interactive communication (IEEE Catalog No.04TH8759)*, Kurashiki, 35–40. doi: 10.1109/ROMAN.2004.1374726

Okuno, H., Nakadai, K., and Kitano, H. (2002). "Social interaction of humanoid robot based on audio-visual tracking," in *Developments in applied artificial intelligence. IEA/AIE 2002. Lecture notes in computer science*, Vol. 2358, eds T. Hendtlass and M. Ali (Berlin: Springer), 140–173. doi: 10.1007/3-540-48035-8_70

Pollack, I. (1975). Auditory informational masking. *J. Acoust. Soc. Am.* 57(Suppl. 1):S5. doi: 10.1121/1.1995329

Polycarpou, P., Andreeva, A., Ioannou, A., and Zaphiris, P. (2016). "Don't read my lips: Assessing listening and speaking skills through play with a humanoid robot," in *HCI international 2016 – Posters' extended abstracts*, ed. C. Stephanidis Berlin: Springer International Publishing, 255–260. doi: 10.1007/978-3-319-40542-1_41

Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., et al. (2011). "The Kaldi speech recognition toolkit," in *In Hilton Waikoloa village IEEE 2011 workshop on automatic speech recognition and understanding*, Piscataway, NJ: IEEE Publications Signal Processing Society.

Rich, C., Ponsler, B., Holroyd, A., and Sidner, C. L. (2010). "Recognizing engagement in human–robot interaction," in *In Proceeding of the 5th ACM/IEEE international conference on human-robot interaction—HRI 2010*, Vol. 375, Osaka: ACM Press, doi: 10.1109/HRI.2010.5453163

Rozin, P., and Cohen, A. B. (2003). High frequency of facial expressions corresponding to confusion, concentration, and worry in an analysis of naturally occurring facial expressions of Americans. *Emotion* 3, 68–75. doi: 10.1037/1528-3542.3.1.68

Saeedvand, S., Jafari, M., Aghdasi, H. S., and Baltes, J. (2019). A comprehensive survey on humanoid robot development. *Knowl. Eng. Rev.* 34:e20. doi: 10.1017/S0269888919000158

Seaborn, K., Miyake, N. P., Pennefather, P., and Otake-Matsuura, M. (2022). Voice in human–agent interaction: A survey. *ACM Comput. Surveys* 54, 1–43. doi: 10.1145/3386867

Semeraro, H. D., Rowan, D., van Besouw, R. M., and Allsopp, A. A. (2017). Development and evaluation of the British English coordinate response measure speech-in-noise test as an occupational hearing assessment tool. *Int. J. Audiol.* 56, 749–758. doi: 10.1080/14992027.2017.1317370

Skuk, V. G., and Schweinberger, S. R. (2014). Influences of fundamental frequency, formant frequencies, aperiodicity, and spectrum level on the perception of voice gender. *J. Speech Lang. Hear. Res.* 57, 285–296. doi: 10.1044/1092-4388(2013/12-0314)

Smith, D. R. R., and Patterson, R. D. (2005). The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and age. *J. Acoust. Soc. Am.* 118, 3177–3186. doi: 10.1121/1.2047107

Smith, M. L., Cesana, M. L., Farran, E. K., Karmiloff-Smith, A., and Ewing, L. (2018). A "spoon full of sugar" helps the medicine go down: How a participant friendly version of a psychophysics task significantly improves task engagement, Performance and data quality in a typical adult sample. *Behav. Res. Methods* 50, 1011–1019. doi: 10.3758/s13428-017-0922-6

Song, H., Barakova, E. I., Markopoulos, P., and Ham, J. (2021). Personalizing HRI in musical instrument practicing: The influence of robot roles (evaluative versus nonevaluative) on the Child's motivation for children in different learning stages. *Front. Robot. AI* 8:699524. doi: 10.3389/frobt.2021.699524

Stroustrup, B. (2000). *The C++ programming language*. London: Pearson Education.

Sulistijono, I. A., Setiaji, O., Salfikar, I., and Kubota, N. (2010). "Fuzzy walking and turning tap movement for humanoid soccer robot EFuRIO," in *International conference on fuzzy systems*, (Barcelona), 1–6. doi: 10.1109/FUZZY.2010.5584423

Ting, C., Yeo, W.-H. Y., King, Y.-J., Chuah, Y.-D., Lee, J.-V., and Khaw, W.-B. (2014). Humanoid robot: A review of the architecture, applications and future trend. *Res. J. Appl. Sci. Eng. Technol.* 7, 1364–1369. doi: 10.19026/rjaset.7.402

Türker, B. B., Buçinca, Z., Erzin, E., Yemez, Y., and Sezgin, M. (2017). "Analysis of Engagement and User Experience with a Laughter Responsive Social Robot," in *Interspeech*, Stockholm: ISCA, 844–848. doi: 10.21437/Interspeech.2017-1395

Uluer, P., Kose, H., Gumuslu, E., and Barkana, D. E. (2023). Experience with an affective robot assistant for children with hearing disabilities. *Int. J. Soc. Robot.* 15, 643–660. doi: 10.1007/s12369-021-00830-5

Van Rossum, G., and Drake, F. L. (2009). *Python 3 reference manual*. Scotts Valley, CA: CreateSpace.

Vestergaard, M. D., Fyson, N. R. C., and Patterson, R. D. (2009). The interaction of vocal characteristics and audibility in the recognition of concurrent Syllablesa. *J. Acoust. Soc. Am.* 125, 1114–1124. doi: 10.1121/1.3050321

Wagenmakers, E.-J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., et al. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications'. *Psychon. Bull. Rev.* 25, 35–57. doi: 10.3758/s13423-017-1343-3

Welch, G. F., Saunders, J., Edwards, S., Palmer, Z., Himonides, E., Knight, J., et al. (2015). Using singing to nurture children's hearing? A pilot study. *Cochlear Implants Int.* 16(Suppl. 3), S63–S70. doi: 10.1179/1467010015Z.000000000276

Zhao, S., Brown, C. A., Holt, L. L., and Dick, F. (2022). Robust and efficient online auditory psychophysics. *Trends Hear.* 26:23312165221118792. doi: 10.1177/23312165221118792

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 21 March 2024
DOI 10.3389/fnins.2024.1321357

Check for updates

OPEN ACCESS

EDITED BY
Karina De Sousa,
University of Pretoria, South Africa

REVIEWED BY
Nihat Yılmaz,
Karabük University, Türkiye
Yuanchia Chu,
Taipei Veterans General Hospital, Taiwan

*CORRESPONDENCE
Milka C. I. Madahana
✉ milka.madahana@wits.ac.za

RECEIVED 13 October 2023
ACCEPTED 06 March 2024
PUBLISHED 21 March 2024

CITATION
Madahana MCI, Ekoru JED, Sebothoma B and Khoza-Shangase K (2024) Development of an artificial intelligence based occupational noise induced hearing loss early warning system for mine workers.
Front. Neurosci. 18:1321357.
doi: 10.3389/fnins.2024.1321357

COPYRIGHT
© 2024 Madahana, Ekoru, Sebothoma and Khoza-Shangase. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Development of an artificial intelligence based occupational noise induced hearing loss early warning system for mine workers

Milka C. I. Madahana[1]*, John E. D. Ekoru[2], Ben Sebothoma[3] and Katijah Khoza-Shangase[3]

[1]School of Mining Engineering, University of the Witwatersrand, Johannesburg, South Africa, [2]School of Electrical and Information Engineering, University of the Witwatersrand, Johannesburg, South Africa, [3]Department of Audiology, University of the Witwatersrand, Johannesburg, South Africa

Introduction: Occupational Noise Induced Hearing Loss (ONIHL) is one of the most prevalent conditions among mine workers globally. This reality is due to mine workers being exposed to noise produced by heavy machinery, rock drilling, blasting, and so on. This condition can be compounded by the fact that mine workers often work in confined workspaces for extended periods of time, where little to no attenuation of noise occurs. The objective of this research work is to present a preliminary study of the development of a hearing loss, early monitoring system for mine workers.

Methodology: The system consists of a smart watch and smart hearing muff equipped with sound sensors which collect noise intensity levels and the frequency of exposure. The collected information is transferred to a database where machine learning algorithms namely the logistic regression, support vector machines, decision tree and Random Forest Classifier are used to classify and cluster it into levels of priority. Feedback is then sent from the database to a mine worker smart watch based on priority level. In cases where the priority level is extreme, indicating high levels of noise, the smart watch vibrates to alert the miner. The developed system was tested in a mock mine environment consisting of a 67 metres tunnel located in the basement of a building whose roof top represents the "surface" of a mine. The mock-mine shape, size of the tunnel, steel-support infrastructure, and ventilation system are analogous to deep hard-rock mine. The wireless channel propagation of the mock-mine is statistically characterized in 2.4–2.5 GHz frequency band. Actual underground mine material was used to build the mock mine to ensure it mimics a real mine as close as possible. The system was tested by 50 participants both male and female ranging from ages of 18 to 60 years.

Results and discussion: Preliminary results of the system show decision tree had the highest accuracy compared to the other algorithms used. It has an average testing accuracy of 91.25% and average training accuracy of 99.79%. The system also showed a good response level in terms of detection of noise input levels of exposure, transmission of the information to the data base and communication of recommendations to the miner. The developed system is still undergoing further refinements and testing prior to being tested in an actual mine.

KEYWORDS

occupational, noise-induced, hearing-loss, artificial intelligence, smart monitoring

Introduction

Occupational noise-induced hearing loss (ONIHL) is a significant concern within the mining industry in South Africa (Khoza-Shangase, 2022), given the documented prevalence of high noise levels (Edwards et al., 2011). This prevalence of ONIHL is attributed to factors such as the nature of mining activities, the confined and reflective work environments, and the use of equipment in mines. These factors significantly increase the risk of exposure to hazardous noise levels, which are the primary cause of hearing problems among mine workers (Matetic, 2005; Strauss et al., 2012). Due to these factors, it has been estimated that one in four mine workers will develop ONIHL. As mine workers proceed to their mid-60's, the incidence increases, with four out of five mine workers presenting with hearing impairment (NIOSH, 2023). To address this issue, South African mines implement hearing conservation programs (HCPs) aimed at protecting workers' hearing health and minimizing the risk of ONIHL. The country has legislation and regulations that mines must adhere to such as the Occupational Health and Safety Act (OHSA) of 1993 (Republic of South Africa, 1995), along with its Noise-Induced Hearing Loss Regulations of 2003, which govern occupational health and safety in South Africa. These regulations set out specific requirements for noise exposure limits, hearing protection, audiometric testing, and the implementation of HCPs. The OHSA sets permissible noise exposure limits (NELs) for different industries and activities, including the mining industry. The regulations specify that the daily personal noise exposure level should not exceed 85 decibels (dB) for an eight-hour work shift.

Legislation and regulations, as part of the hierarchy of controls, also declare that, through risk assessments, employers are required to conduct noise risk assessments to determine the potential for hearing loss and identify areas where noise control measures are necessary. This involves measuring noise levels, evaluating exposure durations, and identifying high-risk areas or job tasks. At the same time, engineering controls measures should be in place to reduce noise levels at their source (NIOSH, 2023). This may involve using quieter machinery and equipment, isolating noisy equipment, or implementing sound insulation measures (Moroe and Khoza-Shangase, 2018a,b).

Additionally, employers are required to implement administrative controls to minimize workers' exposure to excessive noise (Musiba, 2015). These controls may include limiting exposure time, scheduling rest breaks in quieter areas, and implementing job rotation to reduce individual exposure levels.

On the level of the employee and ranked as the last option on the hierarchy of controls, when engineering and administrative controls are insufficient to reduce noise levels to acceptable limits, employers are required to provide suitable hearing protection devices (HPDs) to their employees (NIOSH, 2023), HPDs that are properly selected, maintained, and used in accordance with regulations (Suter, 2002). Furthermore, employees must undergo regular audiometric testing as a crucial component of HCPs (Moroe et al., 2022). Employers are required to provide baseline audiograms for employees exposed to noise levels at or above the action level, followed by periodic audiometric monitoring to detect early signs of hearing loss (Moroe et al., 2022). Additionally, education and training should form part of HCPs where the goal is to raise employees' and supervisors' awareness about the risks of ONIHL (Moroe et al., 2018), proper use of HPDs (Ntlhakana et al., 2015), and the importance of complying with hearing conservation measures. For HCPs to be successful, legislation and regulations dictate that employers must maintain records of noise measurements, risk assessments, audiometric tests, and training provided to employees, and that these records should be readily available for inspection by relevant authorities (Amedofu and Fuente, 2008; Moroe N., 2020). Compliance with and enforcement of these regulations and legislation is the responsibility of the South African Department of Employment and Labour, which is responsible for enforcing occupational health and safety regulations, including those related to ONIHL.

Key points in South African legislation regulations, which comprehensively cover the hierarchy of controls including noise level limits, hearing conservation programs, engineering controls, education and training, monitoring and reporting are similar to those meant to be adhered to globally including in the Americas (Latin America, Canada, and the United States) and the rest of Africa (Arenas and Suter, 2014; Moroe et al., 2018). The main difference is the application and implementation of these, for example what each country's defined values for permissible exposure limit (PEL) is, and if and how legislation enforcement occurs (Moroe et al., 2018). Where some countries ensure effective enforcement of regulations through inspections, penalties for non-compliance, and incentives for compliance; other countries do not (Moroe et al., 2018).

While HCPs in South Africa aim to address ONIHL, several challenges exist in their implementation (Moroe et al., 2018; Khoza-Shangase et al., 2020). Some documented common challenges associated with these programs include; lack of awareness and education among both employers and employees regarding the risks of ONIHL and the importance of hearing conservation measures (Moroe et al., 2018; Kanji et al., 2019). Inadequate and insufficient training and supervision regarding the implementation of HCPs, where employees and supervisors receive no or limited training on identifying noise hazards, selecting appropriate hearing protection, and conducting regular audiometric testing (Moroe and Khoza-Shangase, 2018a,b). Limited resources, including capacity versus demand challenges around audiologists in the country, leading to inadequate noise control measures, insufficient provision of HPDs, and limited access to audiometric testing facilities (Moroe et al., 2018; Pillay et al., 2020). Compliance issues around hearing conservation regulations where employers struggle to meet the requirements for noise measurements, risk assessments, audiometric testing, and recordkeeping, mostly due to some employers not prioritizing hearing conservation or attempting to cut costs by disregarding regulations (Khoza-Shangase, 2022). Linguistic, cultural and behavioral factors where, for example, attitudes towards wearing HPDs pose challenges; and the language used for training and education is incongruent with the employees (Moroe N., 2020). Effective enforcement and monitoring can be a challenge, influenced by insufficient resources and limited inspections by regulatory authorities resulting in inadequate enforcement of regulations and insufficient follow-up on non-compliant mines (Khoza-Shangase, 2022); and cumulative noise exposure and burden of disease (HIV/AIDS and TB) where some employees are exposed to high noise levels from multiple sources, both in their occupational and non-occupational environments, and suffer concurrent toxins exposure where they are on ototoxic treatments for HIV/AIDS and TB (Khoza-Shangase, 2022), thus increasing their risk of ONIHL and making it more challenging to control and mitigate the effects solely through workplace HCPs that

do not take these factors into account. Addressing these challenges requires a multi-faceted approach that can be supported by the use of Internet of Things (IoT)-based hearing loss early monitoring systems as part of HCPs (Mardonova and Choi, 2018).

The main objective of this research work is to present a preliminary development of an AI based early monitoring system that integrates smart hearing protection with smart mine wearable watches. The developed system can provide significant benefits for mine workers as a form ONIHL early warning system. This system combines the capabilities of IoT devices, such as sensors and wearables, to monitor noise exposure levels and facilitate real-time monitoring and protection of the workers' hearing when exposed to hazardous noise levels. By integrating IoT technology, smart hearing protection devices, and wearable watches, the current researchers aim to have a system that enables real-time monitoring, personalized protection, and early detection of hearing loss risks for mine workers. This system aims at enhancing worker safety, promoting proactive hearing health management, and contributing to a culture of prevention in the mining industry.

This early warning system includes numerous factors, for it to be efficient and successful, with positive outcomes for any HCP. Firstly, there has to be IoT sensors for noise monitoring that get strategically deployed in the mining environment to measure and monitor noise levels. These sensors can be placed in key areas or attached to equipment to capture accurate and real-time noise data. In the current study, these sensors are part of smart hearing muffs that transmit the data to a central monitoring system for analysis. The miners wear smart hearing protection devices (SHPDs) with smart watches which are also equipped with sensors. These SHPDs can have built-in noise sensors and connectivity capabilities to communicate with the central monitoring system. The SHPDs can adjust noise attenuation levels based on real-time noise exposure and provide workers with audio cues and alerts. The mineworkers are provided with wearable watches that act as a central hub for integrating various IoT devices and functionalities. These watches can connect to the SHPDs, IoT sensors, and other wearables, consolidating data and enabling real-time monitoring, communication, and alerts. Secondly, the early warning system must have real-time monitoring and feedback capabilities, where the IoT-based system continuously collects noise data from the sensors and SHPDs, transmitting it to the wearable watches. Mine workers can access real-time noise exposure information, receive alerts when noise levels exceed safe thresholds, and obtain feedback on their personal noise exposure. Thirdly, the system can allow for data analysis and insights development, where the collected data is analysed by the central monitoring system to identify patterns, trends, and potential risks. Machine learning algorithms can be employed to recognize patterns of noise exposure and detect early signs of hearing loss. The system can generate personalized reports and insights for individual workers and mine management. This can be done because the system has alert mechanisms, where if the system detects excessive noise levels or potential risks of hearing loss, it triggers alerts through the wearable watches. These alerts can be visual or auditory or vibrotactile, ensuring that mine workers are immediately aware of the hazards and can take necessary actions, such as adjusting their work practices or seeking quieter areas. Lastly, the system is set up to integrate with existing mine management systems and databases, allowing for seamless data sharing and accessibility. This integration facilitates comprehensive analysis, reporting, and decision-making processes related to hearing health and safety in the mining environment. Such an early warning system requires provision of comprehensive training to mine workers on using the IoT-based system, including the proper use of SHPDs and wearable watches. This includes conduction of awareness programs to educate workers about the importance of hearing protection and the benefits of the early monitoring system. These training sessions, and refresher courses, need to be conducted regularly to ensure effective and ongoing usage of the system. Ensuring that privacy and security considerations have been addressed is important as well. Implementation of robust privacy and security measures to protect worker data collected by the IoT devices, with compliance with data protection regulations (POPIA), secure data transmission protocols, and clear communication on data usage and privacy policies are essential to build trust among mine workers where such an IoT-based system is being used as part of HCPs (Ntlhakana et al., 2022). Another important consideration is the making sure that a robust maintenance plan to address issues related to IoT devices, wearables, and sensors is in place. Regular updates, calibration, and technical support are crucial to maintaining the reliability and accuracy of the system.

Once the AI-based ONIHL early warning system is in place, it can bring numerous value and benefits to both workers and the mining industry. This value and advantages include the following: Improved worker safety, early detection and intervention, personalised risk assessment, increased awareness and education, cost reduction, regulatory compliance, long-term data analysis, continuous monitoring, enhanced Occupational Health Programs (OHPs), and technological advancement and innovation. These benefits can be crucial in the context of mines, where noise is excessive.

Background

Mining employees exposed to high noise levels often experience difficulty hearing high-frequency sounds initially (Edwards et al., 2010; Grobler et al., 2020). Regardless of age or gender, the measurement of hearing loss is typically assessed through percentage loss of hearing (PLH) and standard threshold shifts (STS) (Department of Labour, 2001; Department of Mineral Resources, 2016). PLH is determined by calculating the decline in hearing thresholds at specific frequencies (0.5, 1, 2, 3, and 4 kHz), and a baseline audiogram is established using this data (Department of Labour, 2001). South African hearing conservation practitioners employed this method for defining hearing loss for compensation purposes between 2001 and 2016 (Department of Labour, 2001). The STS method, based on the International Organization for Standardization (ISO) standard ISO1999:2013, considers an 8 dB decline as indicative of early ONIHL. Since 2016, South African mines have utilized the STS method to assess miners' hearing, tracking STS deterioration as a precursor to hearing loss (Strauss et al., 2012; Grobler et al., 2020). In 2008, the Department of Mineral Resources and Energy (DMRE) established NIHL milestones for the mining industry, aiming to prevent hearing deterioration beyond 10 % in occupationally exposed individuals after December 2008 (Department of Minerals and Energy, 2008; Msiza, 2014). Despite efforts, hearing loss prevention was not entirely successful (Edwards and Kritzinger, 2012; Moroe and Khoza-Shangase, 2018a,b), leading to revised milestones in 2014, where no employee's STS should exceed 10 dBHL from the baseline

when averaged at 2000, 3000, and 4,000 Hz in one or both ears by December 2016 (MHSC, 2015; Moroe N. F., 2020). Therefore, STS became a prioritized metric for measuring hearing loss in miners.

Normal hearing is denoted as 0 dBHL (Chamber of Mines, 2016), and a STS is defined as an average shift in hearing threshold of 10 dBHL. While no hearing loss occurs at this stage, any shift greater than 10 dBHL should be reported, triggering further investigation and intervention (Chamber of Mines, 2016). A shift exceeding 25 dBHL for one or both ears indicates actual hearing loss, requiring diagnostic audiometry confirmation (Department of Mineral Resources, 2016). The use of STS to describe the hearing function of workers exposed to excessive noise levels has been a global practice since the early 2000s (Heyer et al., 2011; Masterson et al., 2015). The hearing loss prevention efforts in the South African mining industry, aligned with the NIHL 2016 milestones, now mirror those of developed countries like the United States. However, the efficacy of these interventions will only be assessed in 2024 (MHSC, 2015).

Developing an Artificial Intelligence (AI) based Occupational Noise Induced Hearing Loss (ONIHL) early warning system for mine workers can be a valuable initiative to safeguard their hearing health. Such a system can help identify potential risks and provide timely alerts to prevent or mitigate the harmful effects of noise exposure. The development process for such a system requires numerous steps depicted in Figure 1.

At present, various wired (Dohare et al., 2015; Ikeda et al., 2021; Kolade et al., 2021; Kolade and Cheng, 2021) and wireless communication technologies are available that meet the minimum mandatory criteria for the data broadcast speed and range to support remote mining operations and advanced monitoring systems. The data transmission diagram by Ikeda et al. (2021) and Figure 2.

The internet and WIFI technologies that are currently implemented in the mines ensure the efficient transmission and transfer of information. The transmission diagram in Figure 2 demonstrates how the smart technology is integrated with hearing protection.

Materials and methods

Development of the AI based early warning system.

Procedure

A smart system that continuously monitors noise levels in mine environments was developed. This system is made up of noise attenuation headphones, server-based AI algorithms and a smart watch. The headphones are equipped with sound sensors, and they collect information about the sound (sound intensity levels and the frequency) an individual mining employee is being exposed to. The dataset that contains the sound level of exposure for each mine worker is transmitted from the headphones to storage in a database. The collected data (sound intensity levels and frequency) is then fed into the trained AI model on the server.

To develop, train and test the AI subsystem of the smart system, a comprehensive dataset with various features is collected from various environments in a platinum mine. The features of interest in the collected data were noise level measurements, duration of exposure, corresponding audiometric test results, age, and gender. The data was cleaned and relevant features that can be used by the AI model such

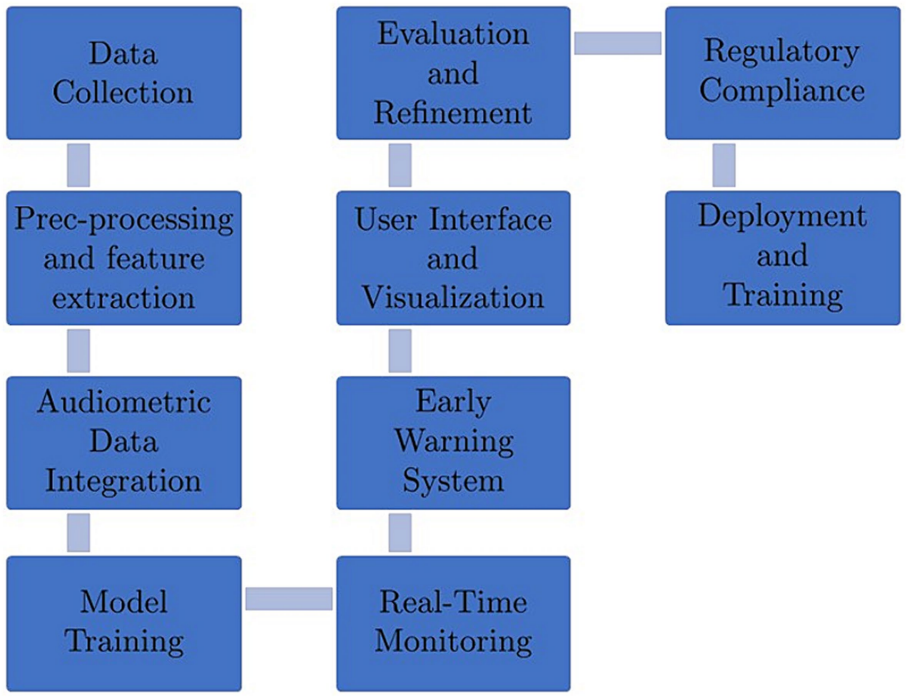


FIGURE 1
Development process for an AI-based ONIHL early warning system.

FIGURE 2
The data transmission diagram (Ikeda et al., 2021).

as duration of exposure, sound intensity of exposure and frequency, were extracted. The audiometric data were combined with the noise exposure data to establish the relationship between noise levels and hearing loss progression. This step was essential in training the AI models to recognize patterns and detect early signs of ONIHL. Machine learning techniques, random forest, support vector machines and logistic regression were utilized to train the AI model. For the AI subsystem, the target feature was the threshold shift of a miner worker defined as the average change in hearing of 10 decibels or more at speech frequencies (2,000–4,000 Hz) in both or one ear in comparison to the mine workers baseline audiogram. A 10-fold cross validation was ran with a split of 80 and 20% randomly shuffled training and testing set, respectively. The K means is used to cluster the mine workers and then using the threshold shift, the mine workers are classified the mine worker according to level of priority. A predicted threshold shift of less than 40 is viewed as low priority, between 40 and 60 is moderate priority, a threshold shift between 60 and 90 has a high priority and a threshold shift of greater than 90 has extreme priority. The various levels of priority are linked to various recommendations messages which are communicated to the mining employee via the smart watch. The low priority does not receive any messages while moderate priority receives a message to remind the mine workers to continue wearing their hearing protection correctly. The high and extreme priority receive a warning message and in addition to that a vibrotactile signal is triggered on the smart watch.

Demographics and inclusion criteria

The initial training of the AI model required data. The data set used was obtained from a platinum mine in South Africa. The demographics of the dataset used to train the AI model is as follows: A total of 12,596 mine workers are in the platinum mine where this study was conducted. 11% of this mining population is female and 89% male. The age distribution indicates appropriate variation with 6,800 workers being younger than 40 years, 4,800 between ages 41 and 55, and 996 being between the ages of 55 and 65 years of age. The designed system targets occupations that are normally exposed to occupational noise for extended periods of time. Therefore, a dataset for 1,350 employees consisting of both male and female mining employees with ages ranging from 18 to 60 years old was used to train the AI model. This sample size for training the AI model was deemed adequate for reliable results as a good sample size is usually approximately 10% of the population, if this does not exceed 1,000 (Carmen and Betsy, 2007).

General description of the subsystems of the developed early warning system

Figures 3, 4 shows the block diagram and the pictorial representation of the Occupational Noise-Induced Hearing Loss and early warning system.

The mining employee is exposed to the level of noise the machine produces. A smart watch that capitalizes on the availability of WIFI and sensors in the mining environment is used to communicate with the hearing protection to provide mine workers with information about their surrounding and to enable communication. With the smart watch, the location of the individual mine worker can be established in real time by the mining administrators on the surface of the mine. This facilitates the ability of the administrators to check the conditions of the location, for example, the level of noise in the location where the mine worker is currently located. Personalised warning or recommendation messages can be sent to the mine worker. The important features the smart watch has for monitoring of the mine worker's state of hearing are: Mine worker's real-time location tracking, incident reporting and feedback-based communication. The mine workers can use the smart watch to report on incidences related

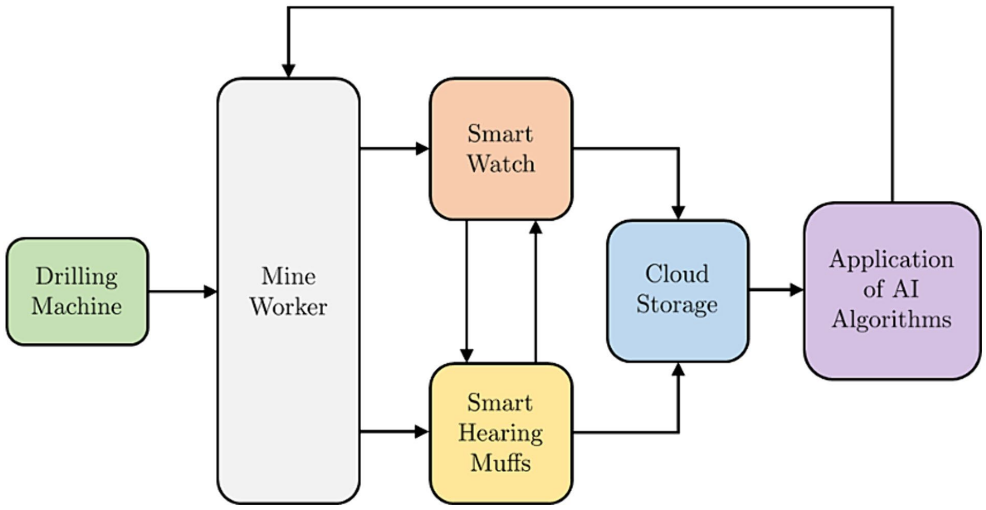


FIGURE 3
The block diagram of the ONIHL early warning and monitoring system.

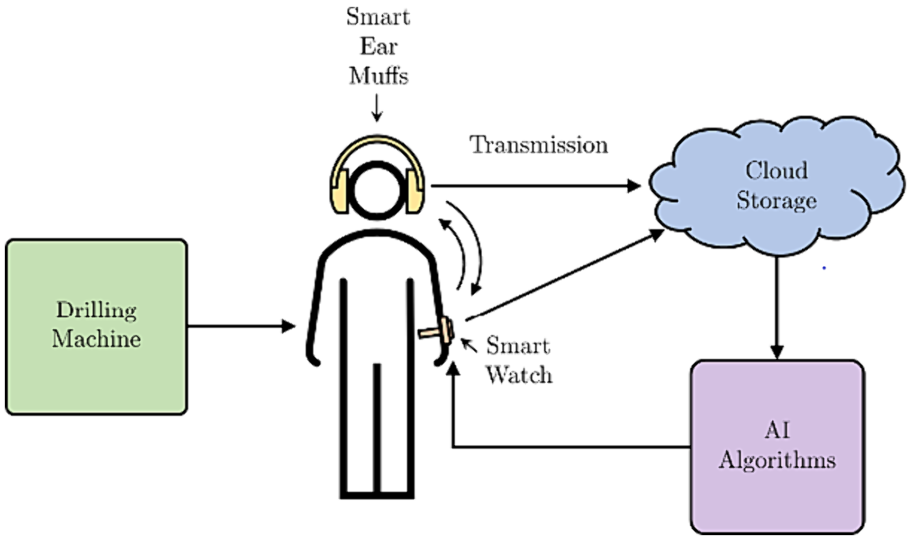


FIGURE 4
A pictorial representation ONIHL early warning and monitoring system.

to excessive noise that could be coming from faulty equipment or malfunctioning hearing protection. The integration of the system with Artificial intelligence permits for the real time automated early warning and recommendations alerts to be sent to the mine workers. The sound sensors set up within the mines provide essential information on areas within the mines with excessive noise that could be due to equipment failure. The mine worker can receive immediate alerts through the vibration of the smart watch. The smart watch applies IoT and can communicate with the hearing protection via Bluetooth technology. Several authors have made use of the ESP-32 as the basis for their smart-watch design (Volsa et al., 2022; Himi et al., 2023; Joseph, 2023; Puckett and Emil, 2023). The smart hearing protection can monitor real time noise levels using the sensors installed on it and cloud technology. The current levels of exposure to sound, which includes the intensity, and the frequency of exposure is collected by the hearing protection and transmitted to the cloud storage via a mobile app.

The mining administrators can conduct real time monitoring of the sound levels and the frequencies using the datasets. With this integrated system, the employees can be informed of their current levels of exposure at any time. With the integrated system, the mine worker can be informed whether the hearing protection is worn correctly or not. The mine worker can also be provided with warnings in the form of vibration of the smart watch which is integrated with the smart hearing device or other visual systems that can also be integrated into this system. The information collected from the smart watch and the hearing protection is stored in a robust data storage solution. To ensure control over the data and the application of the AI models, cloud storage is chosen. This type of storage ensures that the data can be recovered in case there is a problem on site in the

mines. The cloud storage also requires little expertise for implementation with a few resources.

The AI subsystem is used in estimation of the mine worker's threshold shift. The degree of priority is classified with the change in threshold shift. There are four classes of priority of threshold shift. These classes are low, moderate, high and extreme. The mining employee's threshold shift is categorized and depending on the category recommendations are provided. The AI subsystem is also used to process the sound intensity patterns at particular frequencies and to provide the mine workers with recommendations of the actions they should take if necessary. Further details on the AI subsystems have already been published by the authors in previous works (Madahana et al., 2019a,b, 2020). The feedback loops allow for a two-way communication between the mine administrators on the surface and the mine employees. The feedback loops are from the mine administrator to the smart watch and from the smart watch to the hearing protection. These two systems can also be decoupled and in case the smart watch is not functioning, then the hearing muffs can still be used and in this case, the mine worker will depend on other visual warning systems in the mines that have been integrated with the system as a supplement.

Implementation of the laboratory test rig

The laboratory test rig was built to test the proposed system. It consists of a smart watch, smart headsets, computer cluster, cloud storage, hydraulic shaping machine, Variable Direct Current (VDC) machine. The hydraulic shaping machine emits noise between 90 to 110 decibels depending on the various settings and activities. The Variable direct current emits a noise of 91.3–100.7 decibels. The system is connected as shown in Figures 3, 4, the drilling machine is replaced with the hydraulic shaping Machine and the VDC machine. The variable direct current machine shown in Figure 5C. The participants have their hearing checked in the psychometric booth shown in Figure 5D to ensure that their state of hearing health is known. The participant wears both the smart watch and the smart headset. The information obtained from the smart watch and the headset is transmitted via WIFI to a cloud storage. The information is then extracted from the cloud storage, AI models are applied to process the information and the appropriate recommendation is sent to the participant. The Participant tests the systems by moving 60 m away and thereafter, the distance is reduced until the participant is 0.5 m away from the machine. Different recommendation messages are sent to the participant smart watch at the various distance. The sound level metres in Figure 6 are used to measure the sound intensity of the machine. Figure 5A shows the shearing machine, which is simulated using hydraulic shaping machine, shown in Figure 5B. The preliminary integrated prototype is tested by allowing user to wear both the headphone and the smart watch, exposing them to noise at various decibels and frequency and observing the recommendations messages that are sent to the smart watch. One of the significant roles that audiologist plays during the testing of the integrated system is verification and validation of the system. The South African Mines usually have an audiologist who designs the Hearing Conservation Program for the mine. It is therefore imperative that audiologists be involved in the research, designing, testing and implementation of any system that would assist in minimizing the risks of ONIHL in the mines. The audiologist provides valuable feedback on the suitability of the integrated system and whether the user is wearing the hearing protection correctly.

The entire testing is conducted in the presence of audiologists to ensure that the participants are not exposed to any occupational noise and that the smart muffs provide sufficient attenuation.

Figure 7A shows the overall systems diagram of the Smart watch. The functioning of the smart watch is centred around the ESP32-WROOM-32 development board provides computational power as well as wireless internet and Bluetooth connectivity. The user inputs are switches which allow the mine worker to switch the watch on and off as well as toggle between various functions. The watch is powered by a Li-po battery, and a power management system is used to control the charging and discharging of the battery. Various sensor inputs are available to provide functionality during surface mining activities: (1) The ambient light sensor automatically adjusts the brightness of the screen and saves battery life. (2) The magnetometer is to be used for direction (compass). (3) Heart rate and Oxygen Saturation sensing for cardiovascular health. (4) The accelerometer for motion detection. During sub-surface mining, the magnetometer may be affected by the underground environment. The outputs of the smart watch are: (1) The Watch display which can be used to read time as well as notifications related to sound level and warnings related to NIHL due to environmental conditions. (2) The haptic vibration motor will vibrate during notifications as sent to the watch as well as when the mine worker is not wearing the hearing protection. (3) A micro-SD card is also available to log data.

Figure 7B shows the overall systems diagram of the Smart Earmuffs. Traditional earmuffs are equipped with additional sound sensors that can collect information from the environment. Similar to the smart watch, the ESP32-WROOM-32 development board provides computational power, wireless internet connection and Bluetooth connectivity. The user inputs are switches which allow the mine worker to switch the earmuffs on and off. The earmuffs are powered by a Li-po battery and a power management system is used to control the charging and discharging of the battery. The various sensor inputs are available to provide functionality during both surface and sub-surface mining activities: (1) The Microphones are used to pick up ambient sound to be used for a noise level meter that measures ambient sound in decibels. The noise level meter measurements are used to provide the mine worker with an instantaneous warning should the sound level reach dangerous limits. These warning messages appear on the smart watch accompanied by vibrations from the haptic vibration motor. In addition, the readings are sent wirelessly by the ESP32 for further processing in the cloud. (2) The capacitative sensors are used to detect whether the mineworker is correctly wearing the hearing protection. The messages are sent to the cloud and can be seen by administrators and warning messages are sent to the smart watch accompanied by vibration from the haptic vibration motor.

Test environment

Performing experiments in real underground environments is a rigorous process that requires permission from the mining stake holders. In addition to that, it can hinder normal operations from occurring while exposing researchers to unnecessary risks (Hussain

FIGURE 5
(A) The hydraulic shaping machines. **(B)** Shearing machine. **(C)** Variable direct current machines. **(D)** Psychometric booth.

FIGURE 6
Sound level meter.

et al., 2017). For rapid and repetitive testing of the developed prototype the Wits mock mine built under the Chamber of Mines building at the University of the Witwatersrand, Johannesburg (South Africa) was used. Some of the aspects of a mine that have previously been tested in this mock mine are and not limited to mine safety, tunnel economics, improved ventilation, energy savings and communication within a mine (Hussain et al., 2017). Comparable to actual mine, is made up of three sections: an arc shaped tunnel in the basement of the building, a stope panel and a vertical shaft. The tunnel is closed on one side and open on the other side. The roof of the mock mine represents the surface of the mine, and it is shallow from the open end and gets deeper towards the closed end. The mock-mine shape, size of the tunnel, steel-support infrastructure, and ventilation system are analogous to deep hard-rock mines. Actual underground mine material has been used to build the mock mine to ensure it mimics a real mine as close as possible. The mock mine is equipped with a weather station, asset management system, seismometer, crack meter, stress meter, asset management and video analytics system. The wireless channel propagation of the mock-mine is statistically characterized in 2.4–2.5 GHz frequency band (Zaman et al., 2018). Data from various systems is collected and

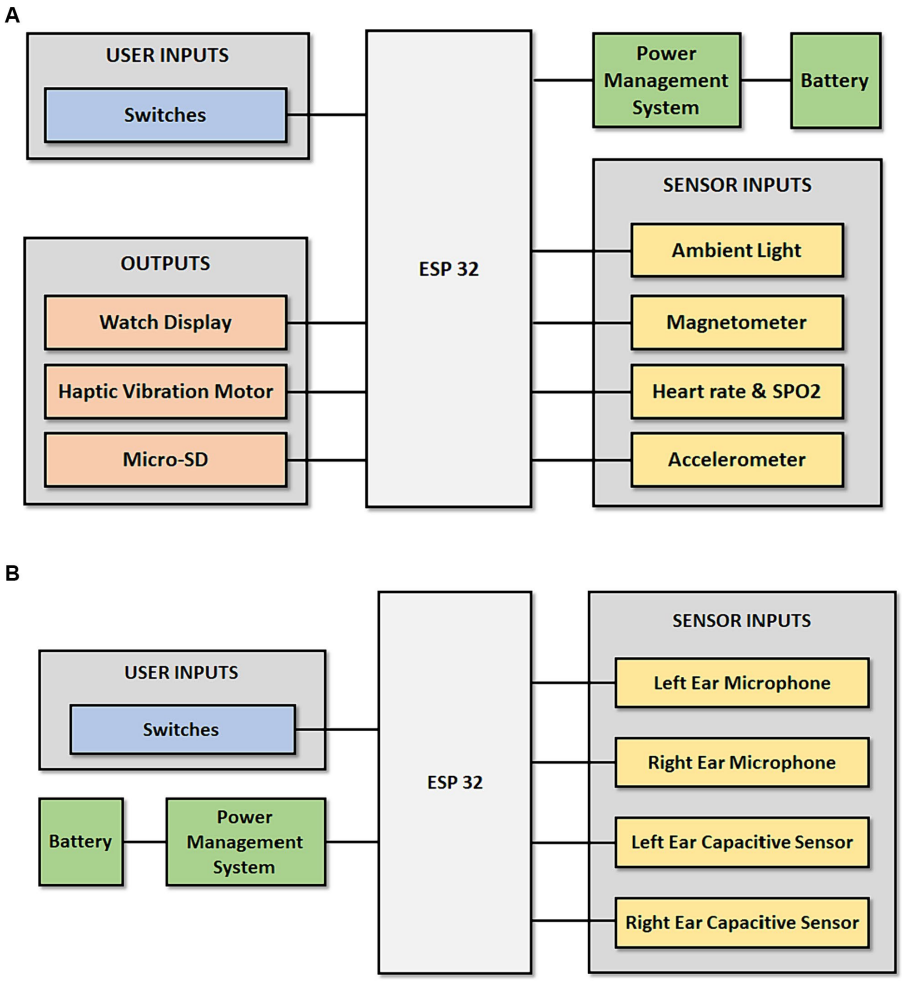


FIGURE 7
(A) Overall systems diagram indicating the subsystems in the ESP32 Smart watch. **(B)** Overall systems diagram indicating the subsystems in the ESP32 Smart Earmuffs.

FIGURE 8
Wits mock mine tunnel.

displayed in the control room adjacent to the mock mine. The mock mine has an intelligent lamp room that prevents miners with inoperative equipment to enter the mine. It also has a rescue chamber, where all the miners can gather in case of a disaster. The tunnel for the mock mine is shown in Figure 8. The sample size used for the first stage of testing was 50 individuals made up of both male and female participants, whose ages ranged from 18 years to 60 years. The experimental protocols employed were approved by the ethics committee (The University of the Witwatersrand) and concur with the Helsinki Declaration. The study falls within the greater scope of another study titled "Feedback based estimation of Noise Induced Hearing Loss in the mines" and it has an ethical clearance number W-CBP-180305-01.

Results and discussion

Results

Table 1 shows the performance of the machine learning algorithm used. The Random forest classifier outperformed the other algorithms.

Table 2 shows the testing of the integrated system.

Discussion

In the context of the ONIHL, early warning and monitoring system for the mining industry, proactive and predictive approaches hold significant importance. By taking a proactive stance, the system identifies

potential risks before they escalate, thus allowing for timely risk identification. It allows for the early detection of elevated noise levels and emerging patterns that could lead to hearing loss among mine workers. Early identification enables the implementation of preventive measures. These measures could include timely interventions, adjustments in work practices, or the use of enhanced PPE to minimize the risk of occupational noise-induced hearing loss. The predictive approach and the system's predictive capabilities, driven by machine learning algorithms, allow for continuous monitoring of noise levels and associated factors. This ensures that any changes or trends in the working environment are promptly detected. Machine learning algorithms used in the developed system are trained to recognize patterns in the data. This includes identifying specific combinations of noise intensity, duration of exposure, and other variables that correlate with an increased risk of hearing loss. Predictive analytics help in forecasting potential issues based on these patterns. The predictive nature of the system enhances the alert mechanism. Instead of responding solely to current conditions, the system can anticipate future risks based on historical data, providing a more optimized and proactive alert system. Predictive analytics assist in the efficient allocation of resources. By forecasting when and where increased noise exposure is likely to occur, mine operators can deploy interventions strategically, focusing resources where they are most needed.

Combining proactive and predictive approaches allows for the development of comprehensive risk mitigation strategies. This involves not only addressing immediate concerns but also planning for long-term measures to reduce the overall risk of ONIHL in the mining environment. The goal is to enhance worker safety. Proactive measures prevent potential risks, while predictive analytics contribute to a more sophisticated and responsive safety infrastructure. This, in turn, minimizes the likelihood of ONIHL incidents. Being proactive in identifying and addressing risks ensures that the developed system aligns with regulatory standards. This is crucial for maintaining compliance with occupational health and safety guidelines specific to noise exposure in mining operations. By adopting proactive and predictive approaches, the ONIHL early warning and monitoring system aims for a lasting impact. Continuous evaluation, refinement, and adherence to safety protocols contribute to sustained worker well-being over the long term.

An alert mechanism that triggers warnings when the AI model detects excessive noise levels or predicts an increased risk of ONIHL for mine workers was implemented. These alerts can be sent to the workers, supervisors, or safety officers through visual or auditory means. Integrated smart hearing protection and wearable mining watches can contribute to hearing loss prevention as they could be categorized as PPE and administration in the hierarchy of controls.

TABLE 1 Performance of the machine learning algorithms.

Model	Average training accuracy	Average testing accuracy
Logistic regression	74.56	77.25
Support vector machines	86.00	99.12
Decision tree	92.25	99.89
Random forest classifier	91.88	99.58

TABLE 2 Testing of the integrated system.

Distance of participant from machine (meters)	Priority level	Recommendation	Observation
Machine off	Low priority	None	No recommendations messages were received
60	low	None	No recommendations messages were received
50	low	None	No recommendations messages were received
40	low	Please wear your hearing protection	Successful SMS
30	Moderate	Hearing protection should be worn correctly	Successful SMS
20	Moderate	Hearing protection should be worn correctly	Successful SMS
10	Moderate	Hearing protection should be worn correctly	Successful SMS
5	High	Hearing protection should be worn correctly or step out of the section	Successful SMS
2	High	Hearing protection should be worn correctly or step out of the section	Successful
1	High	Hearing protection should be worn correctly or step out of the section	Successful
0.5	Extreme	• Hearing protection should be worn correctly or step out of the section • Vibration.	Successful SMS and vibration

These form part of preventative audiology efforts where the focus is on *preventive care* rather than *compensatory care*. This preventive goal is achieved when the smart watch and the hearing protection work collaboratively to ensure preservation of hearing among mine workers by sending alerts and recommendation messages regarding the work context as well as the miner's state of hearing.

Recommendations and conclusions

In summary, the importance of proactive and predictive approaches, as in the proposed, lies in their ability to prevent, identify, and address risks systematically, fostering a safer and healthier working environment for mine workers. The ONIHL early warning and monitoring system employs a holistic approach, integrating advanced technologies, machine learning, and real-time monitoring to effectively address the risk of ONIHL in the mining industry. Bearing the above in mind, several steps remain to be completed in future. Firstly, evaluation and refinement of the system still needs to be done. The performance of the AI model and the effectiveness of the early warning system need to be continuously evaluated. In this process, feedback from mine workers and stakeholders will be collected to identify areas for improvement and refine the system accordingly. Secondly, regulatory compliance needs to be ensured. The developed system requires the researchers to ensure that it aligns with relevant safety regulations and standards for noise exposure in mining operations. Compliance with occupational health and safety guidelines is crucial to ensure the well-being of workers. Lastly, deployment and training of the system is yet to be performed. The system still needs to be deployed in mine sites and adequate training be provided to workers and supervisors on how to interpret and respond to the warnings. Regular training sessions and awareness programs can help promote a safety-conscious culture. The development design and preliminary implementation of a test prototype for a ONIHL early warning and monitoring system has been presented. This system will play a fundamental role in ensuring that the risks of ONIHL in the South African mines is minimized or mitigated. For this system to be work efficiently, mine workers will have to be trained on the correct ways to wear the smart watches and the hearing protection.

Data availability statement

The original contributions presented in the study are included in the article/supplementary materials, further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving humans were approved by University of the Witwatersrand ethics committee. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

Author contributions

MM: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Software, Validation, Writing – original draft. JE: Data curation, Investigation, Methodology, Software, Validation, Writing – review & editing. BS: Formal analysis, Investigation, Validation, Writing – review & editing. KK-S: Conceptualization, Formal analysis, Investigation, Methodology, Supervision, Validation, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The authors would like to thank School of Mining Engineering, Faculty of Engineering, University of the Witwatersrand, Johannesburg, South Africa for providing financial assistance for the publication of this manuscript.

Acknowledgments

We would like to acknowledge the School of Mining, future Tech Laboratory for allowing their facilities to be used. We would also like to acknowledge the Department of Speech Pathology and Audiology for providing some of the required equipment.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Amedofu, G., and Fuente, A. (2008). "Occupational hearing loss in developing countries" in *Audiology in Developing Countries*. eds. B. McPherson and R. Brouillette (New York: Nova Science Publishers), 189–221.

Arenas, J. P., and Suter, A. H. (2014). Comparison of occupational noise legislation in the Americas: an overview and analysis. *Noise Health* 16, 306–319. doi: 10.4103/1463-1741.140511

Carmen, V., and Betsy, M. (2007). Understanding power and rules of thumb for determining sample size. *Tutor. Quant. Methods Psychol.* 3, 43–50. doi: 10.20982/tqmp.03.2.p043

Chamber of Mines. (2016). *Noise Team on the Mine Health and Safety Milestones*. Johannesburg: Chamber of Mines.

Department of Labour. (2001). *Circular Instruction No. 171 – The Determination of Permanent Disablement Resulting From Hearing Loss Caused by Exposure to Excessive Noise and Trauma*. Pretoria: Department of Labour.

Department of Mineral Resources. (2016). *Guidance Note for the Implementation of Standard Threshold Shift in the Medical Surveillance of Noise Induced Hearing Loss*. Pretoria: Department of Mineral Resources

Department of Minerals and Energy. (2008). *Presidential Mine Health and Safety Audit*. Pretoria: Department of Minerals and Energy.

Dohare, Y. S., Maity, T., Das, P. S., and Paul, P. S. (2015). Wireless communication and environment monitoring in underground coal mines–review. *IETE Tech. Rev.* 32, 140–150. doi: 10.1080/02564602.2014.995142

Edwards, A. L., Dekker, J. J., Franz, R. M., Van Dyk, T., and Banyini, A. (2011). Profile of noise exposure levels in South African mining. *J. South. Afr. Inst. Min. Metall.* 111, 315–322. Available at: http://www.scielo.org.za/scielo.php?script=sci_arttext&pid=S2225-62532011001700003&lng=en&tlng=en.

Edwards, A., and Kritzinger, D. (2012). Noise-induced hearing loss milestones: past and future. *J. South. Afr. Inst. Min. Metall.* 112, 865–869.

Edwards, A., van Coller, P., and Badenhorst, C. (2010). Early identification of noise induced hearing loss: a pilot study on the use of distortion product otoacoustic emissions as an adjunct to screening audiometry in the mining industry. *Occupat. Health South. Africa* 6, 28–35.

Grobler, L. M., Swanepoel, D. W., Strauss, S., Becker, P., and Eloff, P. B. Z. (2020). Occupational noise and age: a longitudinal study of hearing sensitivity as a function of noise exposure and age in South African gold mine workers. *S. Afr. J. Commun. Disord.* 67:a687. doi: 10.4102/sajcd.v67i2.687

Heyer, N., Morata, T. C., Pinkerton, L., Brueck, S. E., Stancescu, D., Panaccio, M. P., et al. (2011). Use of historical data and a novel metric in the evaluation of the effectiveness of hearing conservation program components. *Occup. Environ. Med.* 68, 510–517. doi: 10.1136/oem.2009.053801

Himi, S. T., Monalisa, N. T., Whaiduzzaman, M., Barros, A., and Uddin, M. S. (2023). MedAi: a smart watch-based application framework for the prediction of common diseases using machine learning. *IEEE Access* 11, 12342–12359. doi: 10.1109/ACCESS.2023.3236002

Hussain, I., Cawood, F., and van Olst, R. (2017). Effect of tunnel geometry and antenna parameters on through-the-air communication systems in underground mines: survey and open research areas. *Phys. Commun.* 23, 84–94. doi: 10.1016/j.phycom.2017.03.002

Ikeda, H., Kolade, O., Mahboob, M. A., Cawood, F. T., and Kawamura, Y. (2021). Communication of sensor data in underground mining environments: an evaluation of wireless signal quality over distance. *Mining* 1, 211–223. doi: 10.3390/mining1020014

Joseph, J. (2023). A Fully Functional DIY ESP32 Smart Watch with Multiple Watch Faces, Heart Rate Sensor, Compass and Games. Available at: https://circuitdigest.com/microcontroller-projects/diy-smart-watch-using-esp32-final-part (Accessed February 20, 2023)

Kanji, A., Khoza-Shangase, K., and Ntlhakana, L. (2019). Noise-induced hearing loss: what south African mineworkers know. *Int. J. Occup. Saf. Ergon.* 25, 305–310. doi: 10.1080/10803548.2017.1412122

Khoza-Shangase, K. (2022). "Confronting realities to hearing conservation programmes in South African mines" in *Occupational Noise-Induced Hearing Loss: An African Perspective*. eds. K. Khoza-Shangase and N. F. Moroe (Cape Town, South Africa: AOSIS Books, AOSIS Publishing (Pty) Ltd), 17–38.

Khoza-Shangase, K., Moroe, N. F., and Edwards, A. (2020). Occupational hearing loss in Africa: an interdisciplinary view of the current status. *South Afr. J. Commun. Disord.* 67, 1–3. doi: 10.4102/sajcd.v67i2.700363

Kolade, O., and Cheng, L. (2021). Markov model characterization of a multicarrier narrowband Powerline Channel with memory in an underground mining environment. *IEEE Access* 9, 59085–59092. doi: 10.1109/ACCESS.2021.3072669

Kolade, O., Familua, A. D., and Cheng, L. (2021). Channel models for an indoor power line communication system. *IET Commun. Technol. Netw. Smart Cities* 90:67.

Madahana, M., Ekoru, J., Mashinini, T., and Nyandoro, O. T. C. (2019a). Noise level policy advising system for mine workers. *IFAC Papers Online* 52, 249–254. doi: 10.1016/j.ifacol.2019.09.195

Madahana, M., Ekoru, J., Mashinini, T., and Nyandoro, O. T. C. (2019b). Mine workers threshold shift estimation via optimization algorithms for deep recurrent neural networks. *IFAC Papers Online* 52, 117–122. doi: 10.1016/j.ifacol.2019.09.174

Madahana, M., Nyandoro, O. T. C., and Ekoru, J. (2020). Intelligent comprehensive occupational health monitoring system for mine workers. *IFAC Papers Online* 53, 16494–16499. doi: 10.1016/j.ifacol.2020.12.751

Mardonova, M., and Choi, Y. (2018). Review of wearable device technology and its applications to the mining industry. *Energies* 11:547. doi: 10.3390/en11030547365

Masterson, A., Deddens, J., Themann, C., Bertke, S., and Calvert, G. M. (2015). Trends in worker hearing loss by industry sector, 1981-2010. *Am. J. Ind. Med.* 58, 392–401. doi: 10.1002/ajim.22429

Matetic, RJ, 31st International Conference of Safety in Mines Research Institutes, 2-5 October 2005, Brisbane, Queensland, Australia. Australia: Safety in Mines Testing and Research Station (Simtars), pp. 133–137; (2005).

MHSC (2015). 'Every mine worker returning from work unharmed every day. Striving for Zero Harm' - 2014 occupational health and safety summit milestones. *J. Occupat. Health South. Africa* 20:6.

Moroe, N. (2020). Occupational noise induced hearing loss in the mining sector in South Africa: perspectives from occupational health practitioners on how mineworkers are trained. *South Afr. J. Commun. Disord.* 67, e1–e6. doi: 10.4102/sajcd.v67i2.676

Moroe, N. F. (2020). Occupational noise-induced hearing loss in south African large-scale mines: exploring hearing conservation programmes as complex interventions embedded in a realist approach. *Int. J. Occup. Saf. Ergon.* 26, 753–761. doi: 10.1080/10803548.2018.1498183

Moroe, N., and Khoza-Shangase, K. (2018a). Management of occupational noise induced hearing loss in the mining sector in South Africa: where are the audiologists? *J. Occup. Health* 60, 376–382. doi: 10.1539/joh.2018-0020-OA

Moroe, N. F., and Khoza-Shangase, K. (2018b). Research into occupational noise induced hearing loss in south African large-scale mines: access denied? *AAS Open Res.* 1:4. doi: 10.12688/aasopenres.12829.1

Moroe, N., Khoza-Shangase, K., Kanji, A., and Ntlhakana, L. (2018). The management of occupational noise–induced hearing loss in the mining sector in Africa: a systematic review–1994 to 2016. *Noise Vib. Worldw.* 49, 181–190. doi: 10.1177/0957456518781860

Moroe, N. F., Ntlhakana, L., Luisa Petrocchi-Bartal, L., and Khoza-Shangase, K. (2022). "Hearing conservation programmes implementation in african mining contexts: occupational audiology in action" in *Occupational Noise-Induced Hearing Loss: An African Perspective*. eds. K. Khoza-Shangase and N. F. N. F. Moroe (Cape Town: AOSIS Books), 61–73.

Msiza, D. (2014). *The Road to Zero Harm: New Milestones*. Mine Health and Safety Council, Johannesburg.

Musiba, Z. (2015). Classification of audiograms in the prevention of noise-induced hearing loss: a clinical perspective. *South Afr. J. Commun. Disord.* 67, e1–e5. doi: 10.4102/sajcd.v67i2.691

NIOSH. (2023). National Institute for Occupational Safety and Health, Mining Program. Available at: https://www.cdc.gov/niosh/mining/topics/hearinglosspreventionoverview.html. (Accessed June 15, 2023)

Ntlhakana, L., Kanji, A., and Khoza-Shangase, K. (2015). The use of hearing protection devices in South Africa: exploring the current status in a gold and a non-ferous mine. *Occupat. Health South. Africa* 21, 10–15.

Ntlhakana, L., Nelson, G., Khoza-Shangase, K., and Dorkin, E. (2022). Occupational hearing loss for platinum miners in South Africa: a case study of data sharing practices and ethical challenges in the mining industry. *Int. J. Environ. Res. Public Health* 19:1. doi: 10.3390/ijerph19010001390

Pillay, M., Tiwari, R., Kathard, H., and Chikte, U. (2020). Sustainable workforce: south African audiologists and speech therapists. *Hum. Resour. Health* 18, 1–13. doi: 10.1186/s12960-020-00488-6

Puckett, S., and Emil, J. (2023). ecoSync: an energy-efficient clock discipline data synchronization in Wi-fi IoMT systems. *Electronics* 12:4226. doi: 10.3390/electronics12204226

Republic of South Africa (1995). Occupational Health and Safety Act, 1993, Hazardous Chemical Substance Regulations 1995. GNR 1179 of 25 August 1995, Government Printers, Pretoria. Available at: http://www.safetycon.co.za/documents/Hazardous%20Chemical%20Substances%20Regulations. (Accessed March 18, 2018)

Strauss, S., Swanepoel, D. W., Becker, P., Hall, J. W. I. I. I., and Eloff, Z. (2012). Prevalence and degree of noise-induced hearing loss in south African gold miners. *Occupat. Health South. Afr.* 18, 20–25. doi: 10.10520/EJC128495

Suter, A. H. (2002). Construction noise: exposure, effects, and the potential for remediation: a review and analysis. *AIHA J.* 63, 768–789. doi: 10.1080/15428110208984768398

Volsa, S., Batinic, B., and Stieger, S. (2022). Self-reports in the field using smart watches: an open-source firmware solution. *Sensors* 22:1980. doi: 10.3390/s22051980

Zaman, I., Förster, A., Mahmood, A., and Cawood, F., "Finding Trapped Miners with Wireless Sensor Networks," 5th International Conference on Information and Communication Technologies for Disaster Management (ICT-DM), Sendai, Japan, 2018; (2018). pp. 1–8.

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 18 April 2024
DOI 10.3389/fnins.2024.1373729

Check for updates

OPEN ACCESS

EDITED BY
Faheema Mahomed-Asmail,
University of Pretoria, South Africa

REVIEWED BY
Jani Johnson,
University of Memphis, United States
Colleen Zenczak-Magill,
Johns Hopkins University, United States

*CORRESPONDENCE
Tong Sheng
✉ tong.sheng@eargo.com

RECEIVED 24 January 2024
ACCEPTED 04 April 2024
PUBLISHED 18 April 2024

CITATION
Sheng T, Pasquesi L, Gilligan J, Chen X-J and Swaminathan J (2024) Subjective benefits from wearing self-fitting over-the-counter hearing aids in the real world.
Front. Neurosci. 18:1373729.
doi: 10.3389/fnins.2024.1373729

COPYRIGHT
© 2024 Sheng, Pasquesi, Gilligan, Chen and Swaminathan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Subjective benefits from wearing self-fitting over-the-counter hearing aids in the real world

Tong Sheng*, Lauren Pasquesi, Jennifer Gilligan, Xing-Jie Chen and Jayaganesh Swaminathan

Eargo, Inc., San Jose, CA, United States

Introduction: In 2022, the US Food and Drug Administration enacted final regulations to establish the category of over-the-counter (OTC) hearing aids aimed at reducing barriers to access hearing health care for individuals with self-perceived mild to moderate hearing loss. However, given the infancy of this device category, the effectiveness of OTC hearing aids in real-world environments is not yet well understood.

Methods and results: To gain insights into the perceived benefit of self-fitting OTC hearing aids, a two-pronged investigation was conducted. In the primary investigation, 255 active users of a self-fitting OTC hearing aid were surveyed on their perceived benefit using an abridged form of the Satisfaction with Amplification in Daily Living (SADL) scale. The mean global (4.9) and subscale scores (Positive Effect (PE): 4.3; Negative Features (NF): 4.3; Personal Image (PI): 6.1) were within the range of those previously reported for users of prescription hearing aids. In the secondary investigation, 29 individuals with self-reported hearing impairment but no prior experience with the investigational self-fitting OTC hearing aids used the devices and reported their perceived benefit and satisfaction following short-term usage. For this prospective group, the global SADL (5.4) and subscale scores (PE: 4.8; NF: 4.9; PI: 6.5) following a minimum of 10 weeks of real-world use were also within the range of those previously reported for traditional hearing aid users. In addition, this prospective group was also asked quality of life questions which assessed psychological benefits of hearing aid use. Responses to these items suggest hearing aid related improvements in several areas spanning emotional health, relationships at home and at work, social life, participation in group activities, confidence and feelings about one's self, ability to communicate effectively, and romance.

Discussion: Converging data from these investigations suggest that self-fitting OTC hearing aids can potentially provide their intended users with a level of subjective benefit comparable to what prescription hearing aid users might experience.

KEYWORDS

hearing loss, over-the-counter hearing aids, patient satisfaction, quality of life, SADL

1 Introduction

According to the National Institute on Deafness and Other Communication Disorders (NIDCD), approximately 28.8 million adults in the United States could benefit from the use of hearing aids (NIDCD, 2021). Hearing loss is a prevalent condition with approximately 40 million adults (15%) in the United States reporting having at least a little trouble with hearing (Pleis and Lethbridge-Cejku, 2007). Hearing loss is disproportionately overrepresented among

older adults, with nearly 25% of American adults aged 45 years and older reporting trouble hearing (Pleis and Lethbridge-Cejku, 2007), and its prevalence doubling with each additional decade of life (Lin et al., 2011).

In addition to affecting one's ability to communicate, hearing loss has also been associated with adverse physical and mental health outcomes. Among older adults over 70 years of age, those with hearing loss were also more likely to have a history of cardiovascular disease and stroke, resulting in an increased mortality risk (Contrera et al., 2015). Hearing loss is also associated with worse depressive symptoms (West et al., 2023), greater prevalence of dementia (Huang et al., 2023), higher rates of difficulties in activities of daily living (Dalton et al., 2003; Choi et al., 2016), and poorer quality of life (QoL), especially pertaining to social and emotional relationships (Ciorba et al., 2012).

While associations between hearing loss and adverse health outcomes are correlational, hearing loss is considered a modifiable risk factor, and its rehabilitation may have some potential to lessen the deterioration of health and quality of life. Early screening and adoption of hearing aid use can play a role in maintaining a positive quality of life (Brodie et al., 2018), and hearing aid use has also been associated with reduced anxiety and depression symptoms, improved QoL indicators, and reduced hearing-related social and emotional impediments (Ciorba et al., 2012). Among older adults with moderate to severe hearing loss, hearing aid use has been associated with lower prevalence of dementia (Huang et al., 2023), and a recent intervention study showed that older adults who were at greater risk of cognitive decline showed less cognitive decline following hearing aid use than those who did not use hearing aids (Lin et al., 2023).

Although the evidence for the rehabilitative benefits of hearing aids continues to accumulate, widespread hearing aid adoption has been stymied by factors such as lack of awareness and motivation (Angara et al., 2021; Zheng et al., 2023), as well as difficulty accessing hearing healthcare, all of which may contribute to the delayed diagnosis and treatment of hearing loss. Furthermore, while hearing aids are viewed as a relatively cost-effective rehabilitation tool, disparate insurance reimbursement policies and potentially high out-of-pocket costs can present a financial barrier to entry (Jilla et al., 2023). These factors can lead to individuals with hearing loss to not seek hearing healthcare altogether or only begin to use hearing aids after their condition has already worsened significantly.

Although the NIDCD reports the need for hearing aid adoption is high, usage continues to be low with 30% utilization for hearing aid candidates aged 70 and above and 16% utilization for candidates aged 20 to 69 (NIDCD, 2021). In recognition of this, the United States Food and Drug Administration (FDA) established a new category of over-the-counter (OTC) hearing aids for adults with self-perceived mild to moderate hearing impairment. By creating the OTC hearing aid category, the FDA hopes to reduce barriers to access and readily unlock the benefits associated with hearing aid use for those with hearing loss (FDA, 2023).

The OTC category of hearing aids aims to promote hearing aid adoption and use through two main objectives: 1) by establishing a category of devices that is accessible independent of the involvement of a hearing healthcare professional, and 2) by ensuring that the devices can be controlled (i.e., adjusted) directly by the end user (21CFR800.30, n.d.). By definition, OTC hearing aids are air-conduction hearing aids that do not require a hearing healthcare professional to procure or prescribe, and do not require implantation or other surgical means to fit to a user's ears. Such devices must also have user controls that enable the end user to adjust the devices based on their hearing needs.

To mitigate the risks associated with making such medical devices available to a potentially broad user base, device manufacturers must satisfy a range of controls in order for a product to meet the requirements of an OTC hearing aid (21CFR800.30, n.d.). These controls include software labeling, device output limits (i.e., maximum acoustic output limits), electroacoustic performance (e.g., distortion, latency, frequency response), and design requirements (e.g., maximum insertion depth, use of atraumatic materials, user controls).

There are broadly two kinds of OTC hearing aids: those with preset amplification levels, and those that are self-fitting. Self-fitting OTC hearing aids, which can be customized based on an individual's hearing loss, require FDA 510(k) clearance, including submission of clinical data, to validate the effectiveness of the self-fitting strategy. Self-fitting OTC hearing aids aim to be easily-accessible and user-friendly, as they can be obtained and fit without the involvement of a hearing healthcare professional. Removing the need to be seen in-person by a hearing care professional may encourage more people with hearing loss to use hearing aids, and at earlier stages of their hearing loss progression, by providing direct access to OTC devices.

While self-fitting OTC hearing aids intend to be more accessible, whether these devices will be perceived as beneficial in isolation and/or compared to prescription hearing aids fit by an audiologist following real world device wear by its users is yet to be seen. Clinically validated questionnaires have been developed to investigate the satisfaction of hearing aid users with real world device wear. Among the most known and used is the Satisfaction with Amplification in Daily Life questionnaire (SADL; Cox and Alexander, 1999). The SADL was developed to ascertain an overall sense of a user's satisfaction with hearing aids, as well as satisfaction in more specific areas related to hearing aid procurement and use (e.g., positive effect, service and cost, negative features, and personal image). Thus, the SADL aimed to quantify the degree of satisfaction with the use of a hearing aid, its perceived benefit, and allow for the identification of adverse aspects of adaptation of hearing aids.

While the real-world benefits and satisfaction with prescription hearing aids (fit by an audiologist following clinical best practice methods) have been extensively studied using surveys such as SADL, perceived benefit from the use of OTC hearing aids has been minimally studied. In the current study, we seek to gain insight into the extent to which users of a commercially-available, FDA-cleared self-fitting OTC hearing aid system report satisfaction and a sense of perceived benefit following real-world device use. To gain insights into the perceived benefit of OTC hearing aids, we conducted 1) a retrospective satisfaction survey study involving current users of an FDA-cleared self-fitting OTC hearing aids, and 2) a prospective cohort study with individuals who fit the description of the intended users of OTC hearing aids who had not previously used the investigational self-fitting OTC hearing aids.

The retrospective satisfaction survey would provide a better understanding of the impact of self-fitting hearing aids after its users acclimatized to and integrated the devices into their everyday lives, whereas the prospective cohort study would provide insight into the onboarding journey and usage experience of those who are new to the investigational self-fitting hearing aids.

2 Methods

In the context of these investigations, the study devices were commercially available Eargo hearing (www.eargo.com; San Jose, CA) aids that have been cleared by the FDA to be marketed as a Class II self-fitting air-conduction hearing aid (K221698, n.d.) and meet the controls set forth by the OTC rule (21CFR800.30, n.d.). The Eargo self-fitting OTC hearing aid system consists of a pair of completely-in-the-canal (CIC) style hearing aids (left and right), a charging case, and a companion mobile app. The Eargo self-fitting hearing aid uses a proprietary method that requires the user to complete a self-guided hearing assessment using the mobile app while wearing the hearing aids. The hearing aids act as the transducer, emitting tonal stimuli of varying levels at different audiometric frequencies. The measured hearing thresholds are then used as the basis for fitting the appropriate gain settings for the user. Once fit, the user can make additional adjustments (e.g., volume, bass/treble) to the left, right, or both hearing aids using the mobile app to achieve a desired fitting. Eargo's self-fitting hearing aids have been rigorously clinically validated and has been shown to provide adults with mild to moderate hearing loss with functional performance that is non-inferior to that provided by a professional hearing aid fitting (Hu et al., 2022; Urbanski et al., 2022; K221698, n.d.).

2.1 Retrospective self-fitting OTC hearing aid use satisfaction

To gain an understanding of user satisfaction and perceived benefit associated with the use of Eargo self-fitting OTC hearing aids, we leveraged the Eargo user base to identify individuals who had purchased an Eargo self-fitting OTC hearing aid and who had completed the product's self-fitting feature using its companion mobile application. In addition, we limited the query to identify only those who have purchased their devices at least 90 days prior to executing the query to constrain the sample to those who have had a chance to acclimatize to the hearing aids.

A random sample of subjects among those meeting the above criteria were invited to participate in a web-based survey about their experience using Eargo self-fitting OTC devices. Participation in the survey study was completely voluntary, and those who consented to participate in the survey received compensation in the form of a $25 Amazon gift card. The survey consisted of device usability and satisfaction questions, including questions from an abridged form of the Satisfaction with Amplification in Daily Living (SADL) questionnaire (Cox and Alexander, 1999).

The SADL scale is a 15-item questionnaire that assesses satisfaction with the use of hearing aids. With the SADL, hearing aid satisfaction can be interpreted using a global score as well as four subscale scores. The global and subscale satisfaction scores are interpreted on a 7-point scale, with 1 corresponding to least satisfaction ("Not At All") and 7 corresponding to the greatest satisfaction ("Tremendously"). There are four items that are phrased in the negative, and therefore, reverse scored. The scoring of the global and individual subscales is otherwise straight-forward, with the global score calculated as the mean score of all items completed by the participant, and individual subscale scores calculated as the mean score of all items completed by the participant within each subscale. The derived subscale and global satisfaction scores are interpreted with higher scores corresponding to higher satisfaction.

The individual subscales are: Positive Effect (PE): assessing functional benefit and satisfaction with overall hearing aid sound quality and use; Service & Cost (SC): assessing the fitting professional, product cost, and reliability/maintenance of hearing aids; Negative Features (NF): assessing the satisfaction with acoustic performance and feedback in specific challenging conditions; and Personal Image (PI): assessing the satisfaction with the hearing aids' *in-situ* physical appearance and perceived stigma when wearing hearing aids.

The SADL inventory was originally developed to evaluate satisfaction with prescription hearing aids well before the OTC category of hearing aids was established. Therefore, items related to SC may not accurately assess user sentiment in the context of OTC hearing aids, nor offer a meaningful interpretation of its score compared to the published norms for prescription hearing aid use. As such, questions related to SC were not included in the abridged SADL questionnaire administered to the users of Eargo self-fitting OTC hearing aids. The scoring instructions permit the omission of individual items with respect to subscale and global scores. However, as the omission of individual items impacts the calculation of the global SADL score, this metric should be interpreted to exclude the service and cost aspects of obtaining and using hearing aids and with caution while comparing with published normative data. However, the individual subscale satisfaction scores related to PE, NF, and PI do offer a more direct comparison with published norms for prescription hearing aids.

2.2 Prospective cohort: self-fitting OTC hearing aid use satisfaction

To gain an understanding of user satisfaction and perceived benefit associated with the first-time use of Eargo self-fitting OTC hearing aids among OTC hearing aid candidates, we recruited individuals who met the description for OTC hearing aid intended users, and who had no prior experience with Eargo's self-fitting OTC hearing aid products, to participate in a prospective cohort study. Potential candidates were recruited for screening via local advertising, word of mouth, and a customer database search. Intended users of OTC hearing aids were defined by the FDA as adults with self-reported mild-to-moderate hearing impairment, and this included individuals who have trouble hearing speech in noisy places, find it difficult to follow speech in groups, have trouble hearing on the phone, become tired when listening, and need to turn up the volume on the TV or radio to a level where other people complain it's too loud.

Participants who met the criteria described above and who consented to participating in the study were provisioned with retail-equivalent Eargo self-fitting OTC hearing aids (including all product package labeling and instructions for use that would accompany the system as if it were purchased commercially), along with a retail-equivalent investigational companion mobile app.

To approximate the journey of a would-be retail client of Eargo self-fitting OTC hearing aids, we asked participants to wear the devices to the extent that they found appropriate or desirable, and

provided no further instructions apart from requesting that they perform the app-based self-fitting procedure. This was to ensure that participants experienced the self-fitting process and that they would be testing and providing feedback on a self-fit hearing aid system. Otherwise, participants were expected to navigate their own hearing aid onboarding journey by using their devices as often or occasionally as they wished, and to review the included instructional materials for device troubleshooting. Participants were allowed to contact study staff if they had any questions, and research staff provided a scope and extent of support that mirrored those available to retail clients.

While all enrolled participants had to meet the criterion of not having prior experience with Eargo self-fitting OTC hearing aids, they were not excluded if they had previously tried or used other hearing devices.

Participants were given at least 10 56 weeks to become familiar with the study devices and to use the devices as much or as little as they felt appropriate in their everyday lives. At the conclusion of the study, all participants were administered a web-based survey on their experiences and satisfaction with using the study device and provided compensation in the form of a $75 Amazon gift card for participating in the study. As the study enrollment occurred on a rolling basis, and the final survey administration occurred at a fixed time point, several participants spent more than the minimum 10 weeks testing the study devices.

The survey consisted of device usability and satisfaction items, including the same abridged SADL questionnaire described above for the retrospective study. Overall satisfaction (global SADL score) and satisfaction in PE, NF, and PI were assessed. In addition, subjects were also asked quality of life (QoL) questions adapted from the MarkeTrak VIII survey (Kochkin, 2011). These questions assessed whether hearing aid users endorsed improvements across various QoL domains – emotional health, mental ability (memory), physical health, relationships at home, relationships at work, social life, feelings about oneself, ability to participate in group activities, sense of independence, sense of safety, confidence in oneself, sense of humor, romance in one's life, and overall ability to communicate more effectively – that participants believed to be attributable to hearing aid use. These questions were administered by asking respondents to "rate the changes you have experienced in the following areas, that you believe are due to your hearing aids" and each scored on a 4-point scale from 1 = "Worse" to 4 = "A lot better."

3 Results

3.1 Retrospective self-fitting OTC hearing aid use satisfaction

We identified a random sample of 393 Eargo self-fitting hearing aid subjects, and among these, 255 subjects met the inclusion criterion of having completed self-fitting using their hearing aids and the mobile app, and completed the abridged SADL questionnaire (see Table 1 for sample characteristics). Most of the respondents were experienced everyday users of the devices; nearly two-thirds had used their devices for more than 6 months (65.1%), over three-quarters

TABLE 1 Retrospective and prospective sample characteristics.

	Retrospective sample characteristics	Prospective sample characteristics
Sample size:	*N* = 255	*N* = 29
Age:	69 years (median)	70 years (median)
	62–74.5 years (interquartile range)	64–77 years (interquartile range)
Gender:	79.2% male	72.4% male
	20.4% female	27.6% female
	0.4% other/prefer not to say	0% other/prefer not to say
Self-reported hearing difficulty	Mild: 49%	Mild: 37.9%
	Moderate: 46.3%	Moderate: 62.1%
	Severe: 4.7%	Severe: 0%
Self-reported device usage (weekly)	1–2 days/week: 22.4%	1–2 days/week: 14.3%
	3–5 days/week: 28.6%	3–5 days/week: 50%
	6–7 days/week: 49%	6–7 days/week: 35.7%
Self-reported device usage (daily)	<4 h/day: 22%	<4 h/day: 14.3%
	4–8 h/day: 32.5%	4–8 h/day: 28.6%
	8+ hours/day: 45.5%	8+ hours/day: 57.1%
Eargo self-fitting hearing aid use history	3–6 months: 34.9%	<1 month: 6.9%
	6–12 months: 48.2%	1–2 months: 51.7%
	>12 months: 16.9%	>2 months: 37.9%
Self-reported lifetime hearing aid use history	<1 year: 46.3%	<1 year: 28%
	1–10 years: 49%	1–10 years: 69%
	>10 years: 4.7%	>10 years: 3%

TABLE 2 Retrospective and prospective cohort hearing aid satisfaction: SADL global and subscale scores.

SADL	Retrospective sample satisfaction results $N = 255$		Prospective sample satisfaction results $N = 29$		Published norms from Cox and Alexander (1999)
Global	MEAN[a]: 4.9	T = 0.25	MEAN[a]: 5.4	**T = 4.08**	MEAN: 4.9
	SD: 0.9	$p = 0.80$	SD: 0.7	P = **0.0003**	20th–80th range: 4.2–5.9
	IQR: 4.2–5.6		IQR: 4.8–6.1		
Positive effect	MEAN: 4.3	**T = 7.21**	MEAN: 4.8	T = 0.33	MEAN: 4.9
	SD: 1.3	**p < 0.0001**	SD: 1.2	$p = 0.75$	20th–80th range: 3.8–6.1
	IQR: 3.3–5.3		IQR: 3.8–5.8		
Negative features	MEAN: 4.3	**T = 8.65**	MEAN: 4.9	**T = 5.45**	MEAN: 3.6
	SD: 1.3	P < **0.0001**	SD: 1.3	P < **0.0001**	20th–80th range: 2.3–5.0
	IQR: 3.3–5.3		IQR: 4.0–6.0		
Personal image	MEAN: 6.1	**T = 11.77**	MEAN: 6.5	**T = 7.97**	MEAN: 5.6
	SD: 0.7	P < **0.0001**	SD: 0.6	P < **0.0001**	20th–80th range: 5.0–6.7
	IQR: 5.7–6.7		IQR: 6.2–6.8		

[a]The global satisfaction score derived from the abridged SADL questionnaire in the present study omits questions related to Service and Cost. Thus, its interpretation should be considered to not pertain to the Service and Cost aspects of obtaining and using hearing aids. Bold indicates $p < 0.05$ (two-tailed).

reported using their devices at least 3 or more days per week (77.6%; Table 1), and nearly half reported using their hearing aids for at least 8 h a day (45.5%). The sample was evenly split with respect to the severity of self-reported hearing impairment (49% mild; 46.3% moderate).

The global, as well as individual subscales satisfaction scores derived from the abridged SADL questionnaire, were mostly positive (Table 2). The mean modified global satisfaction (absent SC items) score of 4.9 was comparable to published satisfaction scores for traditional hearing aids obtained through private practice and fit by an audiologist following clinical best practice methods (study mean = 4.9 vs. norm mean of 4.9; T = 0.25, $p = 0.80$). For the subscale scores, the Positive Effect was slightly poorer than published norms (study mean = 4.3 vs. norm mean of 4.9), while Negative Features (study mean = 4.3 vs. norm mean of 3.6), and Personal Image (study mean = 6.1 vs. norm mean of 5.6) subscale scores were better than published norms. While the differences in subscale scores were statistically different from published norms (all Ps < 0.05), the interquartile range for each subscale overlapped with the previously reported ranges (i.e., 20th-80th percentile ranges) for users of prescription hearing aids (Cox and Alexander, 1999). A post-hoc power calculation indicated that the study had sufficient power (100%) to detect a satisfaction score difference of 0.5 (with standard deviation of 1.0) at alpha = 0.05 with the 255 respondents.

3.2 Prospective cohort: self-fitting OTC hearing aid use satisfaction

Thirty-three adults were enrolled into the prospective cohort study and twenty-nine subjects provided responses on the final survey. For this cohort, 37.9% self-reported having mild hearing impairment, while 62.1% self-reported having moderate hearing impairment (Table 1). The vast majority reported regularly using the study devices for at least 1 month (89.6%), with 37.9% reporting using the study devices for at least 2 months (Table 1). With respect to device usage, 85.7% reported using the devices at least 3 or more days per week, and 57.1% reporting using the devices for 8 or more hours per day.

Among this cohort of OTC hearing candidates who were new to using Eargo self-fitting hearing aids, the levels of self-reported satisfaction following this short-term device trial were within the expected range of satisfaction scores for prescription hearing aids (Table 2). Notably, the modified global satisfaction (absent SC items) following short-term wear was significantly higher than the global satisfaction score reported for prescription hearing aid users (study mean = 5.4 vs. norm mean of 4.9; T = 4.08, $p = 0.0003$). Satisfaction scores in the Negative Features (mean = 4.9) and Personal Image (mean = 6.5) subscales were significantly higher than published norms for prescription hearing aids (all $ps < 0.05$), although the interquartile range of individual SADL subscales overlapped with the ranges (i.e., 20th-80th percentile ranges) published for prescription hearing aids (Cox and Alexander, 1999).

With respect to self-reported QoL improvements attributable to the short-term use of Eargo self-fitting hearing aids, there was near-unanimous endorsement of stability or improvement in all domains assessed (at least 96% of respondents reported "same" or "better" on all 14 questions). In the following QoL domains, more than half of the responding sample reported improvements stemming from wearing self-fitting hearing aids: emotional health (54.5%), relationships at home (64%), relationships at work (61.1%), social life (65.4%), feeling about oneself (60.9%), ability to participate in group activities (60%), confidence in oneself (54.2%), romance (100%), and overall ability to communicate more effectively (69.2%; Table 3). A post-hoc power calculation indicated that while the initial sample of 33 participants demonstrated sufficient power (>80%) to detect a satisfaction score difference of 0.5 (standard deviation of 1.0) at alpha = 0.05, the sample of 29 respondents yielded a power of 76.8%.

4 Discussion

This study assessed subjective benefits and satisfaction with real-world device wear using a clinically validated questionnaire (SADL) for

TABLE 3 Prospective cohort quality of life changes attributed to hearing aid use.

Quality of life domain (number of respondents)	Worse	Same	Better
Romance in my life ($n = 22$)	0%	0%	100%
Overall ability to communicate more effectively in most situations ($n = 25$)	0%	26.9%	69.2%
Social life ($n = 26$)	3.8%	30.8%	65.4%
Relationships at home ($n = 25$)	0%	36%	64%
Relationships at work ($n = 18$)	5.6%	33.3%	61.1%
Feelings about yourself ($n = 23$)	0%	39.1%	60.9%
Ability to participate in group activities ($n = 25$)	4%	36%	60%
Emotional health ($n = 22$)	0%	45.5%	54.5%
Confidence in yourself ($n = 25$)	0%	45.8%	54.2%
Mental ability (memory) ($n = 21$)	0%	52.4%	47.6%
Sense of independence ($n = 23$)	0%	60.9%	39.1%
Sense of safety ($n = 23$)	0%	60.9%	39.1%
Sense of humor ($n = 24$)	0%	75%	25%
Physical health ($n = 21$)	0%	81%	19%

an FDA-cleared self-fitting OTC hearing-aid system (Eargo) in adults with self-perceived hearing difficulties. Two cohorts were recruited for this study: 1) A retrospective cohort with longer acclimatization and integration of self-fitting OTC hearing-aids into their everyday lives; and 2) A prospective cohort who were new to the investigational self-fitting hearing aids. While comparing between cohorts, the mean global and subscale satisfaction scores were better for the prospective cohort (vs. scores from retrospective cohort). While it may be tempting to interpret these differences as a stabilization of perceived benefit over time (for example, initial excitement may be driving higher satisfaction in the prospective cohort), any comparison and interpretation of SADL scores between the two cohorts should be done with caution due to differences in sample size and characteristics.

However, comparisons between SADL scores from our retrospective and prospective cohorts and those reported in the literature for individuals wearing prescription hearing aids (fit by an audiologist following clinical best practice methods) can offer interesting insights. The mean modified global SADL satisfaction score (absent SC items) from the retrospective group was statistically similar to those reported in the literature for adults fit with prescription hearing aids (Cox and Alexander, 1999, 2001; Shi et al., 2007; Kozlowski et al., 2017), while the global SADL score observed in our prospective cohort was slightly elevated. This suggests that users of our investigational self-fitting OTC hearing aids who have had an opportunity to acclimatize to the devices experience a comparable level of overall satisfaction and benefit as those who have been fit with prescription hearing aids, whereas the brand new investigational device users in our prospective cohort could have exhibited some initial product excitement that may or may not temper over time.

With respect to the SADL subscale scores, the observed mean PE score in the retrospective cohort was slightly poorer than published norms, while the observed mean PE score in the prospective cohort was comparable to published norms. The PE subscale consists of items related to a device's functional performance and whether use of the device is worthwhile to the user. It is possible that hearing aid candidates who sought treatment through the traditional channel may be better aligned with respect to their expectations when embarking on their hearing aid journey. The higher satisfaction in our prospective cohort relative to the retrospective cohort could be due to the fact that these users were provisioned investigational devices as part of a product usability study, and did not obtain them through a retail or prescription channel.

With respect to the NF and PI subscales, converging evidence from both the retrospective and prospective cohorts indicate that the observed NF and PI subscale scores following use of the Eargo self-fitting OTC hearing aids were slightly better than the published norms for prescription hearing aid users. The favorable NF scores in both of our cohorts indicate that users felt the investigational self-fitting OTC hearing aids had good acoustic performance (i.e., adequate gain with acceptable amount of feedback). For the positive PI scores in both cohorts, a reasonable explanation could be that the CIC form factor of the Eargo self-fitting OTC devices were less visually obvious than other form factors when worn *in-situ*, which may in turn alleviate some of our users' concerns related to social stigma around hearing aid wear (Pasquesi et al., 2023).

However, although we have observed some slight differences in the mean scores across the global and subscale scores between our two cohorts and the published norms for prescription hearing aid users, the distributions (i.e., interquartile ranges) of all of our observed scores were largely comparable to the distributions (i.e., published 20th-80th percentile ranges, see Table 2) reported for prescription hearing aids. This, along with a few other study-specific details that may contribute to the interpretation of our data, encourage us to refrain from making absolute statements about satisfaction relative to prescription hearing aids.

For example, only individuals who have ordered a self-fitting OTC hearing aid system at least 90 days prior to the survey administration were eligible to participate in our retrospective study. This meant that any customers who have tried but returned their devices within the initial trial period were not included in the sample. While we did not otherwise exclude any potential participant based on complaints, return requests, customer support cases, or any other obvious indicators of device dissatisfaction, it is reasonable to assume that those who kept their hearing aids past 90 days may be a self-selecting group with a

slightly elevated baseline level of satisfaction with the devices. However, while device usage and experience has been shown to be linked to satisfaction (Uriarte et al., 2005; Vestergaard, 2006; Vestergaard Knudsen et al., 2010; Dashti et al., 2015), as the U.S. norm data also included responses from experienced hearing aid users (Cox and Alexander, 1999), we believe reasonable comparisons could still be made.

Another aspect to consider is that the SADL norms consisted of hearing aid satisfaction data from both private-pay and sponsored samples. While there was no difference observed between private-pay and sponsored respondents on global satisfaction, there was a difference in the SC subscale satisfaction score (Cox and Alexander, 1999). While we excluded the SC subscale items in both the retrospective and prospective cohort studies, we must acknowledge the inherent and implicit impacts of using a provisioned hearing aid on the expectations and perceptions of our prospective cohort participants. Our prospective cohort data were consistent with other studies where hearing aid cost was fully or partially sponsored demonstrating slightly elevated SADL scores compared to the published U.S. norms (Uriarte et al., 2005; Iwahashi et al., 2013; Dashti et al., 2015).

For the prospective cohort, the positive impact of self-fitting hearing aid use on QoL areas varied by area, but ranged from 19% of respondents endorsing improved physical health to 100% of respondents endorsing improved romantic life. Out of the 14 areas surveyed, all but three had unanimous responses of improvement or stability since using the investigational devices. In areas related to social functioning, 65.4% endorsed improvements in social life, 60% reported being better able to participate in group activities, and many reported improved relationships at home (64%) and at work (61.1%). Short-term hearing aid use was associated with an improved ability to communicate effectively in 69.2% of respondents. In areas related to sense of self, 60.9% reported improved feelings about oneself, 54.2% reported improved confidence, and 39.1% reported improved senses of safety and independence, respectively. Overall, improvements were endorsed by users across a number of QoL areas, particularly those related to communication, interpersonal relationships, and social functions. It is not entirely clear if and how the QoL responses may change with an extended duration of device wear. Future work may assess QoL improvements from a larger, real-world cohort. Our intent was to present these QoL data purely as descriptive findings; it would not be appropriate to directly compare our observations against those published in the MarkeTrak reports based on paying hearing aid customers. However, it was still interesting to note, with caution, that the endorsement of improvements observed with wearers of self-fitting OTC hearing-aid from this study were better than the responses reported with prescription hearing-aids fit by an audiologist following clinical best practice methods across several categories (Picou, 2022).

5 Conclusion

Taken together, and in considering some of the limitations of our study above, data involving new and experienced users of an investigational self-fitting OTC hearing aid suggest that users report a level of satisfaction and subjective benefit equivalent to or better than (in most areas assessed, with an exception on the PE subscale where the retrospective cohort reported lower satisfaction), those experienced by users of prescription hearing aids fit by audiologists following clinical best practice methods. Converging evidence from the retrospective and prospective cohorts with respect to consistent user-reported device usage and positive PI scores demonstrate the device category's potential impact on hearing aid use adoption. With the establishment of the OTC category of hearing aids, there is hope that access to hearing healthcare will broaden. Given that untreated hearing loss has negative implications on many aspects of physical, cognitive, and emotional health, breaking down barriers to device access can ultimately have an outsized impact on objective and subjective outcomes; certainly, in our small sample of prospective self-fitting OTC hearing aid users, many endorsed experiencing improvements in relationships and other social situations. While more research is needed to fully characterize the potential positive impact that self-fitting OTC hearing aids may have on health outcomes, here we present preliminary but encouraging evidence that the use of self-fitting OTC hearing aids can play a role in helping to preserve or improve perceived quality of life for adults with mild to moderate hearing loss.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Eargo Legal and Regulatory Review. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

TS: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Supervision, Validation, Writing – original draft, Writing – review & editing. LP: Formal analysis, Investigation, Writing – review & editing. JG: Methodology, Writing – original draft, Writing – review & editing. XC: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Validation, Writing – review & editing. JS: Conceptualization, Formal analysis, Methodology, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

TS, LP, JG, XC, and JS were employed by Eargo, Inc.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

21CFR800.30 (n.d.). Over-the-counter hearing aid controls. Code of Federal Regulations. Available at: https://www.ecfr.gov/current/title-21/part-800/section-800.30

Angara, P., Tsang, D. C., Hoffer, M. E., and Snapp, H. A. (2021). Self-perceived hearing status creates an unrealized barrier to hearing healthcare utilization. *Laryngoscope* 131, E289–E295. doi: 10.1002/lary.28604

Brodie, A., Smith, B., and Ray, J. (2018). The impact of rehabilitation on quality of life after hearing loss: a systematic review. *Eur. Arch. Otorhinolaryngol.* 275, 2435–2440. doi: 10.1007/s00405-018-5100-7

Choi, J. S., Betz, J., Deal, J., Contrera, K. J., Genther, D. J., Chen, D. S., et al. (2016). A comparison of self-report and audiometric measures of hearing and their associations with functional outcomes in older adults. *J. Aging Health* 28, 890–910. doi: 10.1177/0898264315614006

Ciorba, A., Bianchini, C., Pelucchi, S., and Pastore, A. (2012). The impact of hearing loss on the quality of life of elderly adults. *Clin. Interv. Aging* 7, 159–163. doi: 10.2147/CIA.S26059

Contrera, K. J., Betz, J., Genther, D. J., and Lin, F. R. (2015). Association of Hearing Impairment and Mortality in the National Health and nutrition examination survey. *JAMA Otolaryngol. Head Neck Surg.* 141, 944–946. doi: 10.1001/jamaoto.2015.1762

Cox, R. M., and Alexander, G. C. (1999). Measuring satisfaction with amplification in daily life: the SADL scale. *Ear Hear.* 20, 306–320. doi: 10.1097/00003446-199908000-00004

Cox, R. M., and Alexander, G. C. (2001). Validation of the SADL questionnaire. *Ear Hear.* 22, 151–160. doi: 10.1097/00003446-200104000-00008

Dalton, D. S., Cruickshanks, K. J., Klein, B. E. K., Klein, R., Wiley, T. L., and Nondahl, D. M. (2003). The impact of hearing loss on quality of life in older adults. *The Gerontologist* 43, 661–668. doi: 10.1093/geront/43.5.661

Dashti, R., Khiavi, F. F., Sameni, S. J., and Bayat, A. (2015). Satisfaction with hearing aids among aged patients with different degrees of hearing loss and length of daily use. *J. Audiol. Otol.* 19, 14–19. doi: 10.7874/jao.2015.19.1.14

FDA (2023). OTC hearing aids: What you should know. U.S. Food and Drug Administration. Available at: https://www.fda.gov/medical-devices/hearing-aids/otc-hearing-aids-what-you-should-know

Hu, J., Swaminathan, J., Kwan, J., Rodriguez, M., Dalager, A., and Walters, A. (2022). *Verification and validation of a self-fitting hearing device [poster]*. IHCON, Lake Tahoe, CA.

Huang, A. R., Jiang, K., Lin, F. R., Deal, J. A., and Reed, N. S. (2023). Hearing loss and dementia prevalence in older adults in the US. *JAMA* 329, 171–173. doi: 10.1001/jama.2022.20954

Iwahashi, J. H., de Souza Jardim, I., and Bento, R. F. (2013). Results of hearing aids use dispensed by a publicly-funded health service. *Braz. J. Otorhinolaryngol.* 79, 681–687. doi: 10.5935/1808-8694.20130126

Jilla, A. M., Johnson, C. E., and Huntington-Klein, N. (2023). Hearing aid affordability in the United States. *Disabil. Rehabil. Assist. Technol.* 18, 246–252. doi: 10.1080/17483107.2020.1822449

K221698 (n.d.). 510(k) Premarket Notification. Available at: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=k221698

Kochkin, S. (2011). MarkeTrak VIII patients report improved quality of life with hearing aid usage. *Hear. Res.* 64, 25–26. doi: 10.1097/01.HJ.0000399150.30374.45

Kozlowski, L., Ribas, A., Almeida, G., and Luz, I. (2017). Satisfaction of elderly hearing aid users. *Int. Arch. Otorhinolaryngol.* 21, 92–96. doi: 10.1055/s-0036-1579744

Lin, F. R., Niparko, J. K., and Ferrucci, L. (2011). Hearing loss prevalence in the United States. *Arch. Intern. Med.* 171, 1851–1853. doi: 10.1001/archinternmed.2011.506

Lin, F. R., Pike, J. R., Albert, M. S., Arnold, M., Burgard, S., Chisolm, T., et al. (2023). Hearing intervention versus health education control to reduce cognitive decline in older adults with hearing loss in the USA (ACHIEVE): a multicentre, randomised controlled trial. *Lancet (London, England)* 402, 786–797. doi: 10.1016/S0140-6736(23)01406-X

NIDCD (2021). Quick Statistics About Hearing | NIDCD. Available at: https://www.nidcd.nih.gov/health/statistics/quick-statistics-hearing

Pasquesi, L., Douda, J., and Swaminathan, J. (2023). Survey details OTC self-fitting hearing aid users' experiences, perceptions. The Hearing Review. Available at: https://hearingreview.com/hearing-products/hearing-aids/otc/survey-details-11000-otc-self-fitting-hearing-aid-users-experiences-perceptions

Picou, E. M. (2022). Hearing aid benefit and satisfaction results from the MarkeTrak 2022 survey: importance of features and hearing care professionals. *Semin. Hear.* 43, 301–316. doi: 10.1055/s-0042-1758375

Pleis, J. R., and Lethbridge-Cejku, M. (2007). Summary health statistics for U.S. *Adults: Nat. Health Inter. Sur.* 2006, 403882008–403882001. doi: 10.1037/e403882008-001

Shi, L.-F., Doherty, K. A., Kordas, T. M., and Pellegrino, J. T. (2007). Short-term and long-term hearing aid benefit and user satisfaction: a comparison between two fitting protocols. *J. Am. Acad. Audiol.* 18, 482–495. doi: 10.3766/jaaa.18.6.3

Urbanski, D., Nelson, P., Donato, S., Rosenthal, J., and Swaminathan, J. (2022). *Self-adjustment versus prescriptive fitting: How much of a difference really makes a difference? [poster]*. IHCON, Lake Tahoe, CA.

Uriarte, M., Denzin, L., Dunstan, A., Sellars, J., and Hickson, L. (2005). Measuring hearing aid outcomes using the satisfaction with amplification in daily life (SADL) questionnaire: Australian data. *J. Am. Acad. Audiol.* 16, 383–402. doi: 10.3766/jaaa.16.6.6

Vestergaard, M. D. (2006). Self-report outcome in new hearing-aid users: longitudinal trends and relationships between subjective measures of benefit and satisfaction. *Int. J. Audiol.* 45, 382–392. doi: 10.1080/14992020600690977

Vestergaard Knudsen, L., Öberg, M., Nielsen, C., Naylor, G., and Kramer, S. E. (2010). Factors influencing help seeking, hearing aid uptake, hearing aid use and satisfaction with hearing aids: a review of the literature. *Trends Amplif.* 14, 127–154. doi: 10.1177/1084713810385712

West, J. S., Smith, S. L., and Dupre, M. E. (2023). The impact of hearing loss on trajectories of depressive symptoms in married couples. *Soc. Sci. Med.* 321:115780. doi: 10.1016/j.socscimed.2023.115780

Zheng, H., Wong, L. L. N., and Hickson, L. (2023). Barriers to hearing aid adoption among older adults in mainland China. *Int. J. Audiol.* 62, 814–825. doi: 10.1080/14992027.2022.2105263

frontiers | Frontiers in Audiology and Otology

TYPE Original Research
PUBLISHED 01 May 2024
DOI 10.3389/fauot.2024.1389116

Check for updates

OPEN ACCESS

EDITED BY
Laura Coco,
San Diego State University, United States

REVIEWED BY
William Bologna,
Towson University, United States
Lauren Dillard,
Medical University of South Carolina,
United States

*CORRESPONDENCE
J. Riley DeBacker
debacker@ohsu.edu

RECEIVED 21 February 2024
ACCEPTED 01 April 2024
PUBLISHED 01 May 2024

CITATION
McMillan GP, DeBacker JR, Hungerford M and Konrad-Martin D (2024) Serial monitoring of the audiogram in hearing conservation using Gaussian processes.
Front. Audiol. Otol. 2:1389116.
doi: 10.3389/fauot.2024.1389116

COPYRIGHT
© 2024 McMillan, DeBacker, Hungerford and Konrad-Martin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Serial monitoring of the audiogram in hearing conservation using Gaussian processes

Garnett P. McMillan[1], J. Riley DeBacker[1,2]*, Michelle Hungerford[1] and Dawn Konrad-Martin[1,2]

[1]National Center for Rehabilitative Auditory Research, VA Portland Health Care System, Portland, OR, United States, [2]Department of Otolaryngology/Head and Neck Surgery, Oregon Health & Science University, Portland, OR, United States

Most hearing conservation programs repeatedly monitor a subject's pure tone thresholds before, during, and after exposure to audiopathic agents. Changes to the audiogram that meet significant shift criteria such as ASHA, CTCAE, and so forth are considered evidence of audiopathic injury. Despite a wide variety of definitions for significant change, all current serial monitoring methods are biased due to regression to the mean and are prone to inconclusive results. These problems diminish the diagnostic accuracy and utility of serial monitoring. Here we propose adopting Gaussian processes to address these issues in a manner that maximizes time efficiency and can be administered using portable equipment at the point of care.

KEYWORDS

serial monitoring, test-retest, Gaussian process, Bayesian analysis, hearing conservation

1 Introduction

Audiometric serial monitoring is the act of evaluating changes in hearing thresholds. Audiologists identify changes in a patient's hearing by comparing audiogram results over time. The rationale is that pure tone sensitivity, as measured by the audiogram, is susceptible to damage from audiopathic exposures such as noise or ototoxic medications, and reflect changes associated with normal aging and certain disease conditions. A change in pure-tone sensitivity is taken as evidence of potential audiopathic injury, motivating follow-up care and/or removal from the audiopathic exposure.

There are many serial monitoring criteria described in the audiology literature (reviews in King and Brewer, 2018 and Moore et al., 2022). There are three particular difficulties with all existing approaches:

1) **Lack of a gold standard**: since there is no gold standard for audiopathic injury and thus no way to evaluate the accuracy of these various criteria, it is up to the clinician or employer to choose among serial monitoring criteria based on clinical objectives, convention, intuition, invasiveness, time, expense, or any other priority. Priorities differ among the end users such as the audiologist, primary care clinician, employer, and patient. The various monitoring criteria can not simultaneously achieve the objectives of all stakeholders resulting in inefficient care.
2) **Bias due to no response**: the audiologist must also decide how to handle thresholds that exceed audiometer test limits, called "No Response" (NR), or how to handle missing thresholds due to patient non-response. The latter is particularly challenging in pediatric applications, while the former often occurs in older populations of patients. How NR and missing thresholds are handled will impact clinical judgements.

3) **Bias due to regression to the mean**: regression to the mean is the (almost) ineluctable fact that, barring any real changes, bigger than average baselines are *always* expected to get smaller and that smaller than average baselines are *always* expected to get bigger. This is *necessarily* true in (almost) any homeostatic system, real or imaginary. The previous parenthetical statements invoke certain technical points that can be studied in Samuels (1991). In the absence of audiopathic injury, regression to the mean (Royston, 1995) guarantees that on average a "large" or "high" baseline threshold will be followed by a "smaller" or "lower" one, and that a "small" baseline threshold will be followed by a "larger" one. This is expected regardless of any audiopathic injury that may have occurred. This clearly confounds any attempt to judge audiometric changes in terms of potential injury to the patient, because any observed changes are at least partially due to regression to the mean. A proper approach is to statistically condition the expected follow-up measurement on the previously observed baseline (Royston, 1995). The clinical expectation about a patient at follow-up naturally depends on what was observed at baseline, and a proper statistical expectation for a patient at follow-up must also depend on the previously observed baseline. Regression to the mean induces bias in all existing serial monitoring criteria (Royston, 1995).

Point (1) impacts most every facet of audiology or medicine. Points (2) and (3) occur in most hearing monitoring criteria because standard methods of evaluating changes in pure tone sensitivity are based on the computed difference between baseline and follow-up audiograms. While intuitive, the computed difference approach will cause bias and loss of information. The audiologist must manually perform the differencing computations to determine if a given criterion has been met. Subsequently, the audiologist must communicate the results to the patient and other stakeholders in their care (family, care team). There is a need for rapid or even real-time communication of these results, particularly when results indicate the need for care coordination, for example to eliminate or reduce the audiopathic exposure, or promote timely access to treatment. An unbiased, rapid and transparent way to communicate serial monitoring results would promote more efficient care.

We propose a different approach in this paper to address points (2) and (3). In our view, serial monitoring occurs under the assumption that pure tone sensitivity does not change between the baseline and follow-up time point. We call this the "Homeostasis Hypothesis," and audiometric serial monitoring is conducted to evaluate whether or not the Homeostasis Hypothesis is true. In this paper we develop a statistical model of the relationship between the audiogram and a patient's underlying pure tone sensitivity under the assumption that the Homeostasis Hypothesis is true. If the follow-up audiogram is unusual with respect to the expectations of the Homeostasis Hypothesis, then the audiologist has evidence against the assumption that pure tone sensitivity has remained constant over the course of exposure. Follow-up action is therefore warranted.

Figure 1 illustrates our approach as described in this paper. Given the patient's baseline audiogram as an input, we compute the predictive distribution of the follow-up audiogram under the assumption that the Homeostasis Hypothesis is correct. Correlations between thresholds in each ear and at neighboring frequencies, as well as population-based pure-tone threshold data are used to narrow the patient-specific predictive distribution. This predictive distribution is expressed as a simultaneous prediction region, which can be easily interpreted: nine out of 10 follow-up audiograms on the patient will be entirely within the region if the Homeostasis Hypothesis is correct. A follow-up audiogram that exceeds the audiogram prediction region at any frequency in either ear is evidence that the Homeostasis Hypothesis is false and that pure tone sensitivity has changed.

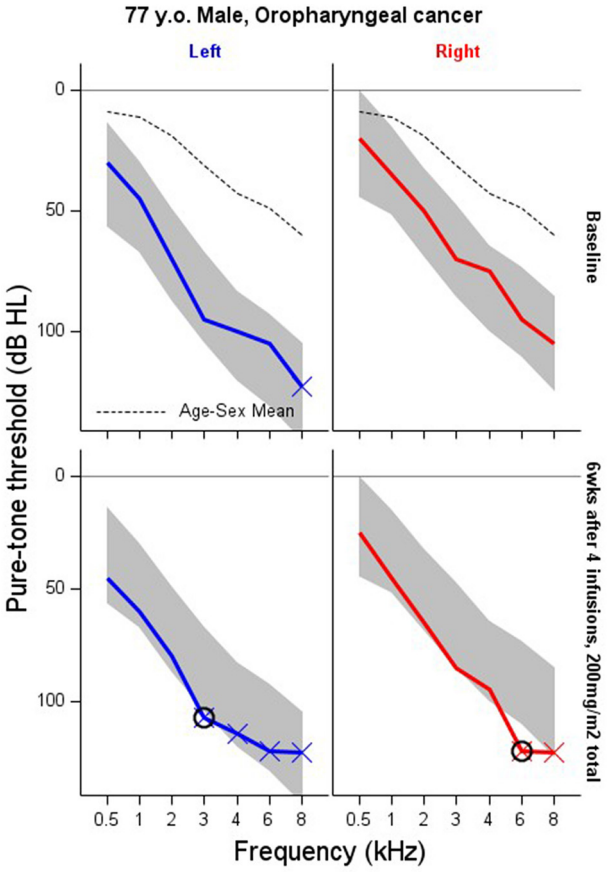


FIGURE 1
Serial monitoring an adult male, 77 years old, being treated for Oropharyngeal cancer with 40 mg/m^2, weekly. Solid lines show the pre-treatment baseline audiogram (top row) and 6-week follow-up (bottom row) after four infusions of cisplatin equaling a cumulative dose 200 mg/m^2. The dashed line is the ISO 1999-2013 (ISO, 2013) age-sex specific mean threshold obtained from population-based pure-tone threshold data. The "X" indicates No Response at that test frequency. The shaded area is a 90% prediction region for this patient's audiogram under the assumption that pure tone sensitivity hasn't changed from their baseline test. Follow-up thresholds exceeding the prediction region (circled) are "significant" changes to the audiogram that warrant follow-up.

In this paper we will take advantage of recent interest in Gaussian Processes in audiology (Song et al., 2015; Bao et al., 2017; Barbour et al., 2019). This methodology provides an alternative to traditional grading or binary scales that are prone to the biases discussed above. This methodology is suitable for patients

and employees at risk of audiopathic injury from any type of exposure (e.g., noise, bactericidal or antineoplastic therapies, etc.) as long as baseline audiometry is available. We do not make recommendations about pure tone test frequencies, testing intervals, or procedures for treating patients with audiopathic injury. These decisions are specific to each exposure and are left to the serial monitoring program. Our approach avoids bias and loss of information that affects current approaches, and we believe it can serve a wider range of clinical objectives and stakeholder priorities than standard criteria currently in use. These benefits are achieved at the cost of computational efficiency; i.e., a computer is required, though this burden is small since all computation is automated and done offline. This is an additional benefit of our approach over existing criteria since it maximizes time efficiency and can be administered using computer-based portable audiometry systems at the point of care. The prediction region such as seen in Figure 1 are computed prior to the follow-up exam, and thus do not impinge on patient-audiologist contact time.

2 Methods

The clinical problem for the audiologist is that of deciding whether a follow-up audiogram measured on a patient demonstrates evidence that pure tone sensitivity has degraded and that the Homeostasis Hypothesis is false. The statistical problem is that of defining (1) the relationship between pure tone sensitivity—a theoretical construct that we cannot observe directly—and its representation as the audiogram, and (2) defining the expected relationship between baseline and follow-up audiograms under the assumption that the Homeostasis Hypothesis is correct.

We assume that pure tone sensitivity δ in each ear e and across the frequency spectrum f at baseline time 0 and follow-up time t are Gaussian Processes with covariance functions K_0 and K_t, and population gender- and age-specific mean function $\mu\left(e,f\right)$:

$$\delta_0\left(e,f\right) \sim GP\left(\mu\left(e,f\right),\ K_0\right)$$

$$\delta_t\left(e,f\right) \sim GP\left(\mu\left(e,f\right),\ K_t\right).$$

This model contains the important assumption that the time that passes between baseline and follow-up measurements (i.e. t) is less than the amount of time that is required before the population age-specific mean function $\mu\left(e,f\right)$ changes. In other words, the model assumes that monitoring occurs over months, during which time normal presbycusis is effectively unmeasurable, and not decades, when many accumulated factors unrelated to the exposure of concern can induce hearing changes. An expanded model is described in Bao et al. (2017).

The pure tone sensitivity $\delta\left(e,f\right)$ is measured by the audiogram at test frequencies defined by the clinical protocol. Viewed in this way, the baseline and follow-up audiograms are each an error-susceptible sample from the pure tone sensitivity processes $\delta_0\left(e,f\right)$ and $\delta_t\left(e,f\right)$. For our purposes, the audiogram $\mathbf{Y}$ is comprised of {Left Ear thresholds at 0.5, 1, 2, 3, 4, 6, 8}, {Right Ear thresholds at 0.5, 1, 2, 3, 4, 6, 8} so that $\mathbf{Y}$ has 14 elements. The ordering of ears and frequencies in $\mathbf{Y}$ must be consistent for intelligibility. $\mathbf{Y_0}$ and $\mathbf{Y_t}$ correspond to audiometry at baseline and follow-up. Moving forward, we use $\delta_\mathbf{0}$ and $\delta_\mathbf{t}$ to represent the functions $\delta_0\left(e,f\right)$ and $\delta_t\left(e,f\right)$ evaluated in the ears and pure tone frequencies specified by the testing protocol. The process mean μ is defined the same way.

By virtue of the Gaussian process model, $\mathbf{Y_0}$ and $\mathbf{Y_t}$ are multivariate normal random variables with respective means $\delta_\mathbf{0}$ and $\delta_\mathbf{t}$ and residual covariance matrices $\sum_\mathbf{0}$ and $\sum_\mathbf{t}$:

$$\mathbf{Y_0}|\delta_\mathbf{0} \sim \mathbf{N}\left(\delta_\mathbf{0}, \sum_\mathbf{0}\right) \text{ and}$$

$$\mathbf{Y_t}|\delta_\mathbf{t} \sim \mathbf{N}\left(\delta_\mathbf{t}, \sum_\mathbf{t}\right).$$

The Homeostasis Hypothesis states that pure tone sensitivity has not changed, i.e. that $\delta_\mathbf{t}=\delta_\mathbf{0}$. An important, but often tacit assumption is that variance components, such as $\sum$ and $\mathbf{K}$ (see below) are assumed constant over the monitoring period, so that any contradiction to the Homeostasis Hypothesis is taken as evidence of audiopathic injury and not to changes model variance components. With these assumptions, the joint distribution of the audiograms at baseline and follow-up time points conditional on $\delta_\mathbf{0}$ according to the Homeostasis Hypothesis is

$$\begin{matrix}\mathbf{Y_0}\\ \mathbf{Y_t}\end{matrix}|\delta_\mathbf{0} \sim \mathbf{N}\left(\begin{pmatrix}\delta_\mathbf{0}\\ \delta_\mathbf{0}\end{pmatrix}, \begin{bmatrix}\sum & \mathbf{0}\\ \mathbf{0} & \sum\end{bmatrix}\right). \quad (1)$$

We don't know the baseline pure tone sensitivity $\delta_\mathbf{0}$ so we integrate expression **(1)** with respect to the distribution of $\delta_\mathbf{0}$. This gives the unconditional joint distribution of $\mathbf{Y_t}$ and $\mathbf{Y_0}$ as

$$\begin{matrix}\mathbf{Y_0}\\ \mathbf{Y_t}\end{matrix} \sim \mathbf{N}\left(\begin{pmatrix}\mu\\ \mu\end{pmatrix}, \begin{bmatrix}\sum+\mathbf{K} & \mathbf{K}\\ \mathbf{K}' & \sum+\mathbf{K}\end{bmatrix}\right), \quad (2)$$

where $\mathbf{K}$ is a matrix of evaluations of the covariance function K at the frequencies and ears specified by the testing protocol. The conventional squared exponential covariance model between binary ear indicators (0 = left, 1 = right) e and e^* and $\log_2$ frequencies f and f^* is used for this purpose:

$$K\left(e,\ e^*,f,f^*\right) = \varphi^2 \bullet \exp\left(-\alpha \bullet \left(e-e^*\right)^2 - \beta \bullet \left(f-f^*\right)^2\right),$$

which implies between-ear correlation of $\frac{1}{e^\alpha}$ and between-octave correlation of $\frac{1}{e^\beta}$. We also assume that $\sum$ is a diagonal matrix with constant diagonal elements σ^2. Expression **(2)** is the multivariate form of the "Linear mixed model" (McCulloch and Searle, 2004).

We can think of two uses of expression **(2)** in serial monitoring. First, we can compute the distribution of the difference between the baseline and follow-up audiograms as

$$\left(\mathbf{Y_t}-\mathbf{Y_0}\right) \sim \mathbf{N}\left(\mathbf{0},\ 2 \bullet \sum\right)$$

and then develop prediction regions based on this model. Application of this approach is hampered by missing data or NR at either baseline or follow-up, and regression to the mean is in effect so that unusually large differences are often incorrectly interpreted as evidence of physiological change (Royston, 1995).

Instead, we use the pre-exposure baseline audiogram as an unbiased estimate of pure tone sensitivity in the absence of any audiopathic injury caused by the exposure. Having first observed the baseline audiogram, prior to any exposure, it's natural to think of the follow-up audiogram as a test of stability within the auditory system that generated the baseline audiogram, i.e. the follow-up audiogram is a test that the Homeostasis Hypothesis is correct. This motivates computing the conditional distribution of the follow-up audiogram given the observed baseline under Homeostasis. The multivariate Normal model of $\mathbf{Y_t}$ and $\mathbf{Y_0}$ implies that the conditional distribution $\mathbf{Y_t}|\mathbf{Y_0}$ is also multivariate Normal with expected value

$$\mu + \mathbf{K} \bullet \left(\sum + \mathbf{K}\right)^{-} \bullet (\mathbf{y_0} - \mu) \tag{3}$$

and covariance

$$\left(\sum + \mathbf{K}\right) - \mathbf{K} \bullet \left(\sum + \mathbf{K}\right)^{-} \bullet \mathbf{K}^{\mathbf{T}} \tag{4}$$

McCulloch and Searle (2004) and Rasmussen and Williams (2005).

The goal is to evaluate whether the follow-up audiogram is consistent with expectations given by the baseline audiogram assuming that the Homeostasis Hypothesis is true. We do this by comparing the follow-up audiogram to the multivariate Normal distribution parameterized by expressions **(3)** and **(4)**. To facilitate clinical applications, the predictive distribution is typically distilled into one or more prediction intervals. A follow-up threshold that lies outside the prediction interval is unexpected and worthy of further consideration, either by clinical referral or even ignoring the result. This decision is left to the attending audiologist.

Ninety percent pointwise prediction intervals for the follow-up threshold at each frequency are given by expression **(3)** $\pm$ 1.64 times the square root of the diagonal elements of expression **(4)** (a 95% prediction interval substitutes 1.96 for 1.64 and so forth). The result is a vector of lower and upper 90% prediction limits for each ear and test frequency within which each follow-up threshold is predicted to lie. These 90% pointwise prediction intervals are called "pointwise" because they provide 90% prediction intervals for that specific test ear and frequency "point" only. This distinction is important: A 90% *pointwise* prediction interval for one ear and test frequency is an interval such that 9 in 10 follow-up thresholds in that ear and test frequency will be within the interval. However, we usually want to monitor multiple frequencies in both ears rather than single frequencies in any one ear. A 90% *simultaneous* prediction region is one in which 9 in 10 *audiograms* are completely inside the region. We can't use the pointwise intervals for this purpose because potentially far more than one in 10 follow-ups will yield one *or more* audiometric thresholds outside the 90% pointwise limits if the Homeostasis Hypothesis is correct. Such a naïve application will yield more false-referrals than expected.

We require a 90% simultaneous prediction region for the entire left and right ear audiogram, and not for each ear and frequency individually. Nine in 10 follow-up audiograms should lie entirely within the prediction region if the Homeostasis Hypothesis is correct. Any frequency in either ear with a threshold outside the interval is cause for concern. We define this prediction region following the "volume tube" methodology outlined in Krivobokova et al. (2010), McMillan and Hanson (2014), and Bao et al. (2017). The idea is to numerically expand the width of all the pointwise prediction intervals until exactly 90% of the predicted audiograms are completely contained within the adjusted intervals in both ears and at each test frequency. Let m_j, l_j, and u_j denote the expected value, upper and lower 90% pointwise prediction limits for the j^{th} ear-by-frequency combination. We first simulate a large number of audiograms from multivariate Normal parameterized by expressions **(3)** and **(4)**. A 90% simultaneous prediction region is found by numerically searching for a constant $c > 1$ that adjusts the lower and upper prediction limits at each frequency by $m_j - c\left(m_j - l_j\right)$ and $m_j + c\left(u_j - m_j\right)$ so that 90% of the simulated audiograms completely lie within the adjusted intervals at all frequencies in both ears.

2.1 Estimation

The predictive distribution of the follow-up audiogram given its baseline is given by the parameters in expressions **(3)** and **(4)** and requires as inputs the age-sex specific population mean audiogram μ, the baseline audiogram y_0 and estimates of σ, φ, α, and β. The population mean thresholds for men and women are taken from ISO 1999-2013 (ISO, 2013). We use these population mean estimates to center the distribution of pure tone sensitivities, though the model allows for considerable variation with respect to the population. Unless there are NR thresholds in the baseline response, the model parameters are easily estimated by maximizing the marginal likelihood in expression **(2)** (Rasmussen and Williams, 2005). We prefer a Bayesian approach so as to easily propagate uncertainty about the parameter estimates into the predictions. This is done through MCMC evaluation of the joint posterior distribution of the model parameters and using those same MCMC evaluations to compute the predictive distribution of $\mathbf{Y_t}$ given $\mathbf{Y_0}$. These predictions are then used in the volume-tube methodology for computing prediction regions for the entire audiogram.

A pure tone threshold that exceeds the audiometer's test limits is called "No Response" (NR) in audiology and more generally called "Right-Censored" at the detection limit d in statistics. This feature is commonly observed in time-to-event data such as patient survivorship in biomedical research or equipment reliability in manufacturing. There are several approaches to handling NR thresholds in hearing research, such as imputing the NR threshold to d plus 5 dB or some other constant. Another approach is to treat the NR measurement as completely missing. Neither of these approaches is appealing because imputation by adding an arbitrary constant implies an observation (the NR limit + 5 dB) that was never made, which implies greater certainty about pure tone thresholds than the audiologist can legitimately claim. This will increase the false-referral rate beyond the nominal levels

TABLE 1 Priors on the parameters and the induced parameters of the proposed model.

Parameter	Prior quantiles		
	5%	50%	95%
σ	2.3	6.0	12.7
$E(\lvert\text{test-retest}\rvert)$ in dB $= \frac{2\sigma}{\sqrt{\pi}}$	2.5	6.8	14.3
φ	1.9	28.3	122.9
$E(\lvert\text{max-min sensitivity}\rvert)$ in dB $= \frac{2\varphi}{\sqrt{\pi}}$	2.2	31.9	138.7
α	0.03	0.34	0.98
Correlation between ears$= \frac{1}{e^{\alpha}}$	0.38	0.71	0.97
β	0.06	0.68	1.95
Correlation between octaves$= \frac{1}{e^{\beta}}$	0.14	0.51	0.94

dictated by the monitoring protocol. Conversely, treating the NR measurement as completely missing isn't a valid approach either, since the audiologist knows that the pure tone threshold exceeds the detection limit d. Thresholds that exceed the audiometer detection limit provide valuable information for making accurate inferences about $\mathbf{K}$ and $\sum$ so that more accurate predictions about the follow-up audiogram can be made.

We approach NR thresholds using censored-data models. Expression **(1)** represents the likelihood σ, φ, α, and β conditional on δ. Without creating additional notation, the Gaussian Process model for $\delta\left(e,f\right)$ implies that δ is also a multivariate normal random variable, $\delta \sim \mathbf{N}(\mu, \mathbf{K})$. In the absence of any NR thresholds, we eliminate dependence on δ by marginalizing the likelihood in **(1)** giving expression **(2)**. However, when there are one or more NR in the audiogram we factor the likelihood into scalar contributions from thresholds that we observe as $N\left(y;\delta,\sigma^2\right)$ and into scalar contributions from NR thresholds as $1-\Phi\left(d;\delta,\sigma^2\right)$. This latter expression is one minus the Normal cumulative distribution function evaluated at the audiometer's detection limit d.

There is no closed form integral of this factored likelihood with respect to the distribution of δ (Ertin, 2007) meaning that the simplicity achieved with a complete baseline audiogram is lost. However, we can use MCMC to evaluate the joint distribution of δ and the parameters σ, φ, α, and β conditional on the baseline audiogram. Each of these MCMC evaluations generate a predicted follow-up audiogram according to expressions **(3)** and **(4)**. The 90% pointwise prediction interval are the 5th and 95th percentiles of the generated predictions at each frequency and ear. The volume tube methodology is applied to these predictive distributions to achieve 90% prediction regions over the entire audiogram. The result is a shaded region (Figures 1, 3–5) that expresses the clinical expectation that 9 in 10 follow-up audiograms will lie completely within the shaded region if the Homeostasis Hypothesis is correct.

The width of the interval can be changed, depending on the clinical application. Chemotherapy monitoring may demand a very low false referral rate so as not to withhold life-saving anti-cancer therapy. Larger apparent changes are admissible before alerting the audiologist to deleterious side effects of the therapy. A 95% reference interval may be preferable in this instance instead of the 90% intervals used throughout this paper. Workplace noise damage monitoring may prefer a higher false-referral rate to avoid financial liability. Smaller threshold changes in the noise exposure context will therefore provoke a response from administrators, so that an 80% reference interval may be preferred. These considerations illustrate the relationship between the consequences of a false-referral and the desired width of the reference interval. If false-referrals provoke little harm, then a narrower interval is acceptable, but if the ramifications of a false-referral are serious, then wider reference intervals are desirable. Our approach provides the user complete control over the nominal false-referral rate.

2.2 Priors

We establish priors on α and β by recalling that the between-ear correlation is $\frac{1}{e^{\alpha}}$ and between-octave correlation is $\frac{1}{e^{\beta}}$. We believe that the between-ear correlation in pure tone sensitivity is likely to be >0.5. We also believe that the between-octave correlation is likely to be >0.5, but we admit much greater uncertainty since this feature may vary widely among patients. These requirements suggest to us that $\alpha \sim Half-Normal(0.5^2)$ and $\beta \sim Half-Normal(1^2)$. We also take advantage of the fact that of $\mathbf{Y_t}$, $\mathbf{Y_0}$, and $\delta_\mathbf{0}$ are multivariate normal random variables, such that the expected value of the absolute difference over time between any two corresponding elements of $\mathbf{Y_t}$ and $\mathbf{Y_0}$ is $\frac{2\sigma}{\sqrt{\pi}}$ and between any two pure tone sensitivities across ears and frequencies on the same person is $\frac{2\varphi}{\sqrt{\pi}}$. Average absolute test-retest differences are expected below about 15 dB and average between 5 and 10 dB. This suggests the prior $\sigma \sim Gamma(4, 0.6)$, which has expected value $\frac{4}{0.6}$ and variance $\frac{4}{0.6^2}$. The range of pure tone sensitivities across frequencies is expected to vary markedly among patients, though we expect no more than about 135 dB range among pure tone sensitivities within a patient. We assume the prior $\varphi \sim Gamma(1, 0.025)$, which is parameterized as for the prior on σ. Summary statistics for each of these priors, as well as the induced priors on the correlations and test-retest differences are shown in Table 1. Prior histograms are shown in Figure 2, along with posterior distributions for the patient shown in Figure 1.

2.3 Computation

The sampler is started at the posterior means from the model in expression **(1)**, initializing NR thresholds at the test limit **d**. These initial values are fed into a new MCMC sampler replacing expression **(2)** with expression **(1)**. We find it sufficient to run the MCMC sampler for 500,000 iterations using SAS Enterprise Guide Software, v. 8.3, PROC MCMC, though visual confirmation of efficient mixing is advisable, particularly for unusual audiograms having, for example, elevated left-right asymmetry or many NR thresholds.

3 Results

Figures 3–5 illustrate model results in the context of additional case studies, following the format of Figure 1. The prediction region

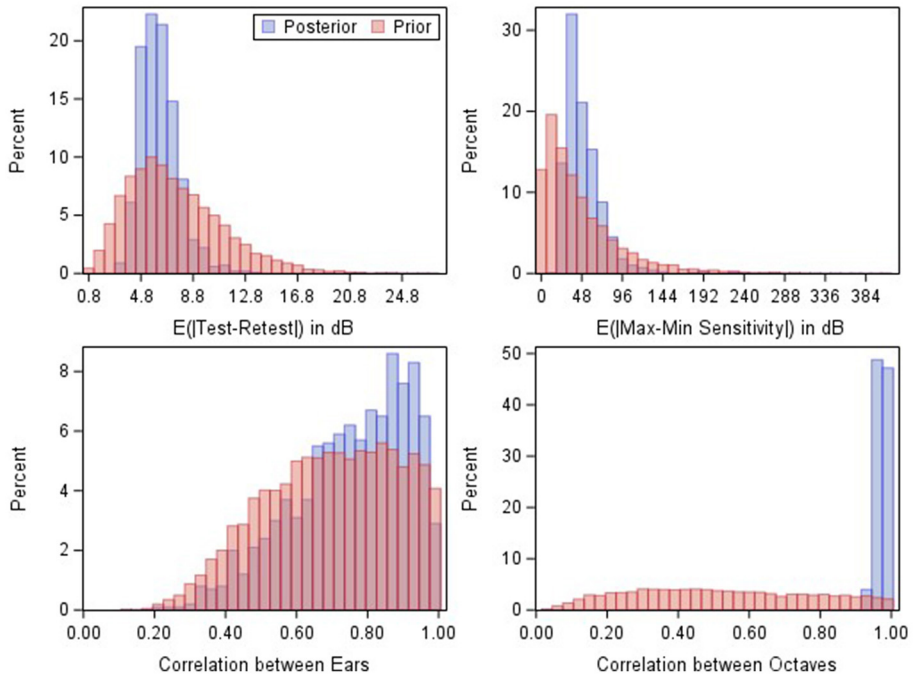


FIGURE 2
Prior (red) and posterior (blue) MCMC samples for the induced parameters in the model fit to the baseline data shown in Figure 1.

shown for these patients were generated by inputting the baseline audiogram thresholds, age, and sex into expressions **(3)** and **(4)**, and following the Volume Tube methodology. Figure 3 shows results for a patient with Cystic Fibrosis who was treated with IV Tobramycin for a bacterial lung infection. Figure 4 shows results for a patient with cancer who was treated with cisplatin, and Figure 5 shows results for an individual exposed to workplace noise over a five-year period. Note that this subject did not provide baseline 3 and 6 kHz thresholds, though the model structure still allows predictions at these frequencies.

4 Discussion

In this paper we describe a Gaussian process regression model of the audiogram that is suitable for serial monitoring in clinical and industrial applications. Additional applications suitably addressed with our approach include monitoring patients for improvements in hearing, for example following surgical intervention such as ossicular reconstruction. The innovative aspects of our approach are three-fold. First, it uses a patient's baseline hearing, known correlations among test frequencies and ears, together with population-based hearing data, to calculate an *individualized* prediction region for that patient. Second, it provides a unified framework for monitoring the audiogram that is much more intuitive than the various shift criteria commonly used in clinical practice. The automated audiogram region estimated using our approach is simply the region where the follow-up audiogram is predicted to land if that patient's hearing has remained stable. Follow-up thresholds that exceed the predicted region at ANY audiometric frequency can be interpreted as evidence for a statistically significant hearing change. Third, our approach overcomes the problem of regression to the mean, which is a nearly ubiquitous but largely overlooked problem in serial monitoring. The flexibility and ease of interpretation of this model allows for the implementation of the criteria directly into audiometers and other computerized hearing testing platforms, increasing the potential user base and uptake of serial monitoring across contexts.

4.1 Limitations

Our proposal doesn't include any explicit model training commonly used in prediction algorithms. We embed information about the population into the informative priors on the model parameters. An expanded approach is to further train the model in a large sample to identify the joint distribution of model parameters. Training must be done in a population for which the Homeostasis Hypothesis is unequivocally true. Furthermore, this is computationally challenging because of the factored likelihood in the presence of NR thresholds. Training model parameters is the subject of ongoing work by our research group.

Our approach mitigates some of the difficulty of NR thresholds in serial monitoring, though it cannot solve the problem entirely. We can generate prediction regions in the presence of baseline NR, however, any NR observed during follow-up measurements can create difficulties. These are illustrated in Figure 1. The baseline, left-ear, 8 kHz threshold is NR, but our methodology still allows one to identify the prediction region for follow-up thresholds at that frequency and ear. The left-ear, 3 kHz threshold is NR at follow-up, which is outside the expectations established by proposed

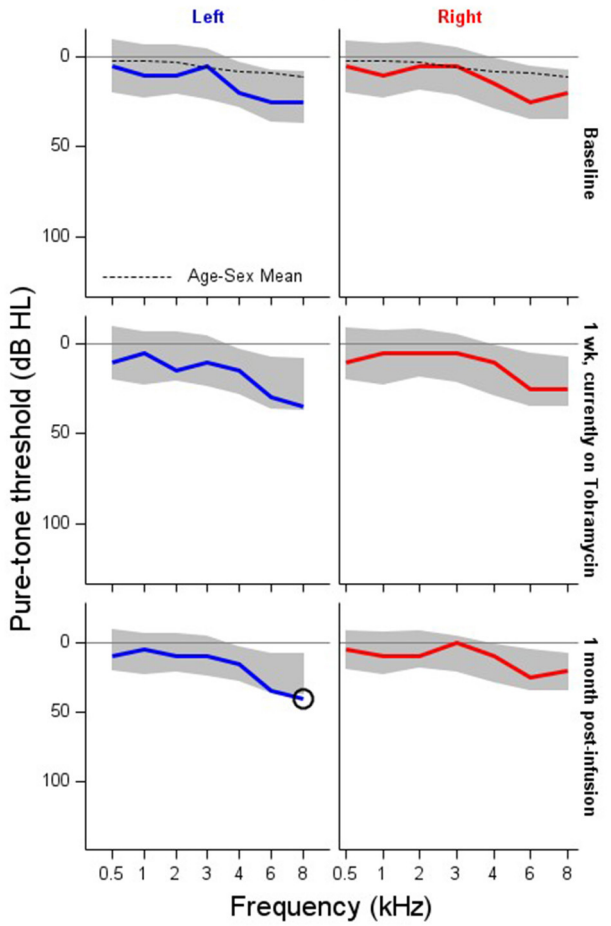


FIGURE 3
Serial monitoring of a 42 year-old male with a history of pancreatic insufficiency and diabetes mellitus related to his Cystic Fibrosis. The patient presented to emergency unit with a pulmonary infection and was admitted to the hospital where they received between 6.75 to 12.5 ml of IV tobramycin. Visit A: baseline audiogram. Visit B: audiogram taken 1 week after the baseline exam during their course on IV tobramycin therapy as an inpatient. Visit C: first follow-up visit conducted about 1 month after end of IV course of treatment. X = no response, NR. Circled thresholds are significant changes.

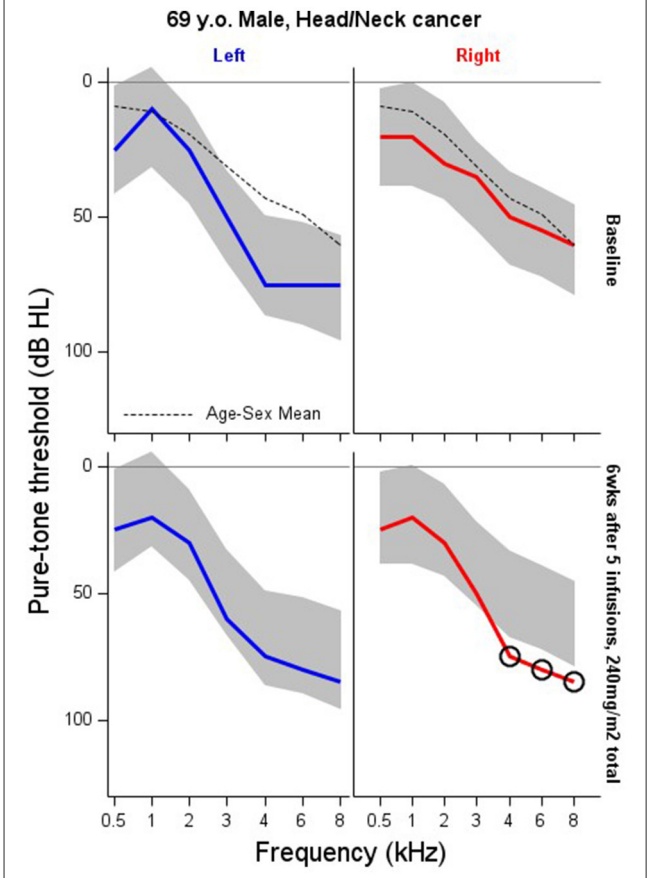


FIGURE 4
Serial monitoring of a 66 year old patient with a tumor of the suraglottis larynx who received cisplatin-based chemotherapy prior to surgical removal. The patient received five infusions of cisplatin for a total of 240 mg/m^2. X = no response, NR. Circled thresholds are significant changes.

methodology. In these instances our approach is handling the NR measurement without any trouble as expected. Difficulties arise when the prediction region "straddles" the NR level such as left-ear, 4, 6, and 8 kHz. The observed NR are consistent with the prediction region that spans the test limit, so that no violation of the Homeostasis Hypothesis is observed. However, this isn't exactly true: an NR threshold may actually be outside the prediction region, but the test limit doesn't permit the audiologist to observe this. There is thus some degree of uncertainty one has to accept in these instances.

Although we have developed and described this approach to address the clinical challenge of determining when audiopathic damage has occurred for an adult patient or worker, the framework is easily extended to pediatric applications as long as suitable priors for this population can be identified. This methodology is also easily generalized to "objective measures" of auditory sensitivity that can be obtained reliably in infants and young children. Otoacoustic emissions are an attractive measure to use due to their sensitivity to noise and ototoxic exposures (Dreisbach et al., 2023) and the large literature of test-retest data in unexposed young controls (Bao et al., 2017; Konrad-Martin et al., 2020). Digital audiometry platforms to determine what constitutes a statistically significant hearing change *for that patient,* will also provide important efficiencies for future clinical trials.

5 Conclusions

Audiogram forecasting such as described in this paper can substantially improve serial monitoring over traditional approaches. Our method avoids sources of bias that reduce diagnostic accuracy and standardizes the definition of a "significant hearing change". This has the added benefit of leaving clinical interpretations about the functional impacts, implications for follow-up, and treatment options up to the treating audiologist and other clinical stakeholders.

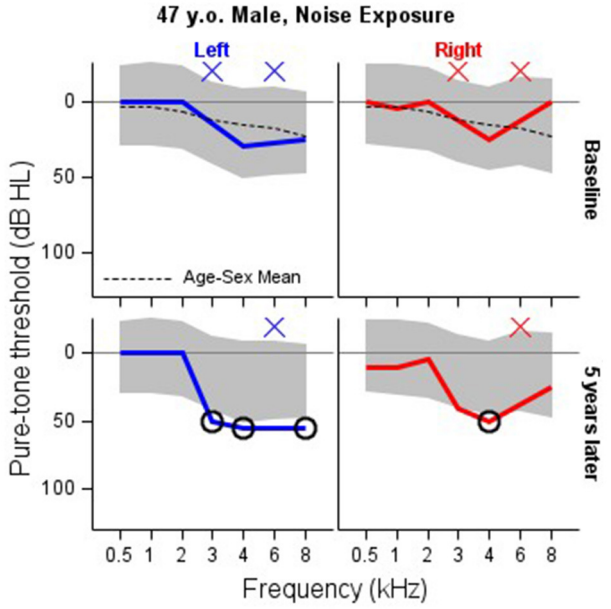


FIGURE 5
Serial monitoring of a 47 year old patient who worked as a chemical engineer and reported exposure to loud machine noise. His initial audiogram was completed at age 47, with a follow up occurring at age 52. Baseline thresholds are missing at 3 and 6 kHz. 3 kHz thresholds are missing at five-year followup. X = no response, NR. Circled thresholds are significant changes.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Oregon Health and Science University/VA Portland Joint IRB. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

GM: Conceptualization, Data curation, Formal analysis, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. JD: Conceptualization, Data curation, Investigation, Methodology, Project administration, Supervision, Visualization, Writing – original draft, Writing – review & editing. MH: Data curation, Investigation, Project administration, Visualization, Writing – review & editing. DK-M: Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Resources, Supervision, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This material was the result of work supported with resources and the use of facilities at the VA Rehabilitation Research and Development (RR&D) National Center for Rehabilitative Auditory Research (NCRAR) [Center Award #C2361C/I50 RX002361] at the VA Portland Health Care System in Portland, Oregon and through funding from a VA RR&D Merit Review Award to DK-M [#C3127R/ I01 RX003127].

Acknowledgments

Edward J. Bedrick provided valuable comments that enhanced the manuscript.

Conflict of interest

DK-M is listed as a co-inventor on a patent for a portable hearing test and testing device.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Author disclaimer

The views expressed in this article are those of the authors and do not necessarily reflect the position or policy of the US Veterans Health Administration or the United States Government.

References

Bao, J., Hanson, T., McMillan, G. P., and Knight, K. (2017). Assessment of DPOAE test-retest difference curves via hierarchical Gaussian processes. *Biometrics* 73, 334–343. doi: 10.1111/biom.12550

Barbour, D. L., Howard, R. T., Song, X. D., Metzger, N., Sukesan, K. A., DiLorenzo, J. C., et al. (2019). Online machine learning audiometry. *Ear Hear.* 40, 918–926. doi: 10.1097/AUD.0000000000000669

Dreisbach, L., Konrad-Martin, D., Gagner, C., Reavis, K. M., and Jacobs, P. G. (2023). Descriptive characterization of high-frequency distortion product otoacoustic emission source components in children. *J. Speech Lang. Hear. Res.* 66, 1–17. doi: 10.1044/2023_JSLHR-23-00013

Ertin, E. (2007). "Gaussian process models for censored sensor readings," in *2007 IEEE/SP 14th Workshop on Statistical Signal Processing* (Madison, WI: IEEE), 665–669. Available online at: http://ieeexplore.ieee.org/document/4301342/ (accessed December 26, 2023).

ISO (2013).14:00–17.00. ISO 1999:2013. Available online at: https://www.iso.org/standard/45103.html (accessed January 8, 2024).

King, K. A., and Brewer, C. C. (2018). Clinical trials, ototoxicity grading scales and the audiologist's role in therapeutic decision making. *Int. J. Audiol.* 57, S89–S98. doi: 10.1080/14992027.2017.1417644

Konrad-Martin, D., Knight, K., McMillan, G. P., Dreisbach, L. E., Nelson, E., Dille, M., et al. (2020). Long-term variability of distortion-product otoacoustic emissions in infants and children and its relation to pediatric ototoxicity monitoring. *Ear Hear.* 41:239. doi: 10.1097/AUD.0000000000000536

Krivobokova, T., Kneib, T., and Claeskens, G. (2010). Simultaneous confidence bands for penalized spline estimators. *J. Am. Stat. Assoc.* 105, 852–863. doi: 10.1198/jasa.2010.tm09165

McCulloch, C. E., and Searle, S. R. (2004). *Generalized, Linear, and Mixed Models*. Hoboken, NJ: John Wiley and Sons, 358.

McMillan, G. P., and Hanson, T. E. (2014). Sample size requirements for establishing clinical test-retest standards. *Ear Hear.* 35, 283–286. doi: 10.1097/01.aud.0000438377.15003.6b

Moore, B. C. J., Lowe, D. A., and Cox, G. (2022). Guidelines for diagnosing and quantifying noise-induced hearing loss. *Trends Hear.* 26:23312165221093156. doi: 10.1177/23312165221093156

Rasmussen, C. E., and Williams, C. K. I. (2005). *Gaussian Processes for Machine Learning*. Cambridge, MA: MIT Press, 266. doi: 10.7551/mitpress/3206.001.0001

Royston, P. (1995). Calculation of unconditional and conditional reference intervals for foetal size and growth from longitudinal measurements. *Stat Med.* 14, 1417–1436. doi: 10.1002/sim.4780141303

Samuels, M. L. (1991). Statistical reversion toward the mean: more universal than regression toward the mean. *Am. Stat.* 45, 344–346. doi: 10.2307/2684474

Song, X. D., Wallace, B. M., Gardner, J. R., Ledbetter, N. M., Weinberger, K. Q., Barbour, D. L., et al. (2015). Fast, continuous audiogram estimation using machine learning. *Ear Hear.* 36, e326–e335. doi: 10.1097/AUD.0000000000000186

frontiers | Frontiers in Audiology and Otology

TYPE Original Research
PUBLISHED 09 October 2024
DOI 10.3389/fauot.2024.1397822

Check for updates

OPEN ACCESS

EDITED BY
Norbert Dillier,
University of Zurich, Switzerland

REVIEWED BY
Niels Henrik Pontoppidan,
Eriksholm Research Centre, Denmark
Razan Alfakir,
Auburn University, United States

*CORRESPONDENCE
De Wet Swanepoel
✉ dewet.swanepoel@up.ac.za

RECEIVED 08 March 2024
ACCEPTED 16 September 2024
PUBLISHED 09 October 2024

CITATION
Fourie C, Mahomed-Asmail F, Oosthuizen I, Manchaiah V, Vercammen C and Swanepoel DW (2024) Hearing aid benefit in daily life: a qualitative ecological momentary assessment study.
Front. Audiol. Otol. 2:1397822.
doi: 10.3389/fauot.2024.1397822

COPYRIGHT
© 2024 Fourie, Mahomed-Asmail, Oosthuizen, Manchaiah, Vercammen and Swanepoel. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Hearing aid benefit in daily life: a qualitative ecological momentary assessment study

Chané Fourie[1], Faheema Mahomed-Asmail[1,2], Ilze Oosthuizen[1,2], Vinaya Manchaiah[1,2,3,4,5], Charlotte Vercammen[6,7,8] and De Wet Swanepoel[1,2,3]*

[1]Department of Speech-Language Pathology and Audiology, University of Pretoria, Pretoria, South Africa, [2]Virtual Hearing Lab, Collaborative Initiative Between University of Colorado School of Medicine, Aurora, CO, United States, [3]Department of Otolaryngology-Head and Neck Surgery, University of Colorado School of Medicine, Aurora, CO, United States, [4]UCHealth Hearing and Balance, University of Colorado Hospital, Aurora, CO, United States, [5]Department of Speech and Hearing, School of Allied Health Sciences, Manipal Academy of Higher Education, Manipal, India, [6]Sonova AG, Research & Development, Stäfa, Switzerland, [7]Manchester Centre for Audiology and Deafness, School of Health Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, United Kingdom, [8]Department of Neurosciences, Research Group Experimental Oto-Rhino-Laryngology, KU Leuven-University of Leuven, Leuven, Belgium

Introduction: Understanding hearing aid wearer experiences in real-world settings is important to provide responsive and individualized hearing care. This study aimed to describe real-life benefits of hearing aids (HAs), as reported by hearing aid wearers through Ecological Momentary Assessment (EMA) in various listening environments.

Method: Qualitative content analysis of 1,209 open-text responses, provided through self-initiated EMAs, was conducted. The de-identified data was collected retrospectively via a smartphone app compatible with these HAs. Only text responses reflecting positive hearing aid experiences were analyzed.The 1,209 open-text responses were categorized into 18 pre-determined sub-categories, further organized into five overarching categories: Conversational, Leisure, Device-related aspects, Lifestyle, and Other factors.

Results: Across these categories, 48 self-generated meaning units highlighted the multifaceted benefits of HAs. In particular, participants reported significant improvements in conversational settings, specifically during phone conversations and meetings, attributed to improved sound quality and speech understanding when wearing their HAs. During leisure activities, particularly TV watching and music listening, clearer sound and ease of Bluetooth streaming contributed to experienced benefits. Lifestyle improvements were reported in occupational and social settings, as hearing aid wearers stated enhanced communication and sound awareness. Device-related factors contributing to positive wearer experiences included extended battery life and the convenience of rechargeable batteries. The most prominent sub-category, other factors, underscored overall satisfaction, comfort with the device, and improved auditory experiences across various environments.

Conclusion: This study reveals the diverse benefits of HAs in improving communication, listening experiences, and quality of life across various settings, as captured through EMA. By emphasizing features like direct streaming and rechargeability, the findings highlight the importance of personalized hearing care and the potential of real-time listener feedback to inform device

enhancements and support strategies, advancing more tailored and effective hearing rehabilitation.

KEYWORDS

hearing aids, hearing aid outcome, Ecological Momentary Assessment (EMA), everyday situations, real-life data

1 Introduction

Hearing rehabilitation aims to enhance hearing functioning, participation, and quality of life for individuals with hearing loss (Boothroyd, 2007). Providing amplification through hearing aids (HAs) is a primary component of hearing care, as HAs amplify sound and improve the clarity of sounds, with a specific emphasis on speech and communication (Ferguson et al., 2017). However, the effectiveness of HAs depends on wearers being familiar with the correct way to handle the devices, understanding the expected benefits and satisfaction, and the clinicians' ability to create personalized and achievable rehabilitation plans with the HAs (Heselton et al., 2022; Humes, 2003; Wong et al., 2003). A recent study that included former HA wearers revealed that almost half of them attributed non-use to device-related issues such as wearing comfort, not liking to wear the HAs, or limited perceived benefit (Franks and Timmer, 2023; Mothemela et al., 2023). It is therefore important to explore the factors contributing to HA wearers' experiences, in order to enhance understanding of HA benefit and satisfaction from the wearer's perspective, provide person-centered care, and validate the effectiveness of treatment with HAs.

Satisfaction is commonly measured through self-reported measures, often referred to as patient-reported outcome measures (PROMs) (Timmer et al., 2018; Oosthuizen et al., 2022). Self-reported questionnaires can, for instance, gather information about the individual's perspective on how well rehabilitation goals have been achieved in real-world settings (von Gablenz et al., 2021). Some limitations of PROMs include that they necessitate client input based on memory and experiences with specific listening conditions. As memory affects recall, this can introduce recall bias. Also, listening situations posed in PROMs might not apply to all HA wearers (Timmer et al., 2018). Ecological momentary assessment (EMA) has been proposed to address some limitations of PROMs, by asking participants to repeatedly report on their experiences during or close in time to an event of interest (Bolger et al., 2003; Shiffman et al., 2008). EMA, also known as ambulatory assessment or experience sampling (Trull and Ebner-Priemer, 2014), is a real-time data collection method, applied in participants' real-world environments. EMA allows the capturing of individuals' daily experiences and changes in their experiences over time (Holube et al., 2020).

In practical terms, EMA research today commonly employs personal digital devices, utilizing auditory or vibratory alerts to prompt participants to respond to a series of questions throughout the day. This prompted EMA approach involves participants receiving messages on their smartphone-based app at regular intervals to complete surveys (Burke et al., 2017). Patients may not always comply with the EMA data collections, as the highly dense data collection can burden participants. This limitation arises when patients do not provide feedback when prompted, thereby restricting the coverage of the analysis. This could lead to inadequate results, as they might fail to accurately portray the diverse range of experiences people have (Holube et al., 2020; Schinkel-Bielefeld et al., 2020). An alternative approach to prompted EMA is self-initiated EMA. During self-initiated EMA, the individuals decide when something of interest has happened, and subsequently fill in a survey on their initiative, without any prompting (Schinkel-Bielefeld et al., 2020).

In addition to being used in research, EMA is also proposed as a valuable clinical tool, enabling patients to monitor daily challenges systematically and contributing to personalized hearing healthcare (Schinkel-Bielefeld et al., 2020). Data collected through EMA could guide healthcare professionals in tailoring HA settings to meet individual patient needs during fitting, fine-tuning, and acclimation (Holube et al., 2020). Combined with a person-centered care approach, such data can enhance understanding, leading to improved health outcomes. These rich data collection methods could also facilitate improved communication between patients and healthcare professionals, particularly concerning patient-specific needs and residual hearing difficulties experience in real life situations (Wu et al., 2015; Brice and Almond, 2022).

Most studies on hearing-related EMA have predominantly utilized quantitative methodologies to capture experiences with HAs and their features (Galvez et al., 2012; Hasan et al., 2014; Timmer et al., 2017; Wu et al., 2018). However, there remains a notable scarcity in the application of EMA for gathering qualitative data, such as personal experiences with HAs in real-life scenarios. Notably, Galvez et al. (2012) undertook a qualitative analysis of prompted EMA data to explore hearing difficulties among HA wearers, providing valuable insights for evaluating HA parameters and enhancing patient care. Similarly, Vercammen et al. (2023) identified key themes in feedback from HA wearers using automated text analysis of self-initiated EMA data, revealing a predominance of positive experiences related to communication and sound quality, in contrast to challenges in HA management.

While recent technological advancements, such as real-time speech-to-text transcription and advanced natural language processing (NLP) techniques, provide unprecedented opportunities for capturing and analyzing qualitative wearer feedback (Manchaiah et al., 2021a,b), their full potential has yet to be explored. To complement insights from such computational methods, this study employs a manual qualitative analysis to delve into a portion of the dataset previously investigated by Vercammen et al. (2023) using NLP techniques. This methodological decision is intentional, addressing the constraints of computational methods, which may lack the depth and nuanced understanding inherent in direct human analysis (Jiang et al., 2021; Baden et al., 2022).

Employing this approach, our study seeks to gain a more thorough understanding of the qualitative EMA data, with a particular emphasis on identifying psychosocial elements of satisfaction with HA usage (Oosthuizen et al., 2022; Knoetze et al., 2023). To this end, we focused on a subset of the original dataset, focusing on positive HA experiences only. Thereby providing a unique perspective, as opposed to the commonly reported challenges with HAs.

2 Materials and methods

2.1 Study design

The study considered a retrospective subset of the data presented in Vercammen et al. (2023) i.e., 1,209 positive open-text statements provided by real-world HA wearers as part of their hearing care. Due to size of the dataset, this manuscript focused on positive responses only, with negative responses being analyzed as part of an upcoming manuscript. Prior to participation, participants were informed of de-identified data analysis for clinical and research purposes per the mobile application's data privacy notice. In addition, no personal identifying information was logged to ensure participant privacy. Institutional Review Board clearance was granted (HUM023/0922) prior to data analysis.

2.2 Study participants, material, and apparatus

The de-identified data was collected through a smartphone mobile application compatible with commercially available HAs, fitted to real-world HA wearers from English-speaking countries, i.e., Australia, Canada, England, Ireland, New Zealand, and the United States. Clinicians activated the EMA application feature within the fitting software during consultations. Clinicians could activate the feature to use as a real-time feedback system whenever they deemed it advantageous for the HA wearer and their hearing care (Vercammen et al., 2023). Participants initiated the mobile application on their own when they had a listening experience that they wanted to report (i.e., self-initiated EMA) and navigated through the windows (see Supplementary material 1) (a) indicate the listening experience as positive or negative; (b) select the listening situation from the list, closest to the experienced situation (i.e., activities, battery or charging, entertainment, hearing children, in meeting, in restaurant, in vehicle, listening music, other, phone conversations, playing games, quiet conversations, shopping, social activities, social event, streaming media, worship, and watching TV); (c) provide description of the listening experience (open text field).

2.3 Data extraction and data cleaning

Between May 2018 and June 2021, an initial sample of 30,127 self-initiated EMAs on real-world HA experiences were collected worldwide and extracted from cloud-based data logging of the smartphone mobile application. Text statements shorter than 20 characters were removed for content quality, and a manual data cleaning process was conducted to correct spelling mistakes, remove nonsense text, and exclude non-English entries. Following data extraction and cleaning, a dataset of 5,331 negative and 3,462 positive EMAs (a total of 8,793 responses) was extracted. Only the 3,462 positive EMA's were considered for further analysis in this study (see Figure 1). During data familiarization, it was found that some of the comments under each pre-determined sub-category (which were derived from the listening situation self-selected by the user—see Supplementary material 1, panel B) were unrelated to the specific situation the participant had chosen (i.e., they were more applicable to another pre-determined sub-category). In addition, some of the comments were negative despite the participant's choice of a positive experience. After a discussion with the research team (IO, FMA, VM, CV, and DWS) a consensus was reached, and 799 open text statements were reclassified (i.e., moved to a more applicable sub-category), and 2,150 comments were moved from positive experiences to negative experiences. Furthermore, 103 comments could not be coded as they were irrelevant to any pre-determined sub-category. A final sample of 1,209 positive self-initiated EMAs was considered for further analysis (see Figure 1).

2.4 Data analysis

The final 1,209 positive self-initiated EMAs were analyzed using qualitative content analysis (Graneheim and Lundman, 2004; Knudsen et al., 2012). This approach was deemed suitable due to the diverse range of responses obtained from the open-ended question, which varied significantly in depth and detail. Qualitative content analysis involves an iterative process of revisiting and refining coding and categorization, facilitating a nuanced and comprehensive understanding of the data. Initially, responses within each of the 18 pre-determined subcategories, derived from the users' selected listening situations in the app (e.g., phone conversations, watching TV etc.), were reviewed, coded and condensed into meaning units, capturing the essence of each participant's experience (Graneheim and Lundman, 2004). These condensed meaning units were further examined and the sub-categories were grouped into broader categories, resulting in five main categories: Conversational, leisure, device-related, lifestyle, and other factors. To ensure clarity, we have detailed the hierarchical classification process used in our data analysis. Seemingly similar responses were placed in distinct categories based on their contextual relevance and the nuances of participant comments. This approach aimed to capture the multifaceted nature of hearing aid experiences. However, we acknowledge the potential overlap of similar responses in different categories and have included this consideration in Section 4.1.

To ensure consistency and reliability, an experienced qualitative researcher cross-checked 50% of the coding, and any discrepancies were resolved through team discussions. This iterative review process allowed for refining the categories, ensuring they accurately represented the data. The final categorization facilitated the identification of 48 self-generated meaning units, providing a comprehensive understanding of the diverse benefits of hearing aids as experienced by participants in their daily lives.

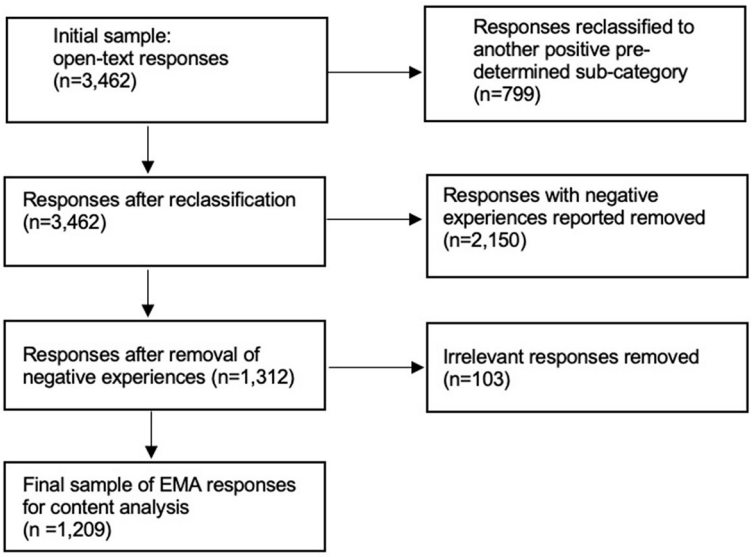


FIGURE 1
Illustration of the manual data cleaning process, leading to the final cample of 1,209 EMA responses that were considered for manual qualitative analysis.

This detailed categorization process ensured that the diverse and nuanced experiences of hearing aid users were systematically captured and analyzed, yielding robust insights into the real-world benefits of HAs.

3 Results

The 18 pre-determined sub-categories were categorized into five categories as part of the manual qualitative content analysis: (1) Conversational, representing diverse settings where participants engaged in conversations; (2) Leisure settings, representing various recreational pursuits for enjoyment and relaxation; (3) Device-related aspects, encompassing different facets of HA functionality; (4) Lifestyle factors, covering settings that contribute to an individual's way of life and daily routines; and (5) Other factors, including settings not falling within the aforementioned domains. From these categories, a total of 48 meaning units were self-generated (see Tables 1–5).

Category 1: Conversational setting benefits of HAs

Six pre-determined sub-categories i.e., phone conversations (n = 86), quiet conversations (n = 64), in vehicle (n = 59), in meeting (n = 45), in restaurant (n = 43), and hearing children (n = 35) were combined into the overarching category "conversational settings" (see Figure 2). Twenty self-generated meaning units were identified (see Table 1 and Supplementary material 2 for elaboration on meaning units). Within phone conversations, the largest identified meaning unit of listening experience (n = 32) focused on aspects such as volume and general satisfaction. The direct streaming capability facilitated seamless call handling, introducing a private and hands-free dimension to phone conversations. In quieter settings, participants enjoyed their ability to engage in one-on-one conversations, emphasizing the significance of both sound quality and improved speech understanding. In vehicle settings, improved sound quality, ease of conversing with passengers, and greater enjoyment of music and audio contributed to enhanced listening experiences. Participants also experienced increased participation and improved speech understanding during group conversations such as meetings and in noisy restaurant environments. Furthermore, participants reported improved communication with children, highlighting enhanced speech understanding and sound quality.

Category 2: Leisure activity benefits of HAs

Four pre-determined sub-categories i.e., watching TV (n = 227), listening music (n = 60), entertainment (n = 6), and playing games (n = 5) were combined into the overarching category "leisure settings" (see Figure 2). Nine self-generated meaning units were identified (see Table 2 and Supplementary material 3 for elaboration on meaning units). Watching TV was the most frequently selected predetermined subcategory with listening experience identified as the largest meaning unit (n = 143) followed by sound quality (n = 46), speech understanding (n = 20), and direct screening (n = 18). These indicate the primary perceived benefits for the HA wearers in this specific leisure activity. Some participants also reported enhanced music listening experiences, noting improved recognition of lyrics, and vibrant sound quality. Greater enjoyment of other media, such as improved listening to radio, gaming, and podcast streaming, was also mentioned.

Category 3: Lifestyle-related benefits of HAs

Five pre-determined sub-categories i.e., activities (n = 64), social event (n = 41), social activities (n = 18), shopping (n = 11), and worship (n = 6) were combined into the overarching category "lifestyle" (see Figure 2). Ten self-generated meaning units were identified (see Table 3 and Supplementary material 4 for elaboration on meaning units). Within Activities, the largest

TABLE 1 Positive experiences in conversational settings ($n = 332$ meaning units).

Sub-category (PD)	Meaning unit (SG)	Meaning unit example
Phone conversations ($n = 86$)	Listening experience ($n = 32$)	*"…telephone use has improved every day!"*
	Direct streaming ($n = 27$)	*"I was able to answer a call by just touching my hearing aid."*
	Sound quality ($n = 19$)	*"I had a phone call with a friend while I was a passenger in the car and the call was very clear and audible"*
	Speech understanding ($n = 8$)	*"I couldn't make out words on my phone without speaker now I can what a difference"*
Quiet conversations ($n = 64$)	General benefit ($n = 51$)	*"…I can hear a quiet conversation without me trying hear out what others was saying…"*
	Sound quality ($n = 9$)	*"Excellent clarity of speech"*
	Speech understanding ($n = 4$)	*"I can understand what they are saying"*
In vehicle ($n = 59$)	Listening experience ($n = 29$)	*"…I could hear the warning beepers without having their volume increased."*
	Conversation ($n = 14$)	*"I could easily converse with others in my car"*
	Entertainment ($n = 12$)	*"Could hear the lyrics on music on car radio"*
	Sound quality ($n = 4$)	*"…the sound quality was very clear. I like the fact that the road noise is limited plus the quality and volume of the sound was very good!"*
In meeting ($n = 45$)	Meeting experience ($n = 17$)	*"I was sitting in a fairly large room with a dozen or so people seated throughout the room, for a meeting, and could hear everyone talk."*
	Additional benefits ($n = 16$)	*"Don't have to reply on reading lips!! and/or what million times."*
	Speech sound quality ($n = 7$)	*"I could hear annunciation very clear."*
	Speech understanding ($n = 5$)	*"understand conversations through the background noise."*
In restaurant ($n = 43$)	Speech understanding ($n = 38$)	*"Having a nice conversation at a restaurant and able to understand some new voices even with accents."*
	Sound quality ($n = 5$)	*"In a quiet restaurant everything was so clear"*
	Hearing children ($n = 35$)	Enhanced communication ($n = 24$)	*"I am able to communicate better with my grandchildren and am not asking "what did you say" all the time…"*
		Speech understanding ($n = 7$)	*"Understanding my grandson [name] is so much easier!!!"*
		Sound quality ($n = 4$)	*"I could hear my grandchildren clearly.."*

(Continued)

TABLE 1 (Continued)

Numbers in brackets are the frequency of the meaning units reported in each category. PD, pre-determined; SG, self-generated.

TABLE 2 Leisure related experiences ($n = 298$ meaning units).

Sub-category (PD)	Meaning unit (SG)	Meaning unit example
Watching TV ($n = 227$)	Listening experience ($n = 143$)	*"I could hear the TV with the volume turned to 11 when mostly it's 25 or higher"*
	Sound quality ($n = 46$)	*"TV voices are clearer than before…"*
	Speech understanding ($n = 20$)	*"I can understand what people in shows are saying"*
	Direct streaming ($n = 18$)	*"My understanding of movies using the TV connect is a lot better and more enjoyable."*
Listening music ($n = 60$)	Listening experience ($n = 33$)	*"Music sounds good in my hearing aids. Nice to have that functionality."*
	Sound quality ($n = 27$)	*"Nice bright and punchy sound quality for music"*
Entertainment ($n = 6$)	Media ($n = 3$)	*"I can hear a podcast I was having difficulty with earlier."*
	Volume ($n = 3$)	*"Wife says radio is not as loud as usual"*
Playing games ($n = 5$)	Listening experience ($n = 5$)	*"Playing pc games on the computer is wonderful as I can hear everything."*

Numbers in brackets are the frequency of the meaning units reported in each category. PD, pre-determined; SG, self-generated.

identified meaning unit was occupational settings ($n = 26$), in which HAs facilitated effective communication, even with face masks on, and heightened awareness of work-related sounds. HAs were valuable in various situations, including recreation, education, travel, social events, shopping, and worship. Participants reported clearer speech, improved hearing, and enhanced satisfaction across these diverse settings.

Category 4: Device-related benefits of HAs

Two pre-determined sub-categories streaming media ($n = 54$) and battery or charging ($n = 34$) were combined into the overarching category "device-related benefits" (see Figure 2).

TABLE 3 Lifestyle related experiences ($n = 140$ meaning units).

Sub-category (PD)	Meaning unit (SG)	Meaning unit example
Activities ($n = 64$)	Occupational ($n = 26$)	*"I can hear my clients in the styling chair, even with masks on…"*
	Recreational ($n = 23$)	*"I could hear conversations better while on a hike."*
	Educational ($n = 11$)	*"Noticeable difference during research seminar, sitting at back and can hear fine"*
	Traveling ($n = 4$)	*"Great sounds on airline for movies chat and other"*
Social event ($n = 41$)	Conversation engagement ($n = 21$)	*"Friends visited able to hear what was being said without too much difficulty"*
	Speech understanding ($n = 12$)	*"I could carry on a conversation with my kids while my 6 grandkids were yelling and screaming in the background…"*
	Sound quality ($n = 8$)	*"Much clearer with voices in a crowded room..."*
Social activities ($n = 18$)	Participation ($n = 18$)	*"Social groups are more enjoyable now"*
Shopping ($n = 11$)	General benefit ($n = 11$)	*"Heard the checkout operator real well!!"*
Worship ($n = 6$)	General benefit ($n = 6$)	*"Hearing the Sunday morning sermon so clearly!"*

Numbers in brackets are the frequency of the meaning units reported in each category. PD, pre-determined; SG, self-generated.

TABLE 4 Device related experiences ($n = 88$ meaning units).

Sub-category (PD)	Meaning unit (SG)	Meaning unit example
Streaming media ($n = 54$)	Listening experience ($n = 41$)	*"Streaming podcasts from an iPhone 7, works great!"*
	Sound quality ($n = 13$)	*"I could hear my children and grandchildren on face time more clearly"*
Battery or charging ($n = 34$)	Duration ($n = 23$)	*"Battery life both in the aids and the charger is great."*
	Rechargeability ($n = 9$)	*"Recharging very convenient."*
	Replacing batteries ($n = 2$)	*"Changed the batteries today to get into routine of changing them every week. Was easy and quick."*

Numbers in brackets are the frequency of the meaning units reported in each category. PD, pre-determined; SG, self-generated.

Five self-generated meaning units, namely listening experience, sound quality, duration, rechargeability, and battery replacement (see Table 4 and Supplementary material 5 for elaboration on meaning units). Battery and charging experiences were characterized by descriptions of extended battery life with some participants maintaining Bluetooth connectivity throughout the day. Rechargeable HAs offered convenience and eliminated the need for disposable batteries. Participants using disposable batteries reported no hindrance, and the process of replacing batteries was deemed straightforward.

TABLE 5 Positive experiences for the pre-determined sub-category, Other ($n = 351$ meaning units).

Sub-category (PD)	Meaning unit (SG)	Meaning unit example
Other ($n = 351$)	General benefits ($n = 200$)	*"I notice some improvement to my range of hearing."*
	Device-related ($n = 83$)	*"Surprised that they're not noticeable…"*
	Environmental sounds ($n = 41$)	*"I have especially enjoyed hearing the spring songbirds a rich experience I have missed."*
	Speech understanding ($n = 27$)	*"It was many conversations at once. I am not overwhelmed, and I can understand the individual conversations."*

Numbers in brackets are the frequency of the meaning units reported in each sub-category. PD, pre-determined; SG, self-generated.

Category 5: Other benefits of HAs

This pre-determined sub-category was the most frequently selected ($n = 351$) (see Figure 2). Four categories were self-generated by the researcher from the responses, encompassing general advantages, device-related aspects, environmental sound considerations, and enhanced speech understanding (see Table 5 and Supplementary material 6 for elaboration on meaning units). Participants reported noticeable improvements in their overall hearing abilities, leading to reduced instances of asking for repetitions and enhancing daily interactions. The comfort and inconspicuous nature of the devices were particularly noteworthy. A richer auditory experience emerged as wearers appreciated sounds not heard as well-before using HAs, such as the melodic songs of springtime birds to the sizzling of bacon.

4 Discussion

This study employed a qualitative content analysis to explore positive real-world HA use experiences of a large sample of HA wearers, who provided self-initiated EMAs through a smartphone application. The pre-determined sub-categories that resulted from the users in-app responses were grouped into categories, as part of the manual qualitative content analysis, and consisted of (1) conversational settings, (2) leisure-related, (3) lifestyle-related, and (4) device-related aspects, and (5) other listening situations (see Supplementary materials 2–6 for more detailed examples of the sub-categories and meaning units).

The cornerstone of hearing rehabilitation is using HAs to improve access to sounds and speech, thereby enhancing communication—a fundamental aspect of daily life (Ferguson et al., 2017). This study particularly highlighted the significant benefits in

FIGURE 2
Overview of 18 pre-determined sub-categories that were derived from the listening situation self-selected by the users in the app (light blue boxes), organized into five overarching categories (dark blue boxes) as part of the qualitative content analysis.

conversational settings, with a notable emphasis on the advantages of smartphone-connected HAs in facilitating effective telephone communication. The rise in smartphone ownership among adults in the United States and the United Kingdom since 2015 has paralleled an increase in research into smartphone-connected HAs, underscoring improvements in phone conversation quality, speech intelligibility, and reduced listening effort thanks to direct streaming capabilities (Maidment et al., 2019; Gomez et al., 2022; Pew Research Center, 2022). Moreover, the capacity for wearers to self-manage their HAs via a smartphone app has not only contributed to enhanced wearer satisfaction and integration into daily routines but has also empowered wearers through improved autonomy and ownership over their hearing experience (Chasin, 2017). Direct streaming, particularly for media and communication via videotelephony platforms, has been identified as a pivotal feature, aligning with the trend toward greater technological integration within HA design. This trend underscores the importance of connectivity in augmenting the wearers experience, fostering enhanced engagement with modern communication platforms (Chasin, 2017), and thereby enriching both listening experiences and social connections. Enhanced speech understanding was consistently reported, particularly in intimate settings such as conversations with spouses, further demonstrating the value of real-time EMA in providing detailed insights into everyday listening environments outside typical scenarios. Across various settings—from quiet conversations to dynamic social events—improvements in sound quality have been central to the wearer's enhanced ability to engage in meaningful interactions (Cox et al., 2014), reinforcing the critical role of sound quality in effective communication (Kaplan-Neeman et al., 2012).

In occupational settings, HAs significantly enhanced communication and sound awareness, aligning with prior research emphasizing the critical role of hearing in professional environments and the heavy reliance of wearers on their devices in such settings (Granberg and Gustafsson, 2021; Timmer et al., 2023). This study extends these insights by demonstrating positive experiences even in traditionally challenging situations such as group meetings and noisy environments (Picou, 2020; Oosthuizen

et al., 2022), where HAs facilitated effective communication and sound source localization in the workplace. These findings, consistent with earlier studies, underscore the efficacy of HAs in improving occupational performance and highlight the unique value of self-initiated EMA data in capturing real-life, wearer-specific experiences, thus advocating for improved person-centered care. The positive feedback from self-initiated EMAs not only highlights the functional benefits of hearing technology in challenging situations but also underscores its impact on social engagements and interactions with close companions. This enhancement of communication and social involvement echoes previous findings on the significance of audiological interventions for improving the quality of life among individuals with hearing impairments (Holman et al., 2021).

It is widely recognized that, in addition to speech clarity, effective communication, and sound quality (Picou, 2020; Oosthuizen et al., 2022), HA wearers value device usage during leisure activities such as watching television (Strelcyk and Singh, 2018), enjoying music, and gaming (Greasley et al., 2020). The same was found in this study's analysis of leisure-related experiences, where watching TV and listening to music were the two most self-selected leisure-related sub-categories. Specifically, the enhanced appreciation of listening to and performing music when using HAs reported in our study, supported the survey results by Greasley et al. (2020). Participants also reported improved speech clarity during several recreational activities, contributing to enhanced participation and enjoyment e.g., while playing netball, hiking, and taking guitar lessons. Improved listening experiences while playing computer games were also mentioned.

In contrast to general situations included in typical self-report questionnaires (Timmer et al., 2018), the findings of this study offer deeper insights into the nuanced benefits and satisfaction derived from hearing aid use in specific lifestyle-related activities, such as airline travel—a context scarcely documented in existing literature. The unique capacity of EMA to capture real-time feedback across diverse life situations not only enriches our understanding of hearing aid utility but also provides a rich dataset for informing device design and clinician support strategies. This granular insight, especially from unique contexts like airline travel, can serve as valuable data, guiding the development of HAs optimized for both common and complex environments. Moreover, the documented psychosocial benefits, including enhanced engagement in social settings and increased self-assurance in communication, underscore the comprehensive impact of HAs on wearers' wellbeing (Holman et al., 2019, 2021; Vercammen et al., 2021; Gomez et al., 2022; Oosthuizen et al., 2022). These findings highlight the transformative potential of hearing rehabilitation, affirming its role in improving not just hearing but the overall quality of life. Thus, by leveraging EMA data, clinicians are empowered to make nuanced adjustments that address the full spectrum of wearers' needs, fostering improved hearing care that is as dynamic as the lives of the individuals it aims to support.

HA wearers also attributed positive experiences to the battery life and convenience of rechargeable HAs in terms of device-related experiences. Similarly, participants in a previous study reported that rechargeable technology is reliable and offers consistent performance (Taneja, 2020). Rechargeable HAs, noted for their reliability and consistent performance (Taneja, 2020), offer an eco-friendly alternative to disposables, simplifying daily routines and reducing costs (Sun, 2019). Despite the ease of recharging, the straightforward replacement of disposable batteries was also appreciated since it can avoid downtime due to batteries needing to be recharged. This emphasizes the importance of person-centered care that supports wearers to make informed choices based on differentiated advantages related to rechargeable and replaceable battery devices and considers proficiency in HA management skills (Campos et al., 2014).

Participants reported a range of benefits from using HAs, including enhanced hearing optimization, the ability to adjust volume for improved hearing range, and fewer needs to ask for repetitions, classified under the "Other" sub-category. Wearers reported wearing HAs comfortably throughout various daily activities, even during sleep, and valued the discreetness of their devices. This inconspicuousness plays a crucial role in diminishing the stigma often associated with HAs, fostering a more positive wearer attitude and enhancing overall device satisfaction (Maidment et al., 2019). Additionally, the enriched perception of environmental sounds—ranging from the natural ambiance of birds and waves to the everyday sounds of home appliances—further underscores the comprehensive benefits of HAs. These improvements contribute to a more engaging and emotionally positive auditory experience, underlining the significant role HAs play in facilitating wearers' active participation in life's diverse scenarios and mitigating communication challenges.

4.1 Limitations

To our knowledge, this study is the first to perform a manual qualitative content analysis on self-initiated EMA data. However, several limitations should be acknowledged. Primarily, the data was collected as part of clinical practice and intended to support HA wearers and clinicians, and thus not initially collected for research (Friedman et al., 2015). This secondary use of the data, while insightful, introduces challenges related to data quality and generalizability, as also acknowledged in studies by Verheij et al. (2018) and Dillard et al. (2020).

The exclusive use of a specific brand of smartphone-connected HAs, including the functionality that had to be activated by the clinician, limited participation, potentially introducing a selection bias toward a more technologically adept and motivated subgroup. Also, we included responses from individuals who provided feedback in English only.The absence of detailed demographic and audiological profiles of participants further restricts the findings' applicability across a broader HA wearers' population.

Additionally, the use of the mobile application's pre-determined sub-categories (i.e., pre-determined listening situations presented in-app for the user to respond to and select) to guide the categorization and grouping of responses may have constrained the coding process. While this classification ensured consistency, it might have limited the exploration of emergent themes not predefined in the app. We acknowledge the potential overlap of similar responses in different categories as a methodological

limitation, as the predefined sub-categories imposed a structure not necessarily shared by all participants. Furthermore, the reliance on self-initiated EMAs introduces potential biases related to participant self-selection and memory recall, as users may selectively report experiences they perceive as significant. The voluntary nature of feedback and the potential for retrospective reporting introduces risks regarding compliamce and recall bias, respectively (Shiffman et al., 2008). In addition, this study focuses exclusively on positive self-initiated EMA responses, presenting a partial view of the overall hearing aid experience. The exclusion of negative responses means that while the study highlights significant benefits, it does not document the full range of user experiences, including potential challenges and negative aspects. Additionally, a follow-up paper is underway to analyze the negative EMA feedback, which will complement this work and provide a more balanced understanding of hearing aid experiences. Furthermore, the free-text EMA method, while rich in detail, can be time-consuming for participants, which may influence engagement and data comprehensiveness.

Future studies could benefit from enhanced app instructions and prompts, alongside efforts to capture a more diverse participant demographic to broaden the research's relevance. Addressing these limitations in future research would help in obtaining more generalized and comprehensive insights into the real-world benefits of hearing aids.

5 Conclusions

Qualitative self-initiated EMA with positive sentiment has demonstrated its potential to uncover the diverse benefits of HAs, offering unique insights into the wearer's experience in real-world settings. The effort participants invest in free-text EMAs yields significant insights, particularly when analyzing positive EMA statements. This study confirms the substantial role of HAs in enhancing listening experiences, sound quality, and communication, even in less documented contexts such as air travel. Features like direct streaming, extended battery life, and rechargeability were particularly valued, bolstering satisfaction and supporting audiologists in delivering personalized auditory solutions. An innovative use of EMA through smartphone apps could enable wearers to contribute feedback spontaneously, allowing for example the immediate analysis of voice notes via NLP strategies. This could extract meaningful themes in real-time, informing clinicians or activating support mechanisms (i.e., specialized chatbots) to assist individuals in those exact moments of need. Employing this technology could lead to more dynamic, responsive, and person-centered hearing care, leveraging personal narratives to address the intricacies of daily life for HA wearers and enhance hearing rehabilitation strategies.

Data availability statement

The datasets presented in this article are not readily available because of the mobile application's data privacy notice. Requests to access the datasets should be directed to Charlotte.Vercammen@sonova.com.

Ethics statement

The studies involving humans were approved by Research Ethics Committee of the Faculty of Humanities, University of Pretoria, South Africa (HUM023/0922). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

CF: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. FM-A: Supervision, Writing – review & editing. IO: Supervision, Writing – review & editing. VM: Supervision, Writing – review & editing. CV: Supervision, Writing – review & editing. DS: Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. CV was supported by the NIHR Manchester Biomedical Research Centre. IO is a post-doctoral fellow at the University of Pretoria supported by a grant from Sonova, AG.

Conflict of interest

CV is professionally employed by Sonova AG. VM serves as the scientific advisor for hearX SA Pty Ltd.

The remaining authors state that the study was done without any commercial or financial links that could be interpreted as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fauot.2024.1397822/full#supplementary-material

References

Baden, C., Pipal, C., Schoonvelde, M., and van der Velden, M. A. C. G. (2022). Three gaps in computational text analysis methods for social sciences: a research agenda. *Commun. Methods Meas.* 16, 1–18. doi: 10.1080/19312458.2021.2015574

Bolger, N., Davis, A., and Rafaeli, E. (2003). Diary methods: capturing life as it is lived. *Annu. Rev. Psychol.* 54, 579–616. doi: 10.1146/annurev.psych.54.101601.145030

Boothroyd, A. (2007). Adult aural rehabilitation: what is it and does it work? *Trends Amplif.* 11, 63–71. doi: 10.1177/1084713807301073

Brice, S., and Almond, H. (2022). Is teleaudiology achieving person-centered care: a review. *Int. J. Environ. Res. Public Health* 19:7436. doi: 10.3390/ijerph19127436

Burke, L. E., Shiffman, S., Music, E., Styn, M. A., Kriska, A., Smailagic, A., et al. (2017). Ecological momentary assessment in behavioral research: addressing technological and human participant challenges. *J. Med. Int. Res.* 19:77. doi: 10.2196/jmir.7138

Campos, P. D., Bozza, A., and Ferrari, D. V. (2014). Hearing aid handling skills: relationship with satisfaction and benefit. *CoDAS* 26, 10–16. doi: 10.1590/s2317-17822014.001-0003

Chasin, M. (2017). *Overview of Smartphone Control of Hearing Aids*. Canadian Audiologist. Available at: https://canadianaudiologist.ca/smartphone-overview-feature/ (accessed November 3, 2023).

Cox, R. M., Johnson, J. A., and Xu, J. (2014). Impact of advanced hearing aid technology on speech understanding for older listeners with mild to moderate, adult-onset, sensorineural hearing loss. *J. Gerontol.* 60, 557–568. doi: 10.1159/000362547

Dillard, L. K., Saunders, G. H., Zobay, O., and Naylor, G. (2020). Insights into conducting audiological research with clinical databases. *Am. J. Audiol.* 29, 676–681. doi: 10.1044/2020_AJA-19-00067

Ferguson, M. A., Kitterick, P. T., Chong, L. Y., Edmondson-Jones, M., Barker, F., and Hoare, D. J. (2017). Hearing aids for mild to moderate hearing loss in adults. *Cochr. Database Syst. Rev.* 9:CD012023. doi: 10.1002/14651858.CD012023.pub2

Franks, I., and Timmer, B. H. B. (2023). Reasons for the non-use of hearing aids: perspectives of non-users, past users, and family members. *Int. J. Audiol.* 1–8. doi: 10.1080/14992027.2023.2270703

Friedman, C., Rubin, J., Brown, J., Buntin, M., Corn, M., Etheredge, L., et al. (2015). Toward a science of learning systems: a research agenda for the high-functioning Learning Health System. *J. Am. Med. Inform. Assoc.* 22, 43–50. doi: 10.1136/amiajnl-2014-002977

Galvez, G., Turbin, M. B., Thielman, E. J., Istvan, J. J., Andrews, J. A., and Henry, J. A. (2012). Feasibility to ecological momentary assessment of hearing difficulties encountered by hearing aid users. *Ear Hear.* 33, 497–507. doi: 10.1097/AUD.0b013e3182498c41

Gomez, R., Habib, A., Maidment, D. W., and Ferguson, M. A. (2022). Smartphone-connected hearing aids enable and empower self-management of hearing loss: a qualitative interview study underpinned by the behavior change wheel. *Ear Hear.* 43, 921–932. doi: 10.1097/AUD.0000000000001143

Granberg, S., and Gustafsson, J. (2021). Key findings about hearing loss in the working-life: a scoping review from a well-being perspective. *Int. J. Audiol.* 60, 60–70. doi: 10.1080/14992027.2021.1881628

Graneheim, U. H., and Lundman, B. (2004). Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. *Nurse Educ. Today* 24, 105–112. doi: 10.1016/j.nedt.2003.10.001

Greasley, A., Crook, H., and Fulford, R. (2020). Music listening and hearing aids: perspectives from audiologists and their patients. *Int. J. Audiol.* 59, 694–706. doi: 10.1080/14992027.2020.1762126

Hasan, S. S., Chipara, O., Wu, Y. H., and Aksan, N. (2014). "Evaluating auditory contexts and their impacts on hearing aid outcomes with mobile phones," in *Proceedings - PERVASIVEHEALTH 2014: 8th International Conference on Pervasive Computing Technologies for Healthcare*, 126–133. doi: 10.4108/icst.pervasivehealth.2014.254952

Heselton, T., Bennett, R. J., Manchaiah, V., and Swanepoel, W. (2022). Online reviews of hearing aid acquisition and use: a qualitative thematic analysis. *Am. J. Audiol.* 31, 284–298. doi: 10.1044/2021_AJA-21-00172

Holman, J. A., Drummond, A., Hughes, S. E., and Naylor, G. (2019). Hearing impairment and daily-life fatigue: a qualitative study. *Int. J. Audiol.* 58, 408–416. doi: 10.1080/14992027.2019.1597284

Holman, J. A., Drummond, A., and Naylor, G. (2021). Hearing aids reduce daily-life fatigue and increase social activity: a longitudinal study. *Trends hear.* 25:23312165211052786. doi: 10.1177/23312165211052786

Holube, I., von Gablenz, P., and Bitzer, J. (2020). Ecological momentary assessment in hearing research: current state, challenges, and future directions. *Ear Hear.* 41, 79–90. doi: 10.1097/AUD.0000000000000934

Humes, L. E. (2003). Modeling and predicting hearing aid outcome. *Trends Amplif.* 7, 41–75. doi: 10.1177/108471380300700202

Jiang, J. A., Wade, K., Fiesler, C., and Brubaker, J. R. (2021). Supporting serendipity: opportunities and challenges for human-ai collaboration in qualitative analysis. *Proc. ACM Hum. Comput. Interact.* 5, 1–23. doi: 10.1145/3449168

Kaplan-Neeman, R., Muchnik, C., Hildesheimer, M., and Henkin, Y. (2012). Hearing aid satisfaction and use in the advanced digital era. *Laryngosc. Investig. Otolaryngol.* 122, 2029–2036. doi: 10.1002/lary.23404

Knoetze, M., Manchaiah, V., Mothemela, B., and Swanepoel, W. (2023). Factors influencing hearing help-seeking and hearing aid uptake in adults: a systematic review of the past decade. *Trends Hear.* 27:23312165231157255. doi: 10.1177/23312165231157255

Knudsen, L. V., Laplante-Lévesque, A., Jones, L., Preminger, J. E., Nielsen, C., Lunner, T., et al. (2012). Conducting qualitative research in audiology: a tutorial. *Int. J. Audiol.* 51, 83–92. doi: 10.3109/14992027.2011.606283

Maidment, D. W., Ali, Y. H. K., and Ferguson, M. A. (2019). Applying the COM-B model to assess the usability of smartphone-connected listening devices in adults with hearing Loss. *J. Am. Acad. Audiol.* 30, 417–430. doi: 10.3766/jaaa.18061

Manchaiah, V., Swanepoel, D. W., Bailey, A., Pennebaker, J. W., and Bennett, R. J. (2021a). Hearing aid consumer reviews: a linguistic analysis in relation to benefit and satisfaction ratings. *Am. J. Audiol.* 30, 761–768. doi: 10.1044/2021_AJA-21-00061

Manchaiah, V., Swanepoel, D. W., and Bennett, R. J. (2021b). Online consumer reviews on hearing health care services: a textual analysis approach to examine psychologically meaningful language dimensions. *Am. J. Audiol.* 30, 669–675. doi: 10.1044/2021_AJA-20-00223

Mothemela, B., Manchaiah, V., Mahomed-Asmail, F., Knoetze, M., and Swanepoel, W. (2023). Factors influencing hearing aid use, benefit and satisfaction in adults: a systematic review of the past decade. *Int. J. Audiol.* 63, 661–674. doi: 10.1080/14992027.2023.2272562

Oosthuizen, I., Manchaiah, V., Launer, S., and Swanepoel, W. (2022). Hearing aid experiences of adult hearing aid owners during and after fitting: a systematic review of qualitative studies. *Trends Hear.* 26:23312165221130584. doi: 10.1177/23312165221130584

Pew Research Center (2022). *Social Media Seen as Mostly Good for Democracy Across Many Nations, But U.S. Is a Major Outlier*. Available at: https://www.pewresearch.org/global/2022/12/06/internet-smartphone-and-social-media-use-in-advanced-economies-2022/ (accessed December 11, 2023).

Picou, E. M. (2020). MarkeTrak 10 (MT10) Survey results demonstrate high satisfaction with and benefits from hearing aids. *Semin. Hear.* 41, 21–36. doi: 10.1055/s-0040-1701243

Schinkel-Bielefeld, N., Kunz, P., Zutz, A., and Buder, B. (2020). Evaluation of hearing aids in everyday life using Ecological Momentary Assessment: what situations are we missing? *Am. J. Audiol.* 29, 591–609. doi: 10.1044/2020_AJA-19-00075

Shiffman, S., Stone, A. A., and Hufford, M. R. (2008). Ecological momentary assessment. *Annu. Rev. Clin. Psychol.* 4, 1–32. doi: 10.1146/annurev.clinpsy.3.022806.091415

Strelcyk, O., and Singh, G. (2018). TV listening and hearing aids. *PLoS ONE* 13:e0200083. doi: 10.1371/journal.pone.0200083

Sun, B. M. (2019). *The Counseling Advantages of Rechargeable Hearing Aid Batteries*. CUNY Academic Works. Available at: https://academicworks.cuny.edu/gc_etds/3188

Taneja, N. (2020). Rechargeable battery slutions for digital hearing aids: a mini review. *Int. J. Otolaryngol.* 5, 14–17. Available at: https://openventio.org/wp-content/uploads/Rechargeable-Battery-Solutions-for-Digital-Hearing-Aids-A-Mini-Review-OTLOJ-SE-5-104.pdf

Timmer, B. H. B., Bennett, R. J., Montano, J., Hickson, L., Weinstein, B., Wild, J., et al. (2023). Social-emotional well-being and adult hearing loss: clinical recommendations. *Int. J. Audiol.* 63, 381–392. doi: 10.1080/14992027.2023.2190864

Timmer, B. H. B., Hickson, L., and Launer, S. (2017). Ecological momentary assessment: feasibility, construct validity, and future applications. *Am. J. Audiol.* 26, 436–442. doi: 10.1044/2017_AJA-16-0126

Timmer, B. H. B., Hickson, L., and Launer, S. (2018). The use of ecological momentary assessment in hearing research and future clinical applications. *Hear. Res.* 369:24228. doi: 10.1016/j.heares.2018.06.012

Trull, T. J., and Ebner-Priemer, U. (2014). The role of ambulatory assessment in psychological science. *Curr. Dir. Psychol. Sci.* 23, 466–470. doi: 10.1177/0963721414550706

Vercammen, C., Bott, A., and Saunders, G. H. (2021). Hearing health in the broader context of healthy living and well-being: changing the narrative. *Int. J. Audiol.* 60, 86–91. doi: 10.1080/14992027.2021.1905893

Vercammen, C., Oosthuizen, I., Manchaiah, V., Ratinaud, P., Launer, S., and Swanepoel, D. W. (2023). Real-life and real-time hearing aid experiences: Insights from self-initiated ecological momentary assessments and natural language analysis. *Front. Digit. Health.* 5:1104308. doi: 10.3389/fdgth.2023.1104308

Verheij, R. A., Curcin, V., Delaney, B. C., and McGilchrist, M. M. (2018). Possible sources of bias in primary care electronic health record data use and reuse. *J. Med. Int. Res.* 20:e185. doi: 10.2196/jmir.9134

von Gablenz, P., Kowalk, U., Bitzer, J., Meis, M., and Holube, I. (2021). Individual hearing aid benefit in real life evaluated using Ecological Momentary Assessment. *Trends Hear.* 25:2331216521990288. doi: 10.1177/2331216521990288

Wong, L. L., Hickson, L., and McPherson, B. (2003). Hearing aid satisfaction: what does research from the past 20 years say? *Trends Amplif.* 7, 117–161. doi: 10.1177/108471380300700402

Wu, Y., Stangl, E., Chipara, O., Hasan, S. S., Wellhaven, A., and Oleson, J. (2018). Characteristics of real-world signal-to-noise ratios and speech listening situations of older adults with mild-to-moderate hearing loss. *Ear Hear.* 39, 293–304. doi: 10.1097/AUD.0000000000000486

Wu, Y. H., Stangl, E., Zhang, X., and Bentler, R. A. (2015). Construct validity of the ecological momentary assessment in audiology research. *J. Am. Acad. Audiol.* 26, 872–884. doi: 10.3766/jaaa.15034

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 04 December 2024
DOI 10.3389/fnins.2024.1491954

Check for updates

OPEN ACCESS

EDITED BY
Laura Coco,
San Diego State University, United States

REVIEWED BY
William Bologna,
Towson University, United States

*CORRESPONDENCE
Bhavisha J. Parmar
✉ bp472@cam.ac.uk

RECEIVED 05 September 2024
ACCEPTED 23 October 2024
PUBLISHED 04 December 2024

CITATION
Parmar BJ, Salorio-Corbetto M, Picinali L, Mahon M, Nightingale R, Somerset S, Cullington H, Driver S, Rocca C, Jiang D and Vickers D (2024) Virtual reality games for spatial hearing training in children and young people with bilateral cochlear implants: the "Both Ears (BEARS)" approach.
Front. Neurosci. 18:1491954.
doi: 10.3389/fnins.2024.1491954

COPYRIGHT
© 2024 Parmar, Salorio-Corbetto, Picinali, Mahon, Nightingale, Somerset, Cullington, Driver, Rocca, Jiang and Vickers. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Virtual reality games for spatial hearing training in children and young people with bilateral cochlear implants: the "Both Ears (BEARS)" approach

Bhavisha J. Parmar[1,2]*, Marina Salorio-Corbetto[1], Lorenzo Picinali[3], Merle Mahon[4], Ruth Nightingale[4], Sarah Somerset[5,6], Helen Cullington[7], Sandra Driver[8], Christine Rocca[8], Dan Jiang[8,9] and Deborah Vickers[1]

[1]SOUND Lab, Department of Clinical Neurosciences, University of Cambridge, Cambridge, United Kingdom, [2]Ear Institute, University College London, London, United Kingdom, [3]Dyson School of Design Engineering, Faculty of Engineering, Imperial College London, London, United Kingdom, [4]Division of Psychology and Language Sciences, Faculty of Brain Sciences, University College London, London, United Kingdom, [5]School of Medicine, University of Nottingham, Nottingham, United Kingdom, [6]Nottingham Hearing Biomedical Research Centre, University of Nottingham, Nottingham, United Kingdom, [7]Auditory Implant Service University of Southampton, Southampton, United Kingdom, [8]St Thomas' Hearing Implant Centre, Guy's and St Thomas' NHS Foundation Trust, London, United Kingdom, [9]Centre for Craniofacial and Regenerative Biology, Faculty of Dentistry, Oral & Craniofacial Sciences, King's College London, London, United Kingdom

Spatial hearing relies on the encoding of perceptual sound location cues in space. It is critical for communicating in background noise, and understanding where sounds are coming from (sound localization). Although there are some monoaural spatial hearing cues (i.e., from one ear), most of our spatial hearing skills require binaural hearing (i.e., from two ears). Cochlear implants (CIs) are often the most appropriate rehabilitation for individuals with severe-to-profound hearing loss, with those aged 18 years of age and younger typically receiving bilateral implants (one in each ear). As experience with bilateral hearing increases, individuals tend to improve their spatial hearing skills. Extensive research demonstrates that training can enhance sound localization, speech understanding in noise, and music perception. The BEARS (Both Ears) approach utilizes Virtual Reality (VR) games specifically designed for young people with bilateral CIs to train and improve spatial hearing skills. This paper outlines the BEARS approach by: (i) emphasizing the need for more robust and engaging rehabilitation techniques, (ii) presenting the BEARS logic model that underpins the intervention, and (iii) detailing the assessment tools that will be employed in a clinical trial to evaluate the effectiveness of BEARS in alignment with the logic model.

KEYWORDS

audiology, cochlear implant, spatial hearing, auditory training, sound localization, speech perception, pediatric audiology, deafness (hearing loss)

Background

Internationally, there are over one million cochlear implant (CI) recipients in the United Kingdom (UK) (Zeng, 2022). Every year, there are ~1,500 new CI recipients in the UK (British Cochlear Implant Group, 2024). Of those who are bilaterally implanted, around 75% are 18 years of age or younger.

Extensive evidence supports the conclusion that early cochlear implantation improves speech and language development outcomes in this population (Geers et al., 2003; Sharma et al., 2020; Peixoto et al., 2013), however they often experience significant challenges in speech perception and sound localization, particularly in noisy environments (Zheng et al., 2022; Badajoz-Davila and Buchholz, 2021). Furthermore, bilateral CI users, particularly those sequentially implanted, may experience difficulties in combining sounds from the two implants to create three-dimensional sound (Sparreboom et al., 2012). Some individuals experience "increased effort" when using the second implant due to perceptible differences in sound quality between the devices, which may lead to the rejection of the second implant (Vickers et al., 2021; Myhrum et al., 2017; Watson et al., 2016; Emond et al., 2013). A lack of rehabilitative support to address these challenges has been documented (Mather et al., 2011).

There are currently no standardized clinical fitting protocols, guidance documents, or rehabilitation tools specifically developed to optimize the fitting of bilateral CIs, either in the UK or internationally. Existing rehabilitation techniques with CIs are often unengaging, do not adequately address real-world hearing challenges, and lack targeted training to maximize the benefits of bilateral implantation.

Recognizing the absence of standardized protocols for fitting bilateral CIs, and the need for ecologically valid outcome measures and resources for multi-modal listening training, the BEARS (Both Ears) programme was established. The aim of this paper is to present the BEARS approach and the underpinning logic model, which extends previous research on the development of the BEARS intervention through participatory design methodologies (Vickers et al., 2021).

BEARS programme logic model

The BEARS programme has involved the development of virtual reality (VR) based spatial hearing games designed to enhance spatial hearing in children and young people (CYP, aged 8–16 years) with bilateral CIs (Vickers et al., 2021). It is informed by a logic model (Figure 1) based on the National Institute for Health and Care Research/Medical Research Council (NIHR/MRC) framework for complex health interventions (Skivington et al., 2021), and has developed both the intervention and outcome measures to rigorously assess intervention effectiveness in a randomized controlled trial (RCT, ISRCTN: 92454702). Logic models are visual representations illustrating the interconnected relationships among various components of a programme or study (Skivington et al., 2021; Funnell and Rogers, 2011).

The BEARS logic model integrates multiple components to assess the intervention and its anticipated outcomes, while also accounting for the specific characteristics of the target patient population. The model outlines the external context for implementation, the mechanisms of change, and the potential effects on healthcare delivery should the intervention demonstrate efficacy. Implementation determinants have also been considered in the development of the BEARS logic model. Moderating and mediating factors include chronological age at first implant, developmental age, training engagement, school setting, duration of hearing before severe-profound deafness, type of intervention device, CI center, number of active CI electrodes, and level of asymmetrical hearing loss. Here, the components of the logic model are presented in more detail.

Target population

The BEARS logic model is grounded in developmental theory, which accounts for the biological, psychological, social, and emotional changes occurring with age (Piaget, 1971). Within our RCT study population of 384 CYP [power calculation based on pilot data using the BEARS primary outcome measure (spatial speech-in-noise)] with bilateral CIs, it is anticipated that participants will have reached either the "concrete operational" stage, characterized by logical thinking about tangible objects, or the "formal operational" stage, marked by the development of abstract thinking and a more complex understanding of the world. They will also have reached a "cognitive stage" which is linked to the proposed change mechanisms. As hearing abilities improve, this should develop knowledge construction of the world, increasing self-confidence and socio-emotional development (i.e., improve experience, expression, management of emotions and ability to establish positive relationships with others). Participants will be bilateral CI users with a minimum of 6 h of daily usage and stable aided hearing levels (within ±10 dB across 500 Hz−4 kHz), confirmed over at least the two most recent clinical review appointments.

Intervention plan

Virtual Reality (VR), which relies on immersive, computer-generated audio-visual environments, is increasingly being applied in health research and healthcare delivery. Users interact with VR environments through a head-mounted display and handheld controllers. In auditory research, VR has been utilized to assess listening abilities (Salanger et al., 2020), train localization skills (Shim et al., 2023; Alzaher et al., 2023), and measure the benefits of hearing aids (Grimm et al., 2016). The advantages of VR can include enhanced experimental reproducibility, a reduced need for additional resources and complex speaker-array equipment, as well as increased applicability to real-world scenarios. These benefits may improve the utility of VR based rehabilitation and diagnostic in clinical scenarios. Furthermore, simulating physical spaces through VR, could be advantageous to train, monitor and potentially improve how hearing device users physically respond to sounds e.g., head turn movements and positioning (Grange et al., 2018), in addition to speech perception and sound localization performance.

Developing VR games to improve spatial hearing

The BEARS intervention is a suite of VR games (Figure 2), delivered via the Meta Quest 2 head-mounted device and a pair of headphones, and specifically designed to enhance the spatial hearing abilities of CYP with bilateral CIs. The BEARS intervention was developed using a participatory design approach, as outlined in Vickers et al. (2021), where stakeholders, including

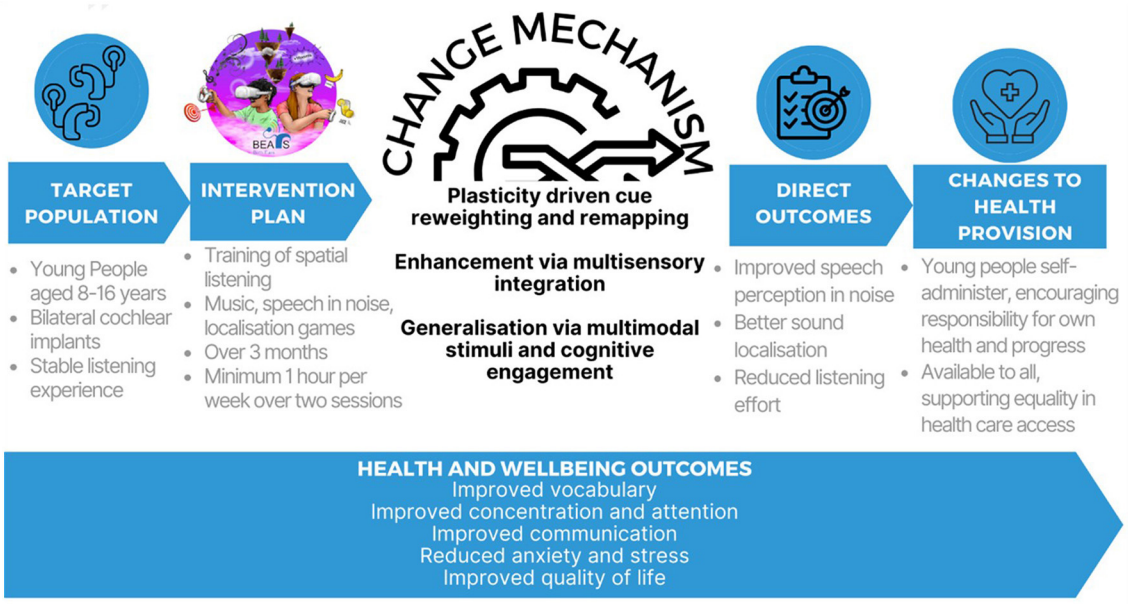


FIGURE 1
The BEARS logic model created using the Medical Research Council (MRC) framework on complex interventions to improve health (Skivington et al., 2021).

CI users, served as co-creators (Vickers et al., 2021). CI users provided valuable feedback on various aspects of the games, such as usability, content, difficulty levels, and settings. Clinicians, including audiologists, speech and language therapists, teachers of the deaf, and music therapists, also played a critical role by evaluating the BEARS training package and suggesting important stimuli for enhancing speech and hearing development. Through collaborative workshops, patients, clinicians, researchers, and engineers reached a consensus that the training package was appropriately designed and ready for use in a randomized controlled trial to evaluate effectiveness.

During this iterative process, modifications were made to ensure the games encompassed a wide range of scenarios, reward systems, lessons and challenges of varying difficulty and clear instructions. They were structured to provide feedback and measure success. Additionally, an iPad version was developed for participants with smaller heads, who find the VR headset uncomfortable, for those who do not like the experience of using the head mounted display, or those with vestibular disorders or significant motion sickness while playing video games. The sound from the VR headset can be presented using inbuilt loudspeakers but due to positioning there was inconsistency in the quality of delivery to the CI processor microphones. There are also options for sound delivery via Bluetooth connection or direct audio-input, but the participatory feedback groups indicated that many individuals were not comfortable in using these listening options and there was greater variability across CI manufacturers. Therefore, headphones were chosen as the most consistent to use for all participants. Various headphones were evaluated for comfort, ease and consistency of use through workshops with the target population, and quality of acoustic representation of the signals and robustness to headphone placement were evaluated by electro-acoustic assessments. AKG k240 headphones were selected for audio delivery with headsets, and with iPads.

The BEARS training package consists of three VR training games to enhance spatial hearing, using target localization, speech perception, and music content (Figure 2). Each game is based on an audio-visual task performed through the VR interface. Players are automatically guided through on-screen visual prompts to support the gameplay with feedback given on their performance. They progress through levels of increasing difficulty. Challenges and lessons are unlocked during the gameplay; the difficulty of the various levels has been calibrated during the participatory design stage and to provide enough content for the whole duration of the trial. The package is designed for self-administration, allowing players the flexibility to engage with the games at any location and time. Workshops with CYP suggested that it is practical to play the BEARS games for at least 1 h per week, divided into a minimum of two 30-min sessions. Clinician workshops recommended that all three games be incorporated into each session to optimize the use of multiple approaches. Informed by these recommendations, relevant literature and device safety guidance (Rechichi et al., 2017; Meta, 2024), no limitations were placed on the number of gaming sessions; however, participants were advised not to exceed 30 min per session. Device datalogging captures detailed gameplay metrics, including session duration, the number of levels unlocked, game points (stars) earned, and the time spent on each game category.

Target game

Originally developed to train normal-hearing individuals in sound localization when using non-personalized rendering (Steadman et al., 2019), the target game was later adapted for CI

FIGURE 2
The BEARS games. **(Left)** Sound localization game, where participants identify targets. **(Middle)** Speech-in-noise game where participants follow instructions to serve café customers with food and beverage items. **(Right)** Music game, for participants to complete tasks of pitch discrimination, rhythm repetition, and instrument selection.

users. In this game, players are initially trained to localize sounds using audiovisual cues. Sounds can originate from any direction around the player, who must identify the target, represented as a bullseye, and direct their controller toward it. At the outset, the bullseye is clearly visible, but as the difficulty levels increase, it gradually disappears, transitioning the task into a purely auditory challenge. Additional challenges involve locating the target signal amidst interfering stimuli or identifying a set of targets in a specific order, further enhancing the training complexity.

Speech in noise game

Players are immersed in a virtual café environment, where they are tasked with progressively challenging speech recognition activities. These tasks require players to interact dynamically with the environment by rotating their heads to localize characters who are speaking, and accurately identify the spoken words in the presence of varying levels of background babble. As customers approach from different directions, players must accurately locate them, take their orders, and select the appropriate items from the café counter. The complexity of the game increases with the introduction of background noise and additional interfering tasks within the café setting. A set of advancing levels are also available in a scenario where the player needs to make pizzas, putting the correct ingredients onto the pizzas in the right order and delivering them directly to customers or to delivery staff.

Music game

The game aims to enhance perception and localization of musical instruments and lyrics in a range of immersive and interactive soundscapes. Players complete a variety of pitch, timbre, and rhythm discrimination tasks. For example, a pitch discrimination task could involve a participant selecting the location of a pitch-shifted popular song and identifying whether the pitch is higher or lower compared to the original. A rhythm-based task may require participants to use VR controllers to replicate a presented rhythmic beat by playing virtual drums. The music game is based on the Musiclarity web-application, created within the 3D Tune-in project (Reactify, 2024; Cuevas-Rodríguez et al., 2019; Levtov et al., 2016).

Change mechanisms

Although individuals with bilateral CIs generally exhibit better sound localization and speech-in-noise perception compared to those with a unilateral implant, their performance remains significantly below that of typically hearing children (Sarant et al., 2014; Sparreboom et al., 2015; Lovett et al., 2015; Zheng et al., 2015; Lammers et al., 2014). Extensive research indicates that sound localization can be enhanced through targeted training, with evidence suggesting that plasticity-induced changes can occur in the auditory pathways of both children and adults, facilitated by appropriate training systems (Firszt et al., 2015; Yu et al., 2018; Killan et al., 2019; Mathew et al., 2018). These improvements are underpinned by cue remapping—using new spatial cues to develop a revised localization map—and cue reweighting, which involves emphasizing unaltered cues while disregarding altered ones (Steadman et al., 2019).

Computer-based training offers substantial potential, particularly due to its remote delivery capability and greater engagement. Such training has been shown to improve speech-in-noise perception in CI users (Casserly and Barney, 2017). Research also highlights the efficacy of combined training stimuli. For instance, Cai et al. (2018) found audio-visual training to be more effective than auditory-only training, while Steadman et al. (2019) emphasized the importance of auditory-based interaction during training. A systematic review by Rayes et al. (2019) identified multimodal interventions or a combination of bottom-up and top-down training tasks as the most effective for children with CIs. Whitton et al. (2017) demonstrated that audio-motor perceptual training can improve speech-in-noise intelligibility by up to 25%. Stitt et al. (2019) also illustrated the use of virtual auditory displays to create training environments that teach users to localize sounds using modified localization cues. The inclusion of audio-visual stimuli facilitates task familiarization, while gamification enhances engagement and performance.

It is anticipated that the BEARS intervention, compared to standard care alone, will improve spatial hearing, speech-in-noise perception, and listening ease. These improvements are expected to be driven by plasticity-related processes, training-induced increases

in performance change rates and maximum performance, auditory-visual integration, multimodal stimuli, and cognitive engagement-driven generalization. The mechanism of action assumes that the games promote learning.

Direct outcomes

The evaluation of BEARS follows a mixed methods approach to determine whether BEARS (i) improves speech-in-noise perception in spatial environments, (ii) enhances quality of life, (iii) is cost-effective, and (iv) increases the perceived benefits of everyday listening. A range of tools and measures are utilized to assess outcomes, including some specifically developed as part of the BEARS project. The primary outcome measure is a spatial speech in noise assessment. The Spatial Speech in Noise Virtual Acoustics (SSiN-VA) test simultaneously assesses word identification and relative localization and can provide information about spatial release from masking. It is based on a test initially developed by Bizley et al. (2015) and has been adapted into a virtual implementation (Bizley et al., 2015; Salorio-Corbetto et al., 2022). The virtual adaptive sentence-in-noise task, utilizing the Spatial Adaptive Sentence List (Sp-ASL; MacLeod and Summerfield, 1990), is administered in accordance with the BKB-SIN task protocol (Bench et al., 1979). These virtual outcome measures are carried out with an iPad and calibrated headphones, and were developed in response to the limited availability of multi-speaker arrays for spatial hearing assessments in many audiology departments (Parmar et al., 2022). They are intended to make speech-in-noise testing more accessible and efficient for audiologists, and can be adapted for different populations and clinical purposes.

Health provision, health, and wellbeing outcomes

A bespoke quality of life measure, the York Binaural Hearing Related Quality of Life—Youth (YBHRQL-Y) has been developed as part of the BEARS programme (Somerset et al., 2023). This measure has been re-operationalized for use with CYP from the original adult YBHRQL developed by Summerfield et al. (2022). Other health economics questionnaires include the Health Utilities Index 3 (HUI-3; Horsman et al., 2003) the Child Health Utility instrument (CHU-9D; Furber and Segal, 2015). The economic evaluation will calculate incremental cost per quality-adjusted life-year (QALY) gained by offering BEARS and usual care compared to usual care, from a National Health Service (NHS), Personal Social Services and Local Education Provider perspective.

A longitudinal qualitative design is being used to explore CYP's experiences of everyday listening, and to contribute to understanding how the BEARS intervention may lead to perceived changes to that experience. Semi-structured online interviews are being carried out with a subset of 40 participants from both BEARS and usual care arms, at baseline and again after 3 months. In addition, all participants in both arms of the trial are asked to respond to open-ended survey questions at successive timepoints throughout the study (baseline, 3 and 12 months). The interview and survey questions have been co-produced in sessions with deaf CYP. Interview and survey data will be analyzed thematically using a Framework approach (Parkinson et al., 2016). Findings will be discussed with deaf CYP to explore whether the trial data resonates and reflects their own lived experiences as users of CIs.

Conclusion

The BEARS programme comprises a suite of VR games specifically designed to enhance spatial hearing in CYP with bilateral CIs. The development of the BEARS intervention is grounded in evidence presented above, supporting the efficacy of sound localization training, the application of VR technologies, multi-modal training approaches, and the necessity for rehabilitation methods that are both effective and engaging for CYP. These games were co-developed with input from bilateral CI users and other key stakeholders, ensuring their relevance and appeal to the target population. This work is aligned with key objectives outlined in the UK's NHS Long Term Plan (National Health Service, 2019), which emphasizes the importance of expanding digital tools and services to empower patients and support healthcare professionals.

The effectiveness of the intervention will be evaluated within an RCT (ISRCTN: 92454702). The unblinded, multi-center RCT is currently underway to evaluate the effectiveness of a 3-month spatial-listening training programme delivered via the BEARS platform, in addition to usual care, compared to usual care alone. The trial aims to assess improvements in spatial hearing abilities, quality of life, and cost-effectiveness. The study is being conducted across 11 cochlear implant centers in the UK, with a target recruitment of 384 bilateral implanted 8- to 16-year-olds. A 12-month follow up session will assess retention and longer-term effects.

In accordance with the NIHR/MRC framework for complex health interventions (Skivington et al., 2021), we are collaborating with participants, clinicians, and researchers to develop a comprehensive scale-up and implementation strategy. This strategy addresses the immediate challenges of integrating the BEARS intervention into clinical practice, alongside long-term considerations such as ongoing game development, equipment maintenance, and ensuring equitable access. Furthermore, we are partnering with international collaborators to explore the feasibility of global implementation of the BEARS intervention. A critical element of this effort is the BEARS process evaluation, which aims to explore trial compliance, and verify the mechanistic assumptions underlying the intervention's outcomes, and to determine opportunities for optimisation (Moore et al., 2015). Insights from the process evaluation will guide the refinement of the implementation strategy and provide essential information for decision-makers seeking to deploy the intervention across varied settings. To mitigate bias, the process evaluation will be conducted independently of the clinical trial, with data collected by individuals not involved in the design or delivery of the intervention.

The BEARS programme plays a significant role in advancing remote care resources, offering novel interventions that empower patients to take greater ownership of their rehabilitation while

potentially alleviating the burden on healthcare providers. For younger populations, the implementation of VR provides a more engaging alternative to traditional auditory rehabilitation methods. The use of participatory design in the development of BEARS games and outcome measures (Vickers et al., 2021) improves their relevance to the target population, thereby enhancing the likelihood of adoption and sustained use.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

BP: Data curation, Investigation, Methodology, Project administration, Writing – original draft, Writing – review & editing. MS-C: Data curation, Investigation, Methodology, Writing – review & editing. LP: Conceptualization, Data curation, Investigation, Writing – review & editing. MM: Data curation, Investigation, Writing – review & editing. RN: Data curation, Formal analysis, Writing – review & editing. SS: Conceptualization, Data curation, Writing – review & editing. HC: Conceptualization, Methodology, Writing – review & editing. SD: Conceptualization, Data curation, Funding acquisition, Methodology, Writing – review & editing. CR: Data curation, Investigation, Methodology, Writing – review & editing. DJ: Funding acquisition, Writing – review & editing. DV: Conceptualization, Data curation, Funding acquisition, Investigation, Supervision, Writing – review & editing.

Funding

The authors declare financial support was received for the research, authorship, and/or publication of this article. All authors were funded by a Programme Grant for Applied Research NIHR201608. MS-C and DV were funded by the Medical Research Council (MRC) United Kingdom, MR/S002537/1.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Author disclaimer

The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

References

Alzaher, M., Valzolgher, C., Verdelet, G., Pavani, F., Farnè, A., Barone, P., et al. (2023). Audiovisual training in virtual reality improves auditory spatial adaptation in unilateral hearing loss patients. *J. Clin. Med.* 12:2357. doi: 10.3390/jcm12062357

Badajoz-Davila, J., and Buchholz, J. M. (2021). Effect of test realism on speech-in-noise outcomes in bilateral cochlear implant users. *Ear Hear.* 42, 1687–1698. doi: 10.1097/AUD.0000000000001061

Bench, J., Kowal, Å., and Bamford, J. (1979). The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. *Br. J. Audiol.* 13, 108–112. doi: 10.3109/03005367909078884

Bizley, J. K., Elliott, N., Wood, K. C., and Vickers, D. A. (2015). Simultaneous assessment of speech identification and spatial discrimination: a potential testing approach for bilateral cochlear implant users? *Trends Hear.* 19:2331216515619573. doi: 10.1177/2331216515619573

British Cochlear Implant Group (2024). *Annual UK Numbers Update*. Available at: https://www.bcig.org.uk/annual_uk_numbers_update.aspx (accessed August 1, 2024).

Cai, Y., Chen, G., Zhong, X., Yu, G., Mo, H., Jiang, J., et al. (2018). Influence of audiovisual training on horizontal sound localization and its related ERP response. *Front. Hum. Neurosci.* 12:423. doi: 10.3389/fnhum.2018.00423

Casserly, E. D., and Barney, E. C. (2017). Auditory training with multiple talkers and passage-based semantic cohesion. *J. Speech Lang. Hear. Res.* 60, 159–171. doi: 10.1044/2016_JSLHR-H-15-0357

Cuevas-Rodríguez, M., Picinali, L., González-Toledo, D., Garre, C., de la Rubia-Cuestas, E., Molina-Tanco, L., et al. (2019). 3D Tune-In Toolkit: an open-source library for real-time binaural spatialisation. *PLoS ONE* 14:e0211899. doi: 10.1371/journal.pone.0211899

Emond, A., Moore, M., Tjornby, C., and Kentish, R. (2013). Factors associated with poor use of sequential bilateral cochlear implants in young people: a preliminary audit of poor users. *Cochl. Implant. Int.* 14, 40–43. doi: 10.1179/1467010013Z.0000000000133

Firszt, J. B., Reeder, R. M., Dwyer, N. Y., Burton, H., and Holden, L. K. (2015). Localization training results in individuals with unilateral severe to profound hearing loss. *Hear. Res.* 319, 48–55. doi: 10.1016/j.heares.2014.11.005

Funnell, S., and Rogers, P. (2011). *Purposeful Program Theory: Effective Use of Theories of Change and Logic Models*.

Furber, G., and Segal, L. (2015). The validity of the Child Health Utility instrument (CHU9D) as a routine outcome measure for use in child and adolescent mental health services. *Health Qual. Life Outcomes* 13:22. doi: 10.1186/s12955-015-0218-4

Geers, A. E., Nicholas, J. G., and Sedey, A. L. (2003). Language skills of children with early cochlear implantation. *Ear Hear.* 24, 46S–58S. doi: 10.1097/01.AUD.0000051689.57380.1B

Grange, J. A., Culling, J. F., Bardsley, B., Mackinney, L. I., Hughes, S. E., Backhouse, S. S., et al. (2018). Turn an ear to hear: how hearing-impaired listeners can exploit head orientation to enhance their speech intelligibility in noisy social settings. *Trends Hear.* 22:2331216518802701. doi: 10.1177/2331216518802701

Grimm, G., Kollmeier, B., and Hohmann, V. (2016). Spatial acoustic scenarios in multichannel loudspeaker systems for hearing aid evaluation. *J. Am. Acad. Audiol.* 27, 557–566. doi: 10.3766/jaaa.15095

Horsman, J., Furlong, W., Feeny, D., and Torrance, G. (2003). The Health Utilities Index (HUI®): concepts, measurement properties and applications. *Health Qual. Life Outcomes* 1:54. doi: 10.1186/1477-7525-1-54

Killan, C., Scally, A., Killan, E., Totten, C., and Raine, C. (2019). Factors affecting sound-source localization in children with simultaneous or sequential bilateral cochlear implants. *Ear Hear.* 40, 870–877. doi: 10.1097/AUD.0000000000000666

Lammers, M. J., van der Heijden, G. J., Pourier, V. E., and Grolman, W. (2014). Bilateral cochlear implantation in children: a systematic review and best-evidence synthesis. *Laryngoscope* 124, 1694–1699. doi: 10.1002/lary.24582

Levtov, Y., Picinali, L., D'Cruz, M., and Simeone, L. (2016). 3D tune-in: the use of 3D sound and gamification to aid better adoption of hearing aid technologies. *Audio Eng. Soc. Conv.* 140.

Lovett, R. E., Vickers, D. A., and Summerfield, A. Q. (2015). Bilateral cochlear implantation for hearing-impaired children: criterion of candidacy derived from an observational study. *Ear Hear.* 36, 14–23. doi: 10.1097/AUD.0000000000000087

MacLeod, A., and Summerfield, Q. A. (1990). procedure for measuring auditory and audio-visual speech-reception thresholds for sentences in noise: rationale, evaluation, and recommendations for use. *Br. J. Audiol.* 24, 29–43. doi: 10.3109/03005369009077840

Mather, J., Gregory, S., and Archbold, S. (2011). The experiences of deaf young people with sequential bilateral cochlear implants. *Deaf. Educ. Int.* 13, 152–172. doi: 10.1179/1557069X11Y.0000000008

Mathew, R., Vickers, D., Boyle, P., Shaida, A., Selvadurai, D., Jiang, D., et al. (2018). Development of electrophysiological and behavioural measures of electrode discrimination in adult cochlear implant users. *Hear Res.* 367, 74–87. doi: 10.1016/j.heares.2018.07.002

Meta (2024). *Meta Quest 2 Health and Safety*. Available at: https://www.meta.com/gb/quest/safety-center/quest-2/ (accessed August 1, 2024).

Moore, G. F., Audrey, S., Barker, M., Bond, L., Bonell, C., Hardeman, W., et al. (2015). Process evaluation of complex interventions: Medical Research Council guidance. *Br. Med.J.* 350:h1258. doi: 10.1136/bmj.h1258

Myhrum, M., Strøm-Roum, H., Heldahl, M. G., Rødvik, A. K., Eksveen, B., Landsvik, B., et al. (2017). Sequential bilateral cochlear implantation in children: outcome of the second implant and long-term use. *Ear Hear.* 38, 301–313. doi: 10.1097/AUD.0000000000000383

National Health Service (2019). *The NHS Long Term Plan*.

Parkinson, S., Eatough, V., Holmes, J., Stapley, E., and Midgley, N. (2016). Framework analysis: a worked example of a study exploring young people's experiences of depression. *Qualit. Res. Psychol.* 13, 109–129. doi: 10.1080/14780887.2015.1119228

Parmar, B. J., Rajasingam, S. L., Bizley, J. K., and Vickers, D. A. (2022). Factors affecting the use of speech testing in adult audiology. *Am. J. Audiol.* 31, 528–540. doi: 10.1044/2022_AJA-21-00233

Peixoto, M. C., Spratley, J., Oliveira, G., Martins, J., Bastos, J., Ribeiro, C., et al. (2013). Effectiveness of cochlear implants in children: long term results. *Int. J. Pediatr. Otorhinolaryngol.* 77, 462–468. doi: 10.1016/j.ijporl.2012.12.005

Piaget, J. (1971). *The Theory of Stages in Cognitive Development. Measurement and Piaget* (New York, NY: McGraw-Hill), 283-ix.

Rayes, H., Al-Malky, G., and Vickers, D. (2019). Systematic review of auditory training in pediatric cochlear implant recipients. *J. Speech Lang. Hear. Res.* 62, 1574–1593. doi: 10.1044/2019_JSLHR-H-18-0252

Reactify (2024). *Adding New Dimensions to Music*. Available at: https://reactify.co.uk/ (accessed August 1, 2024).

Rechichi, C., Mojà, G. D., and Aragona, P. (2017). Video game vision syndrome: a new clinical picture in children? *J. Pediatr. Ophthalmol. Strabismus* 54, 346–355. doi: 10.3928/01913913-20170510-01

Salanger, M., Lewis, D., Vallier, T., McDermott, T., and Dergan, A. (2020). Applying virtual reality to audiovisual speech perception tasks in children. *Am. J. Audiol.* 29, 244–258. doi: 10.1044/2020_AJA-19-00004

Salorio-Corbetto, M., Williges, B., Lamping, W., Picinali, L., and Vickers, D. (2022). Evaluating spatial hearing using a dual-task approach in a virtual-acoustics environment. *Front. Neurosci.* 16:787153. doi: 10.3389/fnins.2022.787153

Sarant, J., Harris, D., Bennet, L., and Bant, S. (2014). Bilateral versus unilateral cochlear implants in children: a study of spoken language outcomes. *Ear Hear.* 35, 396–409. doi: 10.1097/AUD.0000000000000022

Sharma, S. D., Cushing, S. L., Papsin, B. C., and Gordon, K. A. (2020). Hearing and speech benefits of cochlear implantation in children: a review of the literature. *Int. J. Pediatr. Otorhinolaryngol.* 133:109984. doi: 10.1016/j.ijporl.2020.109984

Shim, L., Lee, J., Han, J. H., Jeon, H., Hong, S. K., Lee, H. J., et al. (2023). Feasibility of virtual reality-based auditory localization training with binaurally recorded auditory stimuli for patients with single-sided deafness. *Clin. Exp. Otorhinolaryngol.* 16, 217–224. doi: 10.21053/ceo.2023.00206

Skivington, K., Matthews, L., Simpson, S. A., Craig, P., Baird, J., Blazeby, J. M., et al. (2021). A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. *Br. Med. J.* 374:n2061. doi: 10.1136/bmj.n2061

Somerset, S. P., Kitterick, A., and Developing, P. (2023). "A questionnaire for children with hearing loss—the york binaural hearing related quality of life—youth (YBHRQL-Y)," in *British Academy of Audiology Conference*.

Sparreboom, M., Langereis, M. C., Snik, A. F., and Mylanus, E. A. (2015). Long-term outcomes on spatial hearing, speech recognition and receptive vocabulary after sequential bilateral cochlear implantation in children. *Res. Dev. Disabil.* 36C, 328–337. doi: 10.1016/j.ridd.2014.10.030

Sparreboom, M., Leeuw, A. R., Snik, A. F. M., and Mylanus, E. A. M. (2012). Sequential bilateral cochlear implantation in children: parents' perspective and device use. *Int. J. Pediatr. Otorhinolaryngol.* 76, 339–344. doi: 10.1016/j.ijporl.2011.12.004

Steadman, M. A., Kim, C., Lestang, J.-., H., Goodman, D. F. M., and Picinali, L. (2019). Short-term effects of sound localization training in virtual reality. *Sci. Rep.* 9:18284. doi: 10.1038/s41598-019-54811-w

Stitt, P., Picinali, L., and Katz, B. F. G. (2019). Auditory accommodation to poorly matched non-individual spectral localization cues through active learning. *Sci. Rep.* 9:1063. doi: 10.1038/s41598-018-37873-0

Summerfield, A. Q., Kitterick, P. T., and Goman, A. M. (2022). Development and critical evaluation of a condition-specific preference-based measure sensitive to binaural hearing in adults: the york binaural hearing-related quality-of-life system. *Ear Hear.* 43, 379–397. doi: 10.1097/AUD.0000000000001101

Vickers, D., Salorio-Corbetto, M., Driver, S., Rocca, C., Levtov, Y., Sum, K., et al. (2021). Involving children and teenagers with bilateral cochlear implants in the design of the BEARS (both EARS) virtual reality training suite improves personalization. *Front. Digit. Health* 3:759723. doi: 10.3389/fdgth.2021.759723

Watson, V., Verschuur, C., and Lathlean, J. (2016). Exploring the experiences of teenagers with cochlear implants. *Cochl. Impl. Int.* 17, 293–301. doi: 10.1080/14670100.2016.1257472

Whitton, J. P., Hancock, K. E., Shannon, J. M., and Polley, D. B. (2017). Audiomotor perceptual training enhances speech intelligibility in background noise. *Curr. Biol.* 27, 3237–47.e6. doi: 10.1016/j.cub.2017.09.014

Yu, F., Li, H., Zhou, X., Tang, X., Galvin Iii, J. J., Fu, Q. J., et al. (2018). Effects of training on lateralization for simulations of cochlear implants and single-sided deafness. *Front. Hum. Neurosci.* 12:287. doi: 10.3389/fnhum.2018.00287

Zeng, F. G. (2022). Celebrating the one millionth cochlear implant. *JASA Expr. Lett.* 2:e077201. doi: 10.1121/10.0012825

Zheng, Y., Godar, S. P., and Litovsky, R. Y. (2015). Development of sound localization strategies in children with bilateral cochlear implants. *PLoS ONE* 10:e0135790. doi: 10.1371/journal.pone.0135790

Zheng, Y., Swanson, J., Koehnke, J., and Guan, J. (2022). Sound localization of listeners with normal hearing, impaired hearing, hearing aids, bone-anchored hearing instruments, and cochlear implants: a review. *Am. J. Audiol.* 31, 819–834. doi: 10.1044/2022_AJA-22-00006

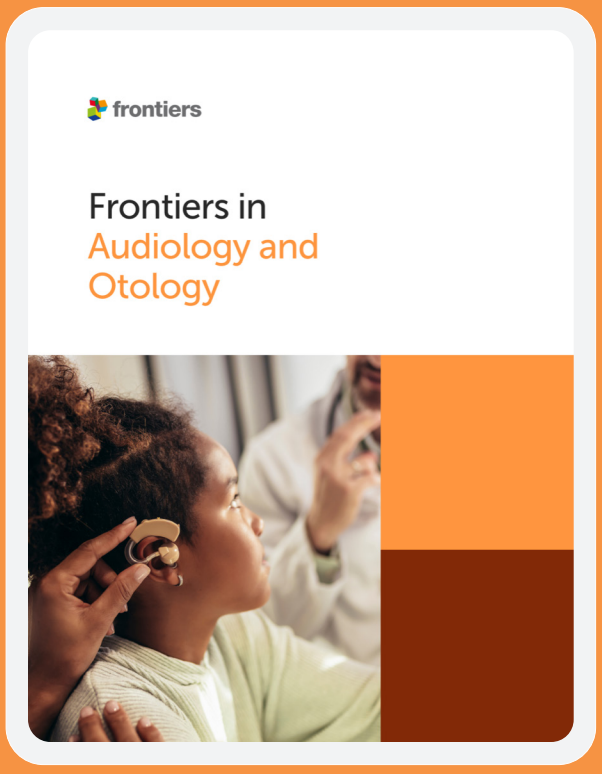
Frontiers in
Audiology and Otology

Explores all aspects of auditory research, otological disorders and the vestibular system

Advances clinical applications and technological solutions to better understand auditory and vestibular disorders, and provide improved solutions for patient treatment and rehabilitation

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact